sensors

Special Issue Reprint

Data, Signal and Image Processing and Applications in Sensors II

Edited by
Manuel José Cabral dos Santos Reis

mdpi.com/journal/sensors

MDPI

Data, Signal and Image Processing and Applications in Sensors II

Data, Signal and Image Processing and Applications in Sensors II

Editor

Manuel José Cabral dos Santos Reis

Basel • Beijing • Wuhan • Barcelona • Belgrade • Novi Sad • Cluj • Manchester

Editor
Manuel José Cabral dos
Santos Reis
University of Trás-os-Montes
e Alto Douro
Vila Real
Portugal

Editorial Office
MDPI AG
Grosspeteranlage 5
4052 Basel, Switzerland

This is a reprint of articles from the Special Issue published online in the open access journal *Sensors* (ISSN 1424-8220) (available at: https://www.mdpi.com/journal/sensors/special_issues/signal_sensors_II).

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

Lastname, A.A.; Lastname, B.B. Article Title. *Journal Name* **Year**, *Volume Number*, Page Range.

ISBN 978-3-7258-1561-6 (Hbk)
ISBN 978-3-7258-1562-3 (PDF)
doi.org/10.3390/books978-3-7258-1562-3

Contents

About the Editor

Manuel José Cabral dos Santos Reis

Manuel José Cabral dos Santos Reis received his PhD in Electrical Engineering from the University of Aveiro in 2001. He is currently an associate professor with "Agregação" at the Department of Engineering, School of Science and Technology, UTAD. His main areas of interest include the study and development of devices and systems for smart friendly environments; signal and image processing and applications; and the development of multimedia methods and tools applicable to teaching/learning, particularly the use of educational resources on the Internet. He is an integrated researcher at the Institute of Electronics and Informatics Engineering of Aveiro (IEETA), University of Aveiro. He has published more than 160 papers in various journals, conference proceedings, and book chapters, among others. Moreover, he has supervised 1 post-doc, 6 doctorate, and 33 master's degree students. He has participated in more than 70 events around the world and served as a research member in more than 30 projects, including 21 projects as the lead researcher. He owns 12 patents and utility models registrations, and he has won 5 awards.

sensors

MDPI

Editorial

Data, Signal and Image Processing and Applications in Sensors II

Manuel J. C. S. Reis [1,2]

[1] Engineering Department, University of Trás-os-Montes e Alto Douro (UTAD), 5001-801 Vila Real, Portugal; mcabral@utad.pt
[2] Institute of Electronics and Informatics Engineering of Aveiro (IEETA), 3810-193 Aveiro, Portugal

A vast and ever-growing amount of data in various domains and modalities is readily available, being the rapid advance of sensor technology one of its main contributor. However, presenting raw signal data collected directly from sensors is sometimes inappropriate, due to the presence of, for example, noise or distortion, among others. In order to obtain relevant and insightful metrics from sensors signals' data, further enhancement of the sensor signals acquired, such as the noise reduction in the one-dimensional electroencephalographic (EEG) signals or colour correction in the endoscopic images, and their analysis by computer-based medical systems, is needed. The processing of the data in itself and the consequent extraction of useful information are also vital and included in the topics of this Special Issue, being this an extension of the first special issue on this subject (https://www.mdpi.com/journal/sensors/special_issues/signal_sensors, accessed on 15 March 2024).

This second edition of this SI of Sensors aims to showcase progress in the advancement, assessment, and implementation of algorithms and methodologies for processing data, signals, and images across diverse sensor types and sensing approaches. Both empirical and theoretical findings, along with review articles, were taken into account.

The quantity of manuscripts submitted directly indicates the significant interest in this topic within the research community, with a total of 42 manuscripts received. Among these, 27 papers of high quality were accepted and published, while 15 papers were rejected. As customary, the Sensors journal upheld its standards by subjecting all submitted manuscripts to a thorough peer-review process.

In the forthcoming presentation, I will utilize the exact wording of the authors to effectively convey the contributions of each paper, as well as trying to provide the readers with a summary of each paper.

Pires et al., in contribution 1, discuss medicine's evolution towards personalized care, focusing on cardiovascular diseases. They propose an AI-based system to empower patients via continuous monitoring and personalized treatment. The system aims to realize 5P (Predictive, Preventive, Participatory, Personalized, and Precision) medicine principles using data from wearables and smart devices. Key features include learning algorithms for data analysis, event prediction, alarm generation, and healthy behaviour promotion. It aims to boost patient engagement and contact with healthcare professionals. Cardiovascular diseases are highlighted as major causes of disability and death, emphasizing proactive management. Computational intelligence and device data integration enable efficient healthcare management, representing a comprehensive approach to disease management. It is positioned to enhance patient well-being and promote global public health.

Hao et al., in contribution 2, introduced a spectrum correction algorithm, decomposition filtering-based dual-window correction (DFBDWC), for improving target distance accuracy in frequency modulated continuous wave (FMCW) laser ranging. Traditional methods face challenges from white Gaussian noise (WGN), spectrum leakage, and the picket fence effect, yielding unsatisfactory results. DFBDWC employs decomposition filtering and a dual-window approach to effectively mitigate these issues. Experimental validation demonstrates its superior performance compared to traditional methods discrete

Citation: Reis, M.J.C.S. Data, Signal and Image Processing and Applications in Sensors II. *Sensors* **2024**, *24*, 2555. https://doi.org/10.3390/s24082555

Received: 27 March 2024
Revised: 8 April 2024
Accepted: 12 April 2024
Published: 16 April 2024

Fourier transform algorithm, phase demodulation, enhanced cross-correlation, Ratio, Chirp Z-transform, and enhanced cross-correlation algorithm. DFBDWC significantly reduces the maximum error from 0.7937 m to 0.0407 m, improving accuracy and frequency resolution while minimizing the impact of noise and spectrum leakage. Accurate WGN estimation and effective filtering contribute to its success. Moreover, a double Hann window reduces spectrum leakage, while utilizing two main spectral lines enhances overall performance.

Gerasimova et al., in contribution 3, propose a novel memristive interface composed of two FitzHugh–Nagumo electronic neurons connected via a metal–oxide memristive synaptic device. A hardware–software complex based on a commercial data acquisition system is developed to record signals from a presynaptic neuron and transmit them to a postsynaptic neuron through the memristive device. Both numerical simulations and experiments demonstrate complex dynamics, including chaos and various types of neural synchronization. The system offers simplicity and real-time performance, with the amplitude of the presynaptic signal leading to potentiation of the memristive device and adaptive modulation of the postsynaptic neuron output. Due to its stochastic nature, the memristive interface simulates real synaptic connections, holding promise for neuro-prosthetic applications. The authors investigate the dynamics of two coupled FitzHugh–Nagumo neuron generators through a metal–oxide memristive device, showcasing stochastic plasticity and various synchronous regimes. The relative compactness and high sensitivity of the proposed neuro-memristive device make it promising for applications in bio-robotics and bioengineering.

Fiedler et al., in contribution 4, introduce Simultaneous Face and Person Detection (SFPD), aiming for real-time detection of faces and persons. Combining these tasks is essential for computer vision applications like face recognition and human–robot interaction. SFPD employs multi-task learning, addressing the lack of datasets with both annotations algorithmically. It utilizes a joint convolutional neural network backbone with shared feature maps and separate detection layers for each task. SFPD doesn't need auxiliary steps during training, such as pre-training individual network parts or additional annotations. Evaluation shows SFPD's effectiveness in detection performance and speed, achieving 40 frames per second. Comparative analysis demonstrates its superiority in processing speed, detection performance, or providing both face and person detections. Overall, SFPD offers a valuable real-time framework for various applications, especially in human–robot interaction scenarios.

Fuentes et al., in contribution 5, tackle real-time IoT data visualization without costly hardware. They propose an augmented reality (AR)-based solution using consumer-grade smartphones. The system enables real-time data visualization from IoT devices via AR, with added security. Tests confirm the solution's effectiveness in accessing IoT data, smartphone-device interactions, and identifying optimized AR markers. Results show the feasibility of using smartphones for IoT device management in diverse environments. Key contributions include an architecture for simplified AR IoT data visualization and a functional prototype validation. Future work may explore an AR marker generator for improved performance and usability.

Fuentes et al., in contribution 6, tackle Portugal's aging population issues, especially in rural areas where elders often face isolation and resource constraints. They propose an affordable Ambient Assisted Living (AAL) system to monitor elderly activities at home, respecting their privacy. Using low-cost IoT sensors and computer vision, the system recognizes elderly activities and was successfully tested in a simulated scenario. It enables remote caregiving, allowing independent living with assistance when needed, while ensuring privacy. The prototype, utilizing Raspberry Pi Zero W, effectively monitors specific home areas. Future enhancements may optimize Raspberry Pi processing, incorporate grey areas to reduce false positives, and introduce automatic alerts for caregivers. Overall, this solution promises to support independent living for elders, enhancing safety and well-being.

Pilyugina et al., in contribution 7, conducted a study to find effective feature extraction methods from auditory steady-state responses (ASSR) data to differentiate between

auditory octave illusion and non-illusion groups. Various feature selection techniques, including univariate selection, recursive feature elimination, principal component analysis, and feature importance, were compared. Machine learning algorithms such as linear regression, random forest, and support vector machine (SVM) were employed to evaluate these methods. The study revealed that combining univariate selection with SVM achieved the highest accuracy of 75%, surpassing the 66.6% accuracy obtained without feature selection. These findings provide a foundation for further research into understanding the mechanism behind the octave illusion phenomenon and developing automatic classification algorithms for octave illusions.

Motor imagery (MI)-based brain–computer interfaces (BCIs) are crucial for device control via brain activity. Despite this, complex inter-communication among brain regions during motor tasks presents challenges for isolating relevant neural patterns. To tackle this, Awais et al., in contribution 8, utilized effective brain connectivity measures like partial directed coherence (PDC) to capture inter-channel/region relationships during motor imagination. Statistical analysis identified significant connectivity pairs, and four classification algorithms (SVM, KNN, decision tree, and probabilistic neural network) predicted MI patterns using PhysioNet EEG data. Results showed the probabilistic neural network (PNN) classifier with PDC features achieved 98.65% accuracy, highlighting PDC's superiority over DTF in classification. Leveraging brain connectivity enhances neural pattern understanding, advancing BCI applications. Future research might explore graph theory and optimization for improved real-time BCI applications, especially for those with motor disabilities.

Bas-Calopa et al., in contribution 9, investigate the impact of low-pressure environments and high-operating frequencies on visual corona discharges, crucial for understanding arc tracking and insulation degradation in aircraft wiring systems as more electric and all-electric aircraft become prevalent. Experimentation employs a rod-to-plane electrode setup across pressure (20–100 kPa) and frequency (50–1000 Hz) ranges relevant to aircraft applications. A low-cost, high-resolution CMOS imaging sensor is utilized for corona detection, offering simplicity and sensitivity, while leakage current analysis serves as a complementary method. Results reveal that corona extinction voltage (CEV) increases notably with air pressure, while frequency exhibits a lesser effect, causing CEV to decrease within certain pressure ranges. The CMOS sensor demonstrates sufficient sensitivity for corona detection in low-pressure environments across various frequencies, offering potential for insulation system design in modern aircraft. Additionally, the study underscores the comparable sensitivity between the CMOS sensor and leakage current analysis, with minor discrepancies diminishing at higher frequencies.

Ebrahimi et al., in contribution 10, investigate the effectiveness of Robust Principal Component Analysis (PCA) matrix decomposition alongside advanced methods (PCT, PPT, and PLST) for analyzing pulsed thermography thermal data in carbon fibre-reinforced polymer (CFRP) materials. Using an academic sample with artificial defects, they assess defect detection and segmentation using CNR and similarity coefficient. Results show significant CNR improvements with Robust PCA pre-processing, enhancing defect detectability by up to 164%, 237%, and 80% for different defect types. Pre-processing notably improves CNR for FBHs and POs, with enhancements ranging from 0.43% to 115.88% and from 13.48% to 216.63%, respectively. Postprocessing enhances results for FBHs and POs by 9.62% to 296.9% and 16.98% to 92.6%, respectively. Robust PCA enhances defect detectability for PCT, PPT, and PLST methods, surpassing PLST for 69% of defects. Pre-processing enhances segmentation potential for all methods, with PLST showing improvements for both pre- and post-processing. The study concludes that Robust PCA pre-processing substantially enhances anomaly detection in pulsed thermography for CFRP materials, with implications for NDT. Further research should extend these techniques to diverse materials for enhanced practicality in NDT.

Jang et al., in contribution 11, presented a new method for detecting vital signals using multiple radar systems to reduce signal degradation from body movement. By analyzing

phase variation in continuous-wave radar signals caused by respiration and heartbeat, the method employs two adjacent radars with different lines-of-sight to capture correlated signals, enhancing differences in organ movement asymmetry. Operating at different frequencies within the 5.8 GHz band and strategically positioned, the radars improve signal-to-noise ratio during vital signal detection. Experimental results showed 97.8% accuracy in vital signal detection, even with subjects moving at velocities up to 53.4 mm/s. The configuration and signal processing method effectively utilize asymmetrical organ movements, improving signal-to-noise ratio and detection accuracy, especially during body movement. Extensive testing demonstrated noise reduction in the low-frequency range and significant enhancements in signal-to-noise ratio and detection accuracy across various radar angles. Overall, this method offers robust vital signal detection, even in dynamic environments with substantial body movement.

Maddirala et al., in contribution 12, tackle the issue of eliminating eye-blink artifacts from single-channel electroencephalogram (EEG) signals, often recorded using portable EEG devices. These artifacts, stemming from eyelid blinking or eye movements, distort EEG measurements, impacting brain activity interpretation. Traditional artifact removal methods are inadequate for single-channel EEG signals, necessitating novel techniques. The proposed approach combines singular spectrum analysis with continuous wavelet transform and k-means clustering to remove eye-blink artifacts while retaining low-frequency EEG data. Assessment on synthetic and real EEG datasets validates the method's superiority over existing techniques. Focused on pre-frontal channel EEG signals, it holds promise for online applications employing such channels. The study highlights successful artifact removal without sacrificing original EEG information, suggesting potential in real-time EEG monitoring and classification tasks. Future research should examine the method's performance in classification scenarios, anticipating favourable outcomes based on demonstrated artifact removal effectiveness.

Murtiyoso et al. in contribution 13, proposed integrating AI-based semantic segmentation into photogrammetric workflows to automate semantically classified point cloud creation. Leveraging deep learning and semantic segmentation advancements, the method uses pretrained neural networks for automatic image masking and dense image matching. By starting with semantic classification in the photogrammetric process, the workflow is streamlined to generate labelled point clouds. Results demonstrate process automation feasibility, with promising assessments for specific classes like building facades and windows. Emphasizing the advantage of abundant 2D image label data for neural network training, challenges remain in handling underrepresented classes and optimizing training data generation. Future research should explore semantic photogrammetry in various settings, refining training data methods for close-range photogrammetry. Overall, the study provides a proof of concept for AI integration into photogrammetric tasks, setting the stage for semantic photogrammetry advancement.

Mohanna et al., in contribution 14, introduced a method to tackle radar shadow effects in FMCW radars, which hinder target discrimination when one target is in another's shadow region. Utilizing CNNs on spectrograms from STFT analysis, the method determines if a target is in another's shadow. Achieving 92% test accuracy with a 2.86% standard deviation, it effectively discerns scenarios with one or two targets. Using MobileNet architecture pretrained on Imagenet, the model attains high accuracy with low parameters, suitable for real-time use. Future research should test the solution on hardware like Raspberry Pi and extend it for tracking multiple moving targets in cluttered settings. While supervised learning is effective, an unsupervised approach may be needed for scenarios with unpredictable classes. Overall, the method offers a promising solution for mitigating radar shadow effects in FMCW radar, with diverse target detection and tracking applications.

Manian et al., in contribution 15, propose a semi-supervised method for labelling and classifying hyperspectral images, addressing the challenge of acquiring ground-truth data, which is time-consuming and resource-intensive. The method comprises two stages:

unsupervised and supervised. In the unsupervised stage, image enhancement and clustering generate ground-truth data. The supervised stage involves pre-processing, feature extraction, and ensemble learning using various machine learning models. The ensemble method achieves high accuracy, with gradient boosting performing the best. It's effective for classifying Lake Erie and Jasper hyperspectral datasets, achieving accuracy rates of 100% and 93.74%, respectively. Additionally, it efficiently detects cloud pixels and water pollutants, useful for environmental monitoring. The choice of normalization scheme and number of PCA bands significantly impacts model performance and efficiency. The method runs significantly faster on cloud servers, making it practical for large-scale image processing tasks. Overall, the semi-supervised ensemble method presents a robust solution for hyperspectral image labelling and classification, with applications in environmental monitoring and remote sensing.

Lee et al., in contribution 16, tackle motion blur in images captured by thermal and photon detectors. They propose a method to synthesize blurry images from sharp ones by analyzing thermal detector mechanisms. Their novel blur kernel rendering method integrates motion blur models with an inertial sensor in the thermal image domain. Evaluation of its accuracy is conducted through thermal image deblurring tasks using a synthetic blurry image dataset constructed from acquired thermal images, the first to include ground-truth images in this domain. Through qualitative and quantitative experiments, the authors demonstrate the superiority of their method over existing techniques. In summary, the paper analyzes differences between thermal and photon detectors, developing a novel motion blur model for thermal images and an effective blur kernel rendering method, validated through rigorous experimentation.

Xiang et al., in contribution 17, introduced a novel approach for multi-sensor data fusion, crucial for information-aware systems with diverse sensory devices. Their method integrates the cloud model and an enhanced evidence theory to handle conflicting and ambiguous data. Quantitative data is converted into qualitative form using the cloud model to construct basic probability assignments (BPA) for each data source's evidence. To resolve conflicts, similarity measures like Jousselme distance, cosine similarity, and Jaccard coefficient are combined to assess evidence similarity, while Hellinger distance calculates evidence credibility. Fusion is performed using Dempster's rule. Experimental results show superior convergence and precision, achieving up to 100% confidence in correct propositions. Applied to early indoor fire detection, the method enhances accuracy by 0.9–6.4% and reduces false alarm rates by 0.7–10.2% compared to traditional methods, validating its effectiveness. Overall, this strategy offers a robust solution for managing conflicting and ambiguous data in information-aware systems, with promising applications across various multi-sensor acquisition systems. Future research should explore its applicability across different systems and integrate homogeneous and heterogeneous data fusion algorithms to further enhance accuracy.

Tropea et al., in contribution 18, introduced an automatic recognition system developed under the SILPI project, aiming to classify stones from quarries in Calabria, Southern Italy. Their two-stage hybrid approach combines Convolutional Neural Networks (CNNs) for feature extraction with Machine Learning (ML) for classification. Transfer Learning (TL) is explored to enhance CNN performance, using pre-trained networks from ImageNet. The system achieves impressive results in predicting stone classes, excelling in image recognition tasks. While granite typologies posed challenges, the hybrid model effectively integrates DL for feature extraction and classical ML algorithms for classification. The ResNet50 CNN model paired with a k-Nearest Neighbors (kNN) classifier emerges as the most promising combination, offering high accuracy, efficient CNN parameter usage, and rapid inference times. Overall, the approach demonstrates the potential for creating user-friendly tools applicable across various fields, including archaeometry, diagnostics, and materials sciences, even for users lacking geological expertise.

Rogers et al., in contribution 19, assessed RF measurement accuracy using Kubios HRV Premium software alongside consumer-grade hardware: Movesense Medical sensor

single-channel ECG (MS ECG) and Polar H10 HR monitor. GE, RR intervals (from H10), and continuous ECG (from MS ECG) were collected from 21 participants during cycling exercises. Results showed strong correlations between reference GE and both H10 and MS ECG-derived RF. Median values differed statistically but were clinically negligible for H10 (about 1 breaths/min) and minimal for MS ECG (about 0.1 breaths/min). ECG-based RF measurement with MS ECG exhibited reduced bias and narrower limits of agreement than H10. The study concludes that MS ECG with Kubios HRV Premium software closely tracked reference RF during exercise, suggesting practical utility for endurance exercise. Additionally, the ECG-centric system outperformed RR interval-derived RF estimation, accurately capturing RF patterns during exercise ramps. Future studies should explore these findings across different exercise types and assess artifact and noise impact.

Benitez-Garcia et al., in contribution 20, introduced a new material translation method using Neural Style Transfer (NST). NST traditionally relies on reference image quality, which may not yield optimal results. To overcome this, their method incorporates automatic style image retrieval, selecting the ideal reference based on semantic similarity and distinctive material characteristics. Excluding style information during retrieval significantly enhances synthesized results. The method combines real-time material segmentation with NST to selectively transfer retrieved style image material to segmented object areas. Evaluation with different NST methods shows effectiveness, validated through human perceptual study indicating synthesized stone, wood, and metal images are perceived as real, surpassing photographs. Applications include creating alternate reality scenarios for users to experience environments with subtly modified objects. Future work should focus on synthesizing more materials and developing real-time material translation applications.

Dziech et al., in contribution 21, introduced a novel data-embedding technique based on the Periodic Haar Piecewise-Linear (PHL) transform. They explain the theoretical basis of the PHL transform and propose a watermarking method that embeds hidden information in the luminance channel of the original image using coefficients with low values. The method's effectiveness is assessed by measuring the visual quality and bit error rate (BER) of watermarked images with different embedded information lengths. Additionally, a method for detecting image manipulation is presented. The technique shows promise for applications in digital signal and image processing, particularly in scenarios requiring high imperceptibility, low BER, and robust information security, such as medical image processing. The proposed method offers a high capacity for hidden information while minimizing image distortion, making it suitable for multimedia systems and services, especially in medical applications. Further research should focus on enhancing the method's robustness against various attacks and exploring its potential applications across different domains.

The task of temporal action detection (TAD) in untrimmed videos is vital across various applications, predicting temporal boundaries and action class labels within videos. Current methods often use stacked convolutional blocks to capture long temporal structures but struggle with redundant information between frames and varying action durations. To tackle these issues, He et al., in contribution 22, propose a non-local temporal difference network with three key modules: chunk convolution, multiple temporal coordination (MTC), and temporal difference (TD). The CC module divides input sequences into chunks, extracting features from distant frames simultaneously. The MTC module aggregates multiscale temporal features without extra parameters, while the TD module enhances motion and boundary features with temporal attention weights. This approach achieves state-of-the-art results on ActivityNet-v1.3 and THUMOS-14 datasets, effectively capturing long-range temporal structures and enhancing TAD accuracy. Discussions underscore TAD challenges and the importance of efficient network design in modelling complex temporal relationships while considering video characteristics like varying action durations and information redundancy.

Conventional methods for repairing old photo damage are often slow and inefficient, relying on manual or semi-automatic processes that involve laborious marking of damaged

areas. Fully automatic repair methods lack control over damage detection, posing risks to preserving historical photos. To overcome these challenges, Kuo et al. propose a deep learning-based architecture in contribution 23 to automatically detect damaged areas in old photos. The model accurately marks damaged regions, reducing damage marking time to less than 0.01 s per photo. By eliminating manual marking, the method enhances efficiency and preserves photo integrity. The use of residual dense block modules improves detection accuracy, ensuring preservation of both damaged and undamaged areas without distortion. Overall, this method provides a more efficient and precise approach to old photo restoration than existing end-to-end methods.

The study by Christou et al., in contribution 24, explores the influence of window size on EEG signal classification for diagnosing epilepsy. Automated analysis using machine learning is essential due to the complexity of EEG waveforms and the sporadic occurrence of epileptic characteristics. Employing various classifiers, including neural networks and k-nearest neighbour, EEG data from the University of Bonn dataset are analyzed with different window lengths. Results reveal that larger window sizes, approximately 21 s, notably enhance classification accuracy across tested methods. Given epilepsy's significant impact, accurate and automated detection methods are crucial. The study underscores the importance of window size in EEG signal analysis and recommends epochs of 20–21 s for optimal classification performance.

Sebastião et al., in contribution 25 investigate pain perception by analyzing physiological responses, aiming to complement self-reporting methods in pain assessment. They recorded various physiological signals, such as ECG, EMG, EDA, and BP, during a pain-inducing protocol. Results demonstrated significant changes in physiological parameters during painful periods compared to non-painful ones, including increased heart rate and decreased PNS influence in ECG data, heightened muscle activity in EMG, and increased SNS activity in EDA. A novel data collection protocol enabled comprehensive analysis of ANS reactivity across body systems. The study highlights the importance of deeper physiological evaluation for understanding pain effects and suggests future research on multimodal classification for more reliable pain measurements. Limitations include the brief recording duration, emphasizing the need for longer protocols to explore SNS influences on the cardiovascular system further.

Wang et al., in contribution 26, introduced a denoising method tailored for partial discharge (PD) signals in mining cables. The method employs genetic algorithm optimization of variational mode decomposition (VMD) and wavelet thresholding to enhance the signal-to-noise ratio (SNR) by effectively separating PD signals from interference. Initially, the genetic algorithm optimizes VMD parameters such as the number of modal components (K) and quadratic penalty factor (α). Subsequently, VMD decomposition generates intrinsic mode functions (IMF), followed by wavelet threshold denoising of each IMF. The denoised IMF are then reconstructed to obtain the cleaned PD signal. Simulation and experimental verification confirm the method's feasibility and efficacy, highlighting the optimized VMD parameters' role in enhancing denoising performance and the synergy between VMD and wavelet thresholding for noise reduction without compromising transient processes. The method shows superior denoising ability, especially for PD signals with lower SNR, making it promising for PD monitoring in mining cables.

Shokouhyan et al., in contribution 27, highlight the significance of neuro-mechanical time delays in sensorimotor control, particularly in individuals with spinal cord injuries (SCI), impacting stabilization efficiency and system stability. Estimating these delays in SCI patients is crucial for designing effective rehabilitation exercises and assistive technologies. The study aims to estimate muscle onset activation in SCI individuals using four strategies on electromyography data. Results show that the total kinetic energy operator technique effectively reduces artifacts compared to classical filtering, while time-frequency techniques estimate longer delays due to lower frequency movement during seated balance. These estimated delays can inform sensory-motor control models and aid in designing tailored exercises and technologies for SCI rehabilitation.

As can be seen form the summaries presented above, 9 of the published works can be classified in the field of image and multidimensional signal processing and applications, 11 in the field of signal processing and applications, and 7 in the field of data processing and applications. Additionally, it is important to highlight the fact that 11 of these works have direct applications in health related areas.

Last, but not least, I want to extend my personal gratitude to all the authors and reviewers who have contributed to this Special Issue. The authors deserve recognition for their innovative ideas and solutions, while the reviewers deserve appreciation for dedicating their time and offering valuable improvement suggestions. Their outstanding efforts have enabled Sensors journal to showcase novel and compelling contributions in the realm of "Data, Signal, and Image Processing and Applications in Sensors II". I final word of thanks goes to the Sensors journal's staff for their continuous support and suggestions. Thank you to each and every one of you!

Conflicts of Interest: The authors declare no conflict of interest.

List of Contributions

1. Pires, I.M.; Denysyuk, H.V.; Villasana, M.V.; Sá, J.; Lameski, P.; Chorbev, I.; Zdravevski, E.; Trajkovik, V.; Morgado, J.F.; Garcia, N.M. Mobile 5P-Medicine Approach for Cardiovascular Patients. *Sensors* **2021**, *21*, 6986. https://doi.org/10.3390/s21216986
2. Hao, Y.; Song, P.; Wang, X.; Pan, Z. A Spectrum Correction Algorithm Based on Beat Signal of FMCW Laser Ranging System. *Sensors* **2021**, *21*, 5057. https://doi.org/10.3390/s21155057
3. Gerasimova, S.A.; Belov, A.I.; Korolev, D.S.; Guseinov, D.V.; Lebedeva, A.V.; Koryazhkina, M.N.; Mikhaylov, A.N.; Kazantsev, V.B.; Pisarchik, A.N. Stochastic Memristive Interface for Neural Signal Processing. *Sensors* **2021**, *21*, 5587. https://doi.org/10.3390/s21165587
4. Fiedler, M.A.; Werner, P.; Khalifa, A.; Al-Hamadi, A. SFPD: Simultaneous Face and Person Detection in Real-Time for Human–Robot Interaction. *Sensors* **2021**, *21*, 5918. https://doi.org/10.3390/s21175918
5. Fuentes, D.; Correia, L.; Costa, N.; Reis, A.; Barroso, J.; Pereira, A. SAR.IoT: Secured Augmented Reality for IoT Devices Management. *Sensors* **2021**, *21*, 6001. https://doi.org/10.3390/s21186001
6. Fuentes, D.; Correia, L.; Costa, N.; Reis, A.; Ribeiro, J.; Rabadão, C.; Barroso, J.; Pereira, A. IndoorCare: Low-Cost Elderly Activity Monitoring System through Image Processing. *Sensors* **2021**, *21*, 6051. https://doi.org/10.3390/s21186051
7. Pilyugina, N.; Tsukahara, A.; Tanaka, K. Comparing Methods of Feature Extraction of Brain Activities for Octave Illusion Classification Using Machine Learning. *Sensors* **2021**, *21*, 6407. https://doi.org/10.3390/s21196407
8. Awais, M.A.; Yusoff, M.Z.; Khan, D.M.; Yahya, N.; Kamel, N.; Ebrahim, M. Effective Connectivity for Decoding Electroencephalographic Motor Imagery Using a Probabilistic Neural Network. *Sensors* **2021**, *21*, 6570. https://doi.org/10.3390/s21196570
9. Bas-Calopa, P.; Riba, J.R.; Moreno-Eguilaz, M. Corona Discharge Characteristics under Variable Frequency and Pressure Environments. *Sensors* **2021**, *21*, 6676. https://doi.org/10.3390/s21196676
10. Ebrahimi, S.; Fleuret, J.R.; Klein, M.; Théroux, L.D.; Ibarra-Castanedo, C.; Maldague, X.P.V. Data Enhancement via Low-Rank Matrix Reconstruction in Pulsed Thermography for Carbon-Fibre-Reinforced Polymers. *Sensors* **2021**, *21*, 7185. https://doi.org/10.3390/s21217185
11. Jang, A.J.; Lee, I.S.; Yang, J.R. Vital Signal Detection Using Multi-Radar for Reductions in Body Movement Effects. *Sensors* **2021**, *21*, 7398. https://doi.org/10.3390/s21217398
12. Maddirala, A.K.; Veluvolu, K.C. SSA with CWT and k-Means for Eye-Blink Artifact Removal from Single-Channel EEG Signals. *Sensors* **2022**, *22*, 931. https://doi.org/10.3390/s22030931

13. Murtiyoso, A.; Pellis, E.; Grussenmeyer, P.; Landes, T.; Masiero, A. Towards Semantic Photogrammetry: Generating Semantically Rich Point Clouds from Architectural Close-Range Photogrammetry. *Sensors* **2022**, *22*, 966. https://doi.org/10.3390/s22030966
14. Mohanna, A.; Gianoglio, C.; Rizik, A.; Valle, M. A Convolutional Neural Network-Based Method for Discriminating Shadowed Targets in Frequency-Modulated Continuous-Wave Radar Systems. *Sensors* **2022**, *22*, 1048. https://doi.org/10.3390/s22031048
15. Manian, V.; Alfaro-Mejía, E.; Tokars, R.P. Hyperspectral Image Labeling and Classification Using an Ensemble Semi-Supervised Machine Learning Approach. *Sensors* **2022**, *22*, 1623. https://doi.org/10.3390/s22041623
16. Lee, K.; Ban, Y.; Kim, C. Motion Blur Kernel Rendering Using an Inertial Sensor: Interpreting the Mechanism of a Thermal Detector. *Sensors* **2022**, *22*, 1893. https://doi.org/10.3390/s22051893
17. Xiang, X.; Li, K.; Huang, B.; Cao, Y. A Multi-Sensor Data-Fusion Method Based on Cloud Model and Improved Evidence Theory. *Sensors* **2022**, *22*, 5902. https://doi.org/10.3390/s22155902
18. Tropea, M.; Fedele, G.; De Luca, R.; Miriello, D.; De Rango, F. Automatic Stones Classification through a CNN-Based Approach. *Sensors* **2022**, *22*, 6292. https://doi.org/10.3390/s22166292
19. Rogers, B.; Schaffarczyk, M.; Gronwald, T. Estimation of Respiratory Frequency in Women and Men by Kubios HRV Software Using the Polar H10 or Movesense Medical ECG Sensor during an Exercise Ramp. *Sensors* **2022**, *22*, 7156. https://doi.org/10.3390/s22197156
20. Benitez-Garcia, G.; Takahashi, H.; Yanai, K. Material Translation Based on Neural Style Transfer with Ideal Style Image Retrieval. *Sensors* **2022**, *22*, 7317. https://doi.org/10.3390/s22197317
21. Dziech, A.; Bogacki, P.; Derkacz, J. A New Method for Image Protection Using Periodic Haar Piecewise-Linear Transform and Watermarking Technique. *Sensors* **2022**, *22*, 8106. https://doi.org/10.3390/s22218106
22. He, Y.; Han, X.; Zhong, Y.; Wang, L. Non-Local Temporal Difference Network for Temporal Action Detection. *Sensors* **2022**, *22*, 8396. https://doi.org/10.3390/s22218396
23. Kuo, T.Y.; Wei, Y.J.; Su, P.C.; Lin, T.H. Learning-Based Image Damage Area Detection for Old Photo Recovery. *Sensors* **2022**, *22*, 8580. https://doi.org/10.3390/s22218580
24. Christou, V.; Miltiadous, A.; Tsoulos, I.; Karvounis, E.; Tzimourta, K.D.; Tsipouras, M.G.; Anastasopoulos, N.; Tzallas, A.T.; Giannakeas, N. Evaluating the Window Size's Role in Automatic EEG Epilepsy Detection. *Sensors* **2022**, *22*, 9233. https://doi.org/10.3390/s22239233
25. Sebastião, R.; Bento, A.; Brás, S. Analysis of Physiological Responses during Pain Induction. *Sensors* **2022**, *22*, 9276. https://doi.org/10.3390/s22239276
26. Wang, Y.; Chen, P.; Zhao, Y.; Sun, Y. A Denoising Method for Mining Cable PD Signal Based on Genetic Algorithm Optimization of VMD and Wavelet Threshold. *Sensors* **2022**, *22*, 9386. https://doi.org/10.3390/s22239386
27. Shokouhyan, S.M.; Blandeau, M.; Wallard, L.; Guerra, T.M.; Pudlo, P.; Gagnon, D.H.; Barbier, F. Sensorimotor Time Delay Estimation by EMG Signal Processing in People Living with Spinal Cord Injury. *Sensors* **2023**, *23*, 1132. https://doi.org/10.3390/s23031132

sensors

MDPI

Article

Sensorimotor Time Delay Estimation by EMG Signal Processing in People Living with Spinal Cord Injury

Seyed Mohammadreza Shokouhyan [1,*], Mathias Blandeau [1], Laura Wallard [1], Thierry Marie Guerra [1], Philippe Pudlo [1], Dany H. Gagnon [2] and Franck Barbier [1,3]

[1] University Polytechnique Hauts-de-France, CNRS, UMR 8201-LAMIH, F-59313 Valenciennes, France
[2] Pathokinesiology Laboratory, Center for Interdisciplinary Research in Rehabilitation of Greater Montréal (CRIR), Montréal, QC H3S 1M9, Canada
[3] INSA Hauts-de-France, F-59313 Valenciennes, France
* Correspondence: seyedmohammadreza.shokouhyan@uphf.fr

Abstract: Neuro mechanical time delay is inevitable in the sensorimotor control of the body due to sensory, transmission, signal processing and muscle activation delays. In essence, time delay reduces stabilization efficiency, leading to system instability (e.g., falls). For this reason, estimation of time delay in patients such as people living with spinal cord injury (SCI) can help therapists and biomechanics to design more appropriate exercise or assistive technologies in the rehabilitation procedure. In this study, we aim to estimate the muscle onset activation in SCI people by four strategies on EMG data. Seven complete SCI individuals participated in this study, and they maintained their stability during seated balance after a mechanical perturbation exerting at the level of the third thoracic vertebra between the scapulas. EMG activity of eight upper limb muscles were recorded during the stability. Two strategies based on the simple filtering (first strategy) approach and TKEO technique (second strategy) in the time domain and two other approaches of cepstral analysis (third strategy) and power spectrum (fourth strategy) in the time–frequency domain were performed in order to estimate the muscle onset. The results demonstrated that the TKEO technique could efficiently remove the electrocardiogram (ECG) and motion artifacts compared with the simple classical filtering approach. However, the first and second strategies failed to find muscle onset in several trials, which shows the weakness of these two strategies. The time–frequency techniques (cepstral analysis and power spectrum) estimated longer activation onset compared with the other two strategies in the time domain, which we associate with lower-frequency movement in the maintaining of sitting stability. In addition, no correlation was found for the muscle activation sequence nor for the estimated delay value, which is most likely caused by motion redundancy and different stabilization strategies in each participant. The estimated time delay can be used in developing a sensory motor control model of the body. It not only can help therapists and biomechanics to understand the underlying mechanisms of body, but also can be useful in developing assistive technologies based on their stability mechanism.

Keywords: spinal cord injury; physiological time delay; Teager–Kaiser Energy Operator; cepstral analysis; power spectrum; EMG

Citation: Shokouhyan, S.M.; Blandeau, M.; Wallard, L.; Guerra, T.M.; Pudlo, P.; Gagnon, D.H.; Barbier, F. Sensorimotor Time Delay Estimation by EMG Signal Processing in People Living with Spinal Cord Injury. *Sensors* **2023**, *23*, 1132. https://doi.org/10.3390/s23031132

Academic Editor: Manuel José Cabral dos Santos Reis

Received: 29 October 2022
Revised: 9 January 2023
Accepted: 17 January 2023
Published: 18 January 2023

1. Introduction

During human sensory motor control, different sensory information is sent to the Central Nervous System (CNS), which processes the data and sends motor commands to various muscles in order to maintain body stability during activities. However, this process is affected by time delays, including the feedback delay due to neural transmission, the motor command delay due to the information process in the CNS [1] and finally an electromechanical time delay due to muscle activation delays [2]. Estimation of these time delays is crucial because higher values of total delay induce stabilization performance degradation leading to system instability [3]. On the other hand, estimating the time delay

can help us to understand the underlying mechanisms of sensory motor control of the body. Many studies have shown that time delay changes with exercise [4] and is longer in patients compared with healthy individuals as well as elderly people compared with young individuals [5,6]. Specifically, time delay estimation allows therapists and biomechanics to have better insight regarding designing exercise in addition to assistive development that can be helpful in the rehabilitation process and in improving their performance during daily activities. Various models or simulations have been developed to figure out the sensory motor control mechanism. Using incorrect parameter values in the system can lead to wrong results, interpretation and subsequently wrong rehabilitation decisions or a nonfunctional assistive device. Physiological time delay is thus a crucial parameter in modeling sitting stability.

People living with a complete spinal cord injury (SCI) have numerous issues in stabilizing their body due to a lack of sensory information and joint torques below their injury level. Specifically, any injury in their lumbar level results in damage to their back and intervertebral muscles, which are crucial in stabilizing the inherently unstable spine [7,8]. Thus, after an SCI, patients use their upper limbs and head rather than their muscles in the lumbar level in order to maintain their stability [9]. Therefore, sitting stability will be the first and most important goal of rehabilitation for them [10]. A better understanding of the underlying mechanism of sitting stability can be helpful in employing the best rehabilitation strategy or assistive technologies for SCI people. In addition, several studies performed multiple experimental tests in the presence of perturbation and developed different models to estimate joint torques, kinematic variables that can be effective in identifying the employed stability mechanisms by SCI individuals. Blandeau et al. [11,12] used a time-delayed 2 DOF H2AT model for sitting stability in SCI people such that the head and both arms could slide relative to the trunk rotating at the lumbar level. In this study, the trunk angle and the position of the head and arm center of mass (COM) were estimated by a nonlinear observer tuned using classical optimization techniques based on Linear Matrix Inequalities (LMI). Convergence towards the experimental trajectories is therefore proven using such methodology. In another investigation [13], they designed a nonlinear PI descriptor observer to estimate the body kinematics and unknown inputs in an H2AT model. Guerra et al. [14] estimated the inputs in the H2AT model by an application-oriented control law. This problem resumes in stabilizing an open-loop unstable underactuated nonlinear system with a time-varying delayed control input, which is a difficult problem to control, and was solved efficiently in [14]. Furthermore, in another study [15], a new model was developed for SCI patients to understand the underlying mechanisms of their body sensory motor control system. However, though these studies developed various models to understand the stability mechanisms in SCI patients, these models cannot be used to estimate the time delay value, which is crucial in stabilizing the employed models, and the stability strategy can be changed with different values of time delay.

Other studies tried to estimate the physiological time delay in healthy people and in patients by different approaches of simulation, experimental and combined strategies. The authors of [16] used an experimental protocol and were able to estimate the time delay between 66 to 99 milliseconds in healthy and lower back pain patients by analyzing the electrical muscle activity (EMG) in the presence of an external perturbation in both anterior–posterior and medio-lateral directions in seated balance. Other investigations were also able to estimate the total physiological time delay by analyzing the EMG data for healthy controls and patients [17–19] in seated balance and stance balance [4,20,20,21]. Instead of EMG, other studies estimated longer time delays by focusing on COP and kinematic data [22,23]. On the other hand, numerous investigations performed data analysis to estimate the time delay by using multiple clinical data including EMG, center of pressure or joint torque and kinematic data [24–33].

In addition, several studies [34–40] developed models to estimate not only the time delay but also other parameters such as joint torques, stiffness, damping, etc. In these studies, a model with multiple unknown parameters was developed in which the parameters were

determined using experimental trajectories and an optimization approach. Furthermore, some other investigations used different techniques such as Kalman filter [41], Cepstral analysis in the time–frequency domain [42] and frequency analysis [43] to estimate the time delay. Despite the fact that numerous studies estimate the time delay by different signal processing approaches in healthy and varieties of patients, to the best of our knowledge, no study was conducted for time delay estimation in SCI people during sitting stabilization. Therefore, the main motivation of this study is to estimate the physiological time delay in SCI patients during seated balance through four classical methodologies found in the literature. Regarding the novelty of this work, to the best of our knowledge, we found no study in the literature dealing with the following: 1. the time delay estimation in SCI people during sitting stability and 2. the comparison of various methods for time delay estimation. The article is organized as follows. Section 2 presents the Materials and Methods section with participants' characteristics, data acquisition and experimental protocol. Results and all estimated values of time delay are shown in the Section 3. In the Section 4, results are discussed and compared with other studies. Finally, the Section 5 presents perspectives and closes the paper with a conclusion.

2. Materials and Methods

2.1. Participants

Seven complete SCI subjects (ASIA-A, level of injury above T6) with mean age 39.7 years (SD 12.4) participated in this study. Ethical approval was obtained from the Research Ethics Committee of the Center for Interdisciplinary Research in Rehabilitation of Greater Montreal (CRIR-1083-0515R). The participants read and signed the informed consent form prior to initiating the measurements. Physical characteristics of participants are shown in Table 1.

Table 1. Physical characteristics of participants.

ID	Age	Sex	Weight (kg)	Height (cm)	IMC	Injury Level	ASIA	TIC	Injury Age (Months)
1	33	F	58.5	162.6	22.1	T6	A	0.5	112
2	33	M	59.9	177.8	18.9	T4	A	0	126
3	35	M	76.2	178	24.0	T6	A	0	147
4	31	F	51.7	157.5	20.8	T6	B	0.5	95
5	44	M	63	165	23.1	T6	A	0.5	161
6	57	M	94.8	185	27.7	T4	A	0.5	185
7	45	F	72.1	168	25.5	T4	A	0	131

2.2. Experimental Protocol

Participants were asked to maintain their sitting stability on a height-adjustable table without back support with hip and knees flexed to 90°, feet resting on the floor and upper limbs flexed to 90° at the elbow level. When sitting stability was achieved, a light destabilizing force was randomly applied at the level of the third thoracic vertebra between the scapulas. The destabilizing force was generated via an impact with a foam-coated wooden pole such that a pressure sensor was added on the tip to define the contact instant (see Figure 1). After one or two familiarization trials with the destabilizing force, each subject completed a minimum of 11 acquisitions. The start time of the trial was vocally announced to participants, at which time they rose their arms and maintained their stability before the perturbation. Their stability was visually evaluated by the examiner, and the time instant was recorded by a synchronized hand switch. Then, the perturbation was exerted at a random time, and participants should have regained their stability. Their status was again visually assessed, and the time instant recorded when they achieved their stability.

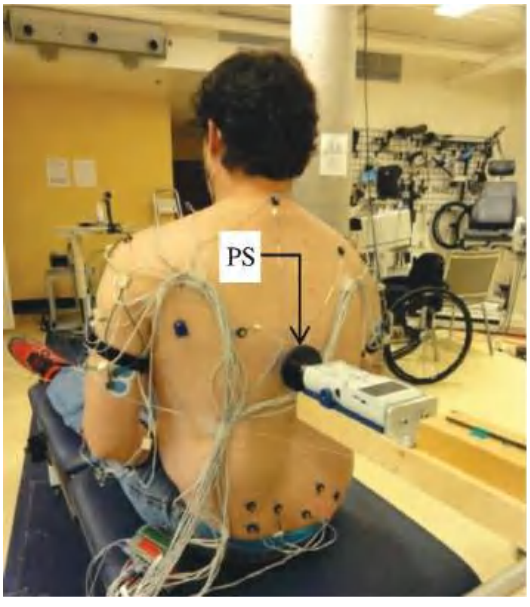


Figure 1. Experimental setup. The perturbation was applied at the level of the third thoracic vertebra.

The experimental protocol total duration was about one hour. The first half hour was dedicated to welcoming the subject, receiving his/her agreement for participating in the experiment and finally installing the EMG. The acquisition lasted for approximately 20 min with up to 1 min break between each acquisition. The last 10 min were dedicated to instrumentation removing and obtaining feedback from the subject.

2.3. Instruments and Data Acquisition

EMG signals were recorded from the following eight upper limb and trunk muscles: Deltoid Anterior (DA), Deltoid Posterior (DP), Pectoralis Major Clavicular (PMC), Pectoralis Major Sternal (PMS), Biceps Brachii (BB), Triceps Brachii (TB), Trapezius Descending (TD) and Latissimus Dorsi (LD). The skin area was cleaned with alcohol wipes and the electrodes were attached in pairs with a center-to-center distance of 25 mm, based upon recommendations reported in the previous literature [44]. After similar skin preparations, a ground electrode was attached to the anterior aspect of the leg over the tibial bone. The EMG signals were recorded with a commercially available EMG system (TeleMyo 900, Noraxon, Scottsdale, Arizona, USA). All EMG signals and hand switch data were sampled at 1200 Hz.

2.4. Data Analysis

In this study, two strategies in the time domain (first and second strategies) and two strategies (third and fourth strategies) in the time–frequency domain were used in order to estimate the time delay in SCI patients by analyzing the EMG data. In addition, the time between the earliest and latest muscle onset was computed as the range of muscle onset. All data analyses were performed with Matlab R2022b software.

2.4.1. First Strategy

At first, all EMG signals were analog filtered using a band pass filter between 30 to 500 Hz by 6th order Butterworth filter, rectified and then low-pass filtered at 100 Hz [45]. The mean and standard deviation (SD) of the signal were computed between 1.5 to 0.5 s immediately before the perturbation. Response onset latencies were

determined as the time at which the rectified EMG signal exceeded a threshold of 2×SD above the mean baseline for a period of at least 25 data points [4,46,47]. EMG onset latencies were computed for all muscles and then the average and SD were calculated for all trials in all subjects.

2.4.2. Second Strategy (TKEO)

In this strategy, the raw data were first rectified and high pass filtered at 20 Hz by 6th order Butterworth filter to remove motion and electrocardiogram (ECG) artifacts. Then, the nonlinear Teager–Kaiser Energy Operator (TKEO) [48] was employed and the data were filtered again (6th order, zero-phase low-pass filter at 50 Hz) for smoothing the signal. The TKEO function (T) is defined as below:

$$T[x(n)] = x^2(n) - x(n+1)x(n-1) \quad (1)$$

where x represents the rectified and filtered EMG signal and n the sample value. The onset of the muscles defined when the mean value of the smoothed signal exceeded a threshold of the mean plus two standard deviations away from the baseline for more than 25 consecutive samples [32,49]. The mean and SD of baseline were computed from 1.5 to 0.5 s right before the perturbation. Eventually, the response latency was defined as the time between the perturbation instant and onset of each muscle. The response latencies were measured for all muscles and then averaged, and the standard deviations were calculated for all trials and participants. For some acquisitions, the threshold of mean ± 2SD of baseline was not reached by the EMG signal, yielding no onset found. Moreover, when the onset was found below 20 ms, the delay was considered as not found because it was inconsistent with the physiological signal.

2.4.3. Third Strategy (Cepstral Analysis)

The feature of neutral delay-differential equations is mainly that the delay of the neutral part can be detected in the cepstrum of the output signal, which motivated one study [42] to estimate the delay of the acceleration feedback term in stick balancing tasks on kinematic data for healthy individuals. Thus, the cepstral analysis was used in this study as the third strategy for time delay estimation in SCI people. At first, the cepstral transformed signal of each EMG signal was obtained from the smoothed signals of the second strategy (TKEO) as shown in equation 2, in which $\mathcal{F}$ and $\mathcal{F}^{-1}$ represent Fourier transform and $T(n)$ is the signal time series after performing the TKEO technique. The frequency domain of 0–0.5 s was examined to find the sharp peaks. The instant of the maximum value was defined as the response onset and the response delay was identified as the time between the perturbation instant and response onset for each muscle. Mean and SD of all muscle onsets were then computed for all trials and subjects.

$$C_p = \mathcal{F}^{-1}\{log[\mathcal{F}\{T(n)\}]\} \quad (2)$$

2.4.4. Fourth Strategy (Power Spectrum)

In this approach, the power spectrum analysis was used to estimate the physiological time delay in SCI patients. The power spectrum of each smoothed signal [43] from the second strategy (TKEO) was extracted over time and frequency as shown in Equation (3), where $|P(f)|^2$ equals the energy density function over frequency. It was observed that most of the signal power is less than 10 Hz, thus the signal power was averaged between 0 to 10 Hz. It was assumed that the instant of power peaks could demonstrate the response onset. Therefore, the time domain of 0 to 0.5 s was investigated to find the instant of the maximum value. Eventually, the physiological time delay was defined as the time between the perturbation instant and when the averaged power signal reaches its maximum value. Mean and SD of the estimated values were then computed for all trials and participants. In

addition, in all strategies, the number of estimated muscle onsets higher than 20 ms and less than 500 ms were found as consistent values with physiological time delay.

$$E = \int_{-\infty}^{\infty} |T(n)|^2 dt = \int_{-\infty}^{\infty} |P(f)|^2 df \quad (3)$$

The algorithm of each strategy is shown in the Figure 2.

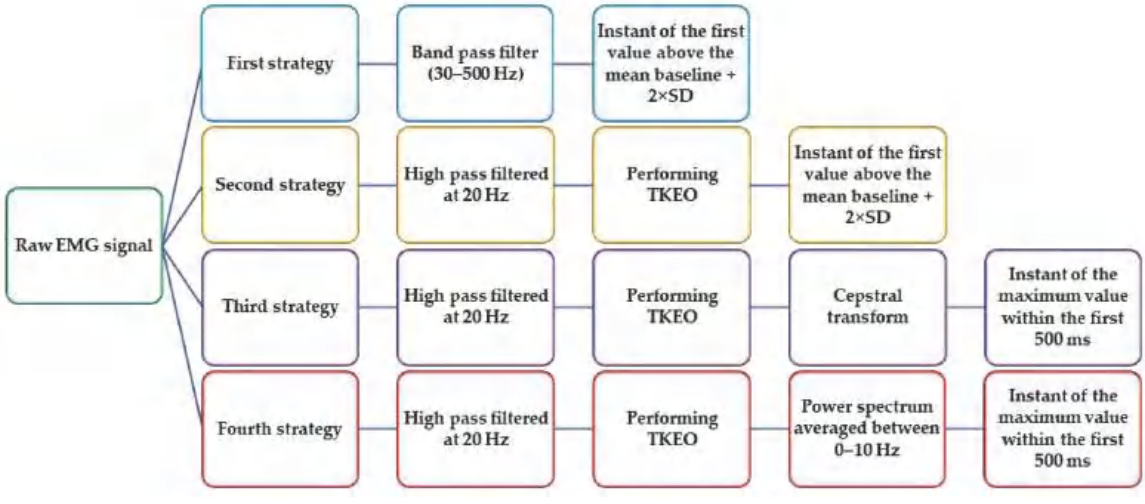


Figure 2. Flow chart of all four strategies used for time delay estimation.

2.5. Statistical Analysis

For evaluation of the ECG removing artifacts to form the EMG signals, a statistical metric of the Robust Measures of Kurtosis (KR_2) [50–52] was used in this study using the equation below:

$$KR_2 = \frac{F^{-1}(0.975) - F^{-1}(0.025)}{F^{-1}(0.75) - F^{-1}(0.25)} - 2.91 \quad (4)$$

where F^{-1} is the inverse cumulative distribution function (quantile function) of the time series data x. Values F^{-1} (0.975) = $-F^{-1}$ (0.025) = 1.96 and F^{-1} (0.75) = $-F^{-1}$ (0.25) = 0.6745 were obtained for the standard Gaussian distribution. Thus, KR_2 is zero if the data x has Gaussian distribution. The methods to evaluate statistical characteristics in estimating the Probability Density Function (PDF) shapes of EMG signals were composed of two stages. First, the PDF was estimated by kernel smoothing with a Gaussian kernel [53] from all time points, and this smooth density was discretized to 1001 bins of width 0.01 that partitioned the range from -1 to +1. Eventually, the average and standard deviation of KR_2 were calculated over all trials and subjects. Spearman and Pearson correlation coefficients were calculated to evaluate correlation in muscle activation sequence and delay value in all trials and subjects, respectively [54]. In addition, the hypothesis of distribution in the normal family was examined for the values of all estimated values for 8 muscles and 4 strategies. Significant effects of muscles and strategies (8×4) were evaluated by a two-way ANOVA on the dependent variable of estimated time delay. The effect was considered significant if the p-value was less than 0.05.

3. Results

The EMG signal and power spectrum of one subject are shown in Figures 3 and 4, respectively, and all dashed lines represent the perturbation instant. During the first 4 s, the subject keeps his arms down on his lower limbs, which explains the low EMG activity.

Mean and SD values of all estimated time delays based on four strategies are demonstrated in Figure 5. The results show that the third and fourth strategies estimated longer time delays compared with the other strategies in most muscles. In addition, the sequence of muscle activation is shown by numbers in each column bar of mean value. It is clear that the sequence of muscle activation changes based on the employed strategies for the estimation. However, the Trapezius Descending and Deltoid Anterior are activated later than other muscles during the posture stabilization for each strategy.

The results of Kurtosis robustness analysis are shown in Figure 6. It can be observed that the TKEO technique could appropriately remove the ECG and motion artifacts in muscle activities. In contrast, the results for the first strategy showed that the KR_2 value is far from zero as well for the Gaussian distribution, and its value is even closer to the unfiltered data, which shows less performance in removing ECG and motion artifacts compared with the TKEO strategy.

Descriptive results of four strategies are shown in Table 2. It can be seen that the first and second strategies sometimes failed to find the muscle onset. Furthermore, the result shows that more detections of Latissimus Dorsi onset were found compared with other muscles in the first and second strategies. On the other hand, the third and fourth strategies were able to estimate more time delays consistent with the actual physiological value.

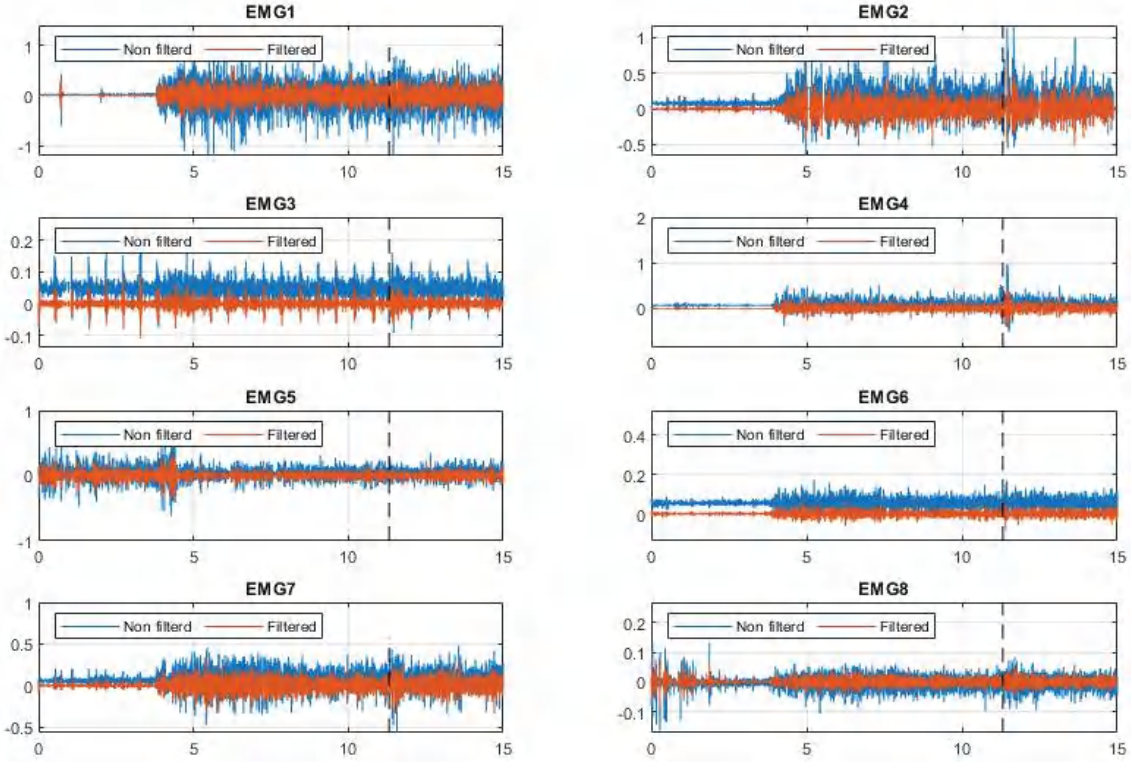


Figure 3. Raw and band pass filtered EMG data for one trial of subject number 7. EMG1 to EMG8 represent Deltoid Anterior, Pectoralis Major Clavicular, Pectoralis Major Sternal, Biceps Brachii, Trapezius Descending, Deltoid Posterior, Latissimus Dorsi and Triceps Brachii, respectively. The black dashed line represents the perturbation instant.

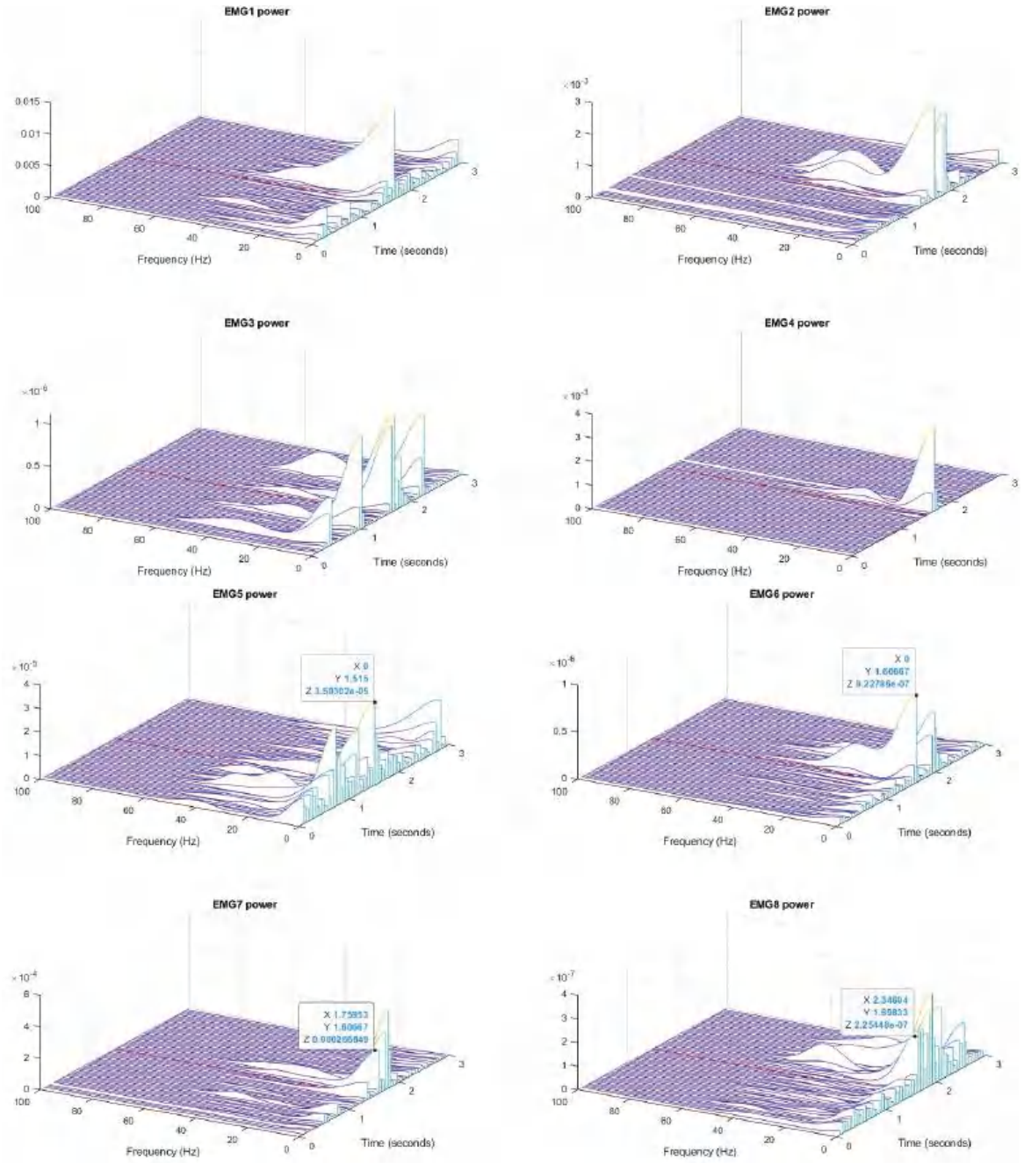


Figure 4. Power spectrum for one trial of subject number 7. EMG1 to EMG8 represent Deltoid Anterior, Pectoralis Major Clavicular, Pectoralis Major Sternal, Biceps Brachii, Trapezius Descending, Deltoid Posterior, Latissimus Dorsi and Triceps Brachii, respectively. The red dashed line represents the perturbation instant.

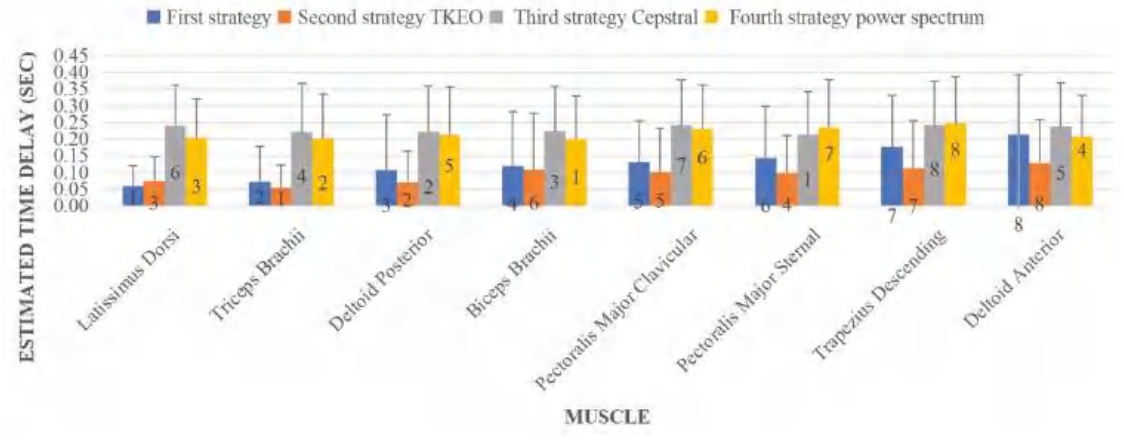


Figure 5. Estimated muscle onset based on different strategies.

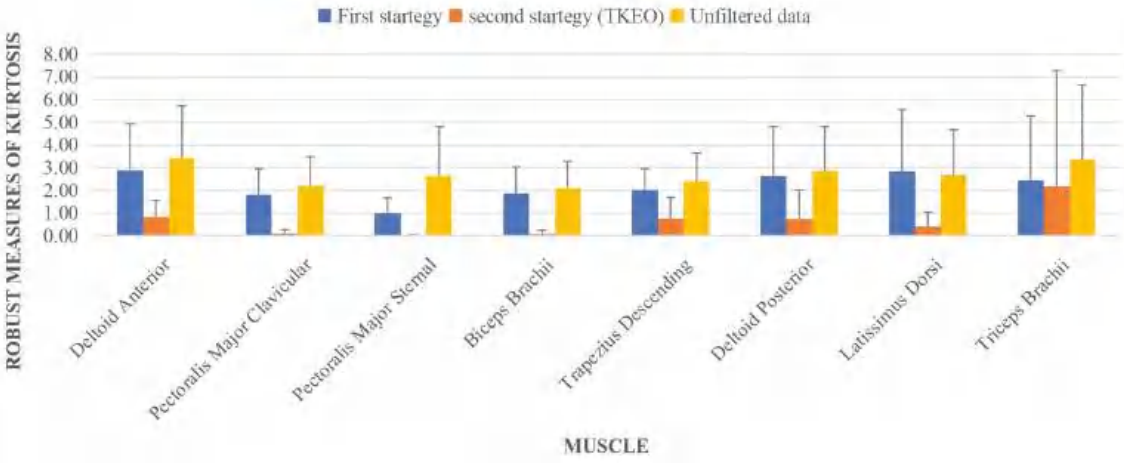


Figure 6. Kurtosis robustness values for both data smoothing techniques and raw data.

Table 2. Descriptive results for all 4 strategies.

	Muscles / Parameters	LD *	TB *	DP *	BB *	PMC *	PMS *	TD *	DA *
First strategy	Found value	82	79	77	71	72	79	67	65
	% out of found 82	100	96.3	93.9	86.6	87.8	96.3	81.7	79.3
	Mean value (s)	0.06	0.07	0.11	0.12	0.13	0.14	0.18	0.21
	SD (s)	0.06	0.11	0.16	0.16	0.12	0.16	0.15	0.18
	Consistent value	61	56	63	63	65	64	58	59
	% out of found values	74.4	70.9	81.8	88.7	90.3	81	86.6	90.8
	Range of EMG onset (s)				0.245 ± 0.199				
Second strategy	Found value	80	79	80	77	75	79	73	75
	% out of found 82	97.6	96.3	97.6	93.9	91.5	96.3	89	91.5
	Mean value (s)	0.08	0.05	0.07	0.11	0.10	0.1	0.11	0.13
	SD (s)	0.07	0.07	0.09	0.17	0.13	0.11	0.14	0.13
	Consistent value	70	53	63	64	56	64	58	63
	% out of found values	87.5	67.1	78.8	83.1	74.7	81	79.5	84
	Range of EMG onset (s)				0.274 ± 0.222				

Table 2. *Cont.*

	Parameters \ Muscles	LD *	TB *	DP *	BB *	PMC *	PMS *	TD *	DA *
Third strategy	Found value	82	82	82	82	82	82	82	82
	% out of found 82	100	100	100	100	100	100	100	100
	Mean value (s)	0.24	0.22	0.22	0.22	0.24	0.21	0.24	0.24
	SD (s)	0.12	0.15	0.14	0.13	0.14	0.13	0.13	0.13
	Consistent value	81	80	81	81	77	79	80	80
	% out of found values	98.8	97.6	98.8	98.8	93.9	96.3	97.6	97.6
	Range of EMG onset (s)				0.351 ± 0.084				
Fourth strategy	Found value	82	82	82	82	82	82	82	82
	% out of found 82	100	100	100	100	100	100	100	100
	Mean value (s)	0.21	0.2	0.21	0.2	0.23	0.24	0.25	0.21
	SD (s)	0.12	0.13	0.12	0.13	0.13	0.14	0.14	0.12
	Consistent value	78	77	79	79	78	72	75	75
	% out of found values	95.1	93.9	96.3	96.3	95.1	87.8	91.5	91.5
	Range of EMG onset (s)				0.307 ± 0.125				

* LD (Latissimus Dorsi), TB (Triceps Brachii), DP (Deltoid Posterior), BB (Biceps Brachii), PMC (Pectoralis Major Clavicular), PMS (Pectoralis Major Sternal), TD (Trapezius Descending) and DA (Deltoid Anterior).

Results of ANOVA test are shown in Table 3. Both muscle and strategy main effects were significant, although their interaction did not show any significant difference.

Table 3. ANOVA analysis result.

Independent Variable	**Estimated Time Delay**	
	F-value	*p*-value
Main Effect		
Muscle	3.79	$p < 0.05$
Strategy	62.34	$p < 0.05$
Interaction		
Muscle × *Strategy*	1.23	0.21

In the muscle main effect, the estimated time delay of Triceps Brachii was significantly different with Deltoid Anterior, Pectoralis Major Clavicular and Trapezius Descending muscles. Figure 7 presents mean and SD values of all estimated time delays for all muscles in each strategy. The estimated time delay values by each the first and second strategies are significantly different compared to the third and fourth strategies.

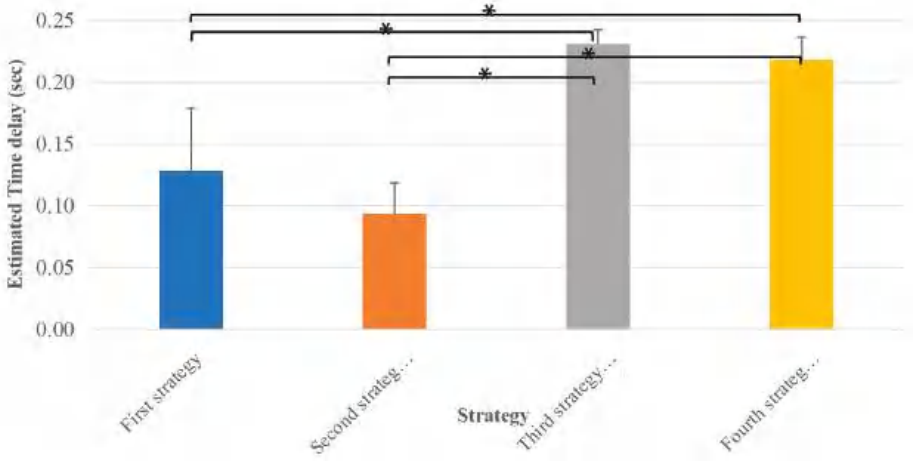


Figure 7. Mean value of all estimated muscle onsets for each strategy. The * stands for significative difference.

4. Discussion

As previously specified, the aim of this study was to estimate the muscle onset activation in SCI people by EMG data. To the best of our knowledge, no study has evaluated different EMG signal processing for muscular onset estimation during seated stability of people living with an SCI. According to our analysis, the first and second strategies estimated shorter time delays (mean = 130 ms and 90 ms, respectively) compared with the third and fourth strategies (mean = 230 ms and 220 ms, respectively) in all muscles. It can be interpreted that third and fourth strategies identify the muscle onset in the time–frequency domain and estimate it by using frequency analysis. In addition, the power spectrum showed that the signal power is less than 10 Hz, thus the movement during stability maintenance occurs in low frequency and it takes more time to reach its peak value. Furthermore, the first and second strategies failed to find the onset threshold, showing that the activity of these muscles does not change much compared to the baseline. Hence, these two strategies may not be appropriate in the estimation of time delay in only the time domain during seating stability in patients with an SCI. The EMG signal contains both ECG artifacts and measurement noise. ECG artifacts can affect the first and second strategies more than the others, because any artifacts within the signal frequency bandwidth can increase the amplitude of the measurement and can be mistakenly identified as muscle onset. The measurement noise frequency is much higher than the activation signal frequency, thus the muscle onset can be detected accurately. It seems that the motor control time delay may be identified better in the time–frequency domain compared with the time domain, which is more vulnerable to noises and artifacts. Other studies in the literature also used different multiples of the SD (1, 3 or 4) [17,47,55] to determine the muscle onset, which can change the value of the time delay. In this regard, the third and fourth strategies may be appropriate for time delay estimation with less variability in identifying the muscle onset. However, we found no study estimating the time delay in SCI people, and the results of this study are consistent with the fact that muscle onset happens earlier than torque or body angle response [33]. In addition, the results showed that the estimated time delay in SCI patients is mostly higher compared with healthy individuals [40].

Otherwise, the Kurtosis robustness analysis demonstrated that the TKEO technique could efficiently remove the ECG and motion artifacts from the EMG signal, thus resulting in an accurate muscle onset identification. On the other hand, the results have shown that the first strategy could not remove these artifacts appropriately. Artifacts can then be detected as muscle onset, leading to an erroneous reading of the data, in particular at the level of the command–contraction temporality. In addition, the ANOVA test demonstrated significant differences in each main effect of muscles and strategy on the value of estimated time delay. Each of the first and second groups were significantly different regarding the third and fourth strategies, and the value of estimated time delay in the Triceps Brachii was significantly different compared to the Deltoid Anterior, Pectoralis Major Clavicular and Trapezius Descending muscles. No Spearman correlation coefficient more than 0.5 was found for the sequence of muscle activation in each pair of different strategies. No significative Pearson correlation coefficient was found between the different methods. This can be due to employing different sequence muscle activations for each participant, resulting in different muscle synergies to compensate for the disruption achieved. There was no restriction in the arms motion so that everyone could maintain his/her stability by moving arms in sagittal or axial planes. Thus, it seems rational that no correlation was found due to motion redundancy. The range of EMG onset mean value was highest for the third strategy and lowest for the first strategy.

Several limitations should be mentioned. At first, it should be reminded that only seven persons participated in this study. A high number of repetitive trials were therefore chosen to cope with this small population. Secondly, the perturbation amplitude may change the stabilization strategy employed by the participants; for example, a high amplitude of perturbation can be detected at the cortical level where the time delay is shorter compared with response from CNS. The perturbation amplitude was not normalized in

this study as performed in [32,49], and this methodological choice was made in our study to cope with the subjects' high variability in injury level thus in stabilization performance, which can change the results. Furthermore, participants did not use specific instructions on how to stabilize their body during seated balance, which caused variability in upper limb motion. Last, each participant performed at least 11 trials, which can increase the learning effect. For future works, the value of the estimated time delay will be used in developing models of people living with SCI maintaining their sitting stability. Stability analysis will be studied using a different controller for the CNS. In addition, time delay and other passive elements of their bodies will be estimated by developing a model so that its trajectory is optimized using experimental data, which can help us to estimate more accurate values.

5. Conclusions

Two strategies in time domain and two strategies in time–frequency domain were investigated in this study for time delay estimation for people living with an SCI. The TKEO technique efficiently reduced the ECG and motion artifacts compared to the classical filtering approach. However, the first and second strategies failed to find muscle onset in several trials. Time–frequency techniques of cepstral and power spectrum estimated longer time delays due to the lower frequency of motion compared with the two other strategies in the time domain during seated balance. The time–frequency approach appears as a better option when the EMG signal includes artifacts and noises. No Spearman or Pearson correlation coefficient was found in the muscle sequence or delay value in each pair of strategies, which shows each participant used a different strategy and different sequence of muscle activation in maintaining the seated stability. The estimated time delay can help therapists and biomechanics to design more appropriate exercise and develop assistive technologies during or after rehabilitation procedures by better understanding the underlying mechanism of the body sensorimotor control system.

Author Contributions: Conceptualization, S.M.S. and M.B.; methodology, S.M.S., M.B., D.H.G. and P.P.; software, S.M.S. and M.B.; validation, T.M.G., L.W. and F.B.; formal analysis, S.M.S., M.B. and L.W.; investigation, S.M.S. and M.B.; resources, P.P., D.H.G., T.M.G. and F.B.; data curation, M.B., D.H.G. and P.P.; writing—original draft preparation, S.M.S. and M.B.; writing—review and editing, T.M.G., L.W. and F.B.; visualization, S.M.S. and M.B.; supervision, S.M.S., M.B. and L.W.; project administration, P.P., D.H.G., T.M.G. and F.B.; funding acquisition, P.P., D.H.G. and T.M.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work is supported by the French Ministry of Higher Education and Research, the National Center for Scientific Research (CNRS), the Nord-Pas-de-Calais Region, Zodiac Seats France and Direction Générale de l'Aviation Civile (project no. 2014 930181). The equipment and material required for the research completed at the Pathokinesiology Laboratory were financed by the Canada Foundation for Innovation (CFI).

Institutional Review Board Statement: Ethical approval was obtained from the Research Ethics Committee of the Centre for Interdisciplinary Research in Rehabilitation of Greater Montreal (CRIR-1083-0515R).

Informed Consent Statement: The participants read and signed the informed consent form prior to initiating the measurements.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors kindly acknowledge Ciska Molenaar from LAMIH and Martin Vermette, Youssef El Khamlichi, Philippe Gourdou and Daniel Marineau from the Pathokinesiology Laboratory for their time and expertise.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Vette, A.H.; Masani, K.; Kim, J.-Y.; Popovic, M.R. Closed-Loop Control of Functional Electrical Stimulation-Assisted Arm-Free Standing in Individuals with Spinal Cord Injury: A Feasibility Study. *Neuromodulation Technol. Neural Interface* **2009**, *12*, 22–32. [CrossRef]
2. Winter, E.M.; Brookes, F.B.C. Electromechanical Response Times and Muscle Elasticity in Men and Women. *Eur. J. Appl. Physiol. Occup. Physiol.* **1991**, *63*, 124–128. [CrossRef]
3. Chagdes, J.R.; Rietdyk, S.; Jeffrey, M.H.; Howard, N.Z.; Raman, A. Dynamic Stability of a Human Standing on a Balance Board. *J. Biomech.* **2013**, *46*, 2593–2602. [CrossRef]
4. Curuk, E.; Lee, Y.; Aruin, A.S. Individuals with Stroke Improve Anticipatory Postural Adjustments after a Single Session of Targeted Exercises. *Hum. Mov. Sci.* **2020**, *69*, 102559. [CrossRef]
5. Kanekar, N.; Aruin, A.S. Aging and balance control in response to external perturbations: Role of anticipatory and compensatory postural mechanisms. *Age* **2014**, *36*, 9621. [CrossRef]
6. Kemoun, G.; Thoumie, P.; Boisson, D.; Guieu, J. Ankle Dorsiflexion Delay Can Predict Falls in the Elderly. *J. Rehabil. Med.* **2002**, *34*, 278–283. [CrossRef]
7. Crisco, J.J.; Panjabi, M.M. Euler Stability of the Human Ligamentous Lumbar Spine. Part I: Theory. *Clin. Biomech.* **1992**, *7*, 19–26. [CrossRef]
8. Silfies, S.P.; Cholewicki, J.; Radebold, A. The Effects of Visual Input on Postural Control of the Lumbar Spine in Unstable Sitting. *Hum. Mov. Sci.* **2003**, *22*, 237–252. [CrossRef]
9. Potten, Y.J.M.; Seelen, H.A.M.; Drukker, J.; Reulen, J.P.H.; Drost, M.R. Postural Muscle Responses in the Spinal Cord Injured Persons during Forward Reaching. *Ergonomics* **1999**, *42*, 1200–1215. [CrossRef]
10. Grangeon, M.; Gagnon, D.; Gauthier, C.; Jacquemin, G.; Masani, K.; Popovic, M.R. Effects of Upper Limb Positions and Weight Support Roles on Quasi-Static Seated Postural Stability in Individuals with Spinal Cord Injury. *Gait Posture* **2012**, *36*, 572–579. [CrossRef]
11. Blandeau, M.; Guerra, T.-M.; Pudlo, P.; Gabrielli, F.; Estrada-Manzo, V. How a Person with Spinal Cord Injury Controls a Sitting Situation Unknown Input Observer and Delayed Feedback Control with Time-Varying Input Delay. In Proceedings of the 2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Vancouver, BC, Canada, 24–29 July 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 2349–2356.
12. Blandeau, M.; Estrada-Manzo, V.; Guerra, T.-M.; Pudlo, P.; Gabrielli, F. Unknown Input Observer for Understanding Sitting Control of Persons with Spine Cord Injury. *IFAC-PapersOnLine* **2016**, *49*, 175–181. [CrossRef]
13. Nguyen, A.-T.; Pan, J.; Guerra, T.-M.; Blandeau, M.; Zhang, W. Designing Fuzzy Descriptor Observer with Unmeasured Premise Variables for Head-Two-Arms-Trunk System. *IFAC-PapersOnLine* **2020**, *53*, 8007–8012. [CrossRef]
14. Guerra, T.M.; Blandeau, M.; Ngyuen, A.T.; Pan, J. Practical Approach of Input Delay Nonlinear Systems: Application to Spinal Cord Injury Sitting Stability. *IFAC-PapersOnLine* **2019**, *52*, 67–72. [CrossRef]
15. Guerra, T.-M.; Blandeau, M.; Nguyen, A.-T.; Srihi, H.; Dequidt, A. Stabilizing Unstable Biomechanical Model to Understand Sitting Stability for Persons with Spinal Cord Injury. *IFAC-PapersOnLine* **2020**, *53*, 8001–8006. [CrossRef]
16. Reeves, N.P.; Cholewicki, J.; Milner, T.E. Muscle Reflex Classification of Low-Back Pain. *J. Electromyogr. Kinesiol.* **2005**, *15*, 53–60. [CrossRef] [PubMed]
17. Radebold, A.; Cholewicki, J.; Panjabi, M.M.; Patel, T.C. Muscle Response Pattern to Sudden Trunk Loading in Healthy Individuals and in Patients with Chronic Low Back Pain. *Spine* **2000**, *25*, 947–954. [CrossRef]
18. Granata, K.P.; Slota, G.P.; Bennett, B.C. Paraspinal Muscle Reflex Dynamics. *J. Biomech.* **2004**, *37*, 241–247. [CrossRef]
19. Cholewicki, J.; Silfies, S.P.; Shah, R.A.; Greene, H.S.; Reeves, N.P.; Alvi, K.; Goldberg, B. Delayed Trunk Muscle Reflex Responses Increase the Risk of Low Back Injuries. *Spine* **2005**, *30*, 2614–2620. [CrossRef]
20. Pereira, S.; Silva, C.C.; Ferreira, S.; Silva, C.; Oliveira, N.; Santos, R.; Vilas-Boas, J.P.; Correia, M.V. Anticipatory Postural Adjustments during Sitting Reach Movement in Post-Stroke Subjects. *J. Electromyogr. Kinesiol.* **2014**, *24*, 165–171. [CrossRef]
21. Aruin, A.S.; Kanekar, N.; Lee, Y.-J.; Ganesan, M. Enhancement of Anticipatory Postural Adjustments in Older Adults as a Result of a Single Session of Ball Throwing Exercise. *Exp. Brain Res.* **2015**, *233*, 649–655. [CrossRef]
22. Redfern, M.S.; Müller, M.L.; Jennings, J.R.; Furman, J.M. Attentional Dynamics in Postural Control during Perturbations in Young and Older Adults. *J. Gerontol. Ser. A Biol. Sci. Med. Sci.* **2002**, *57*, B298–B303. [CrossRef]
23. Reeves, N.P.; Luis, A.; Chan, E.C.; Sal, Y.R.V.G.; Tanaka, M.L. Assessing Delay and Lag in Sagittal Trunk Control Using a Tracking Task. *J. Biomech.* **2018**, *73*, 33–39. [CrossRef]
24. Brown, L.A.; Jensen, J.L.; Korff, T.; Woollacott, M.H. The Translating Platform Paradigm: Perturbation Displacement Waveform Alters the Postural Response. *Gait Posture* **2001**, *14*, 256–263. [CrossRef]
25. Leinonen, V.; Kankaanpää, M.; Luukkonen, M.; Hänninen, O.; Airaksinen, O.; Taimela, S. Disc Herniation-Related Back Pain Impairs Feed-Forward Control of Paraspinal Muscles. *Spine* **2001**, *26*, E367–E372. [CrossRef]
26. Radebold, A.; Cholewicki, J.; Polzhofer, G.K.; Greene, H.S. Impaired Postural Control of the Lumbar Spine Is Associated with Delayed Muscle Response Times in Patients with Chronic Idiopathic Low Back Pain. *Spine* **2001**, *26*, 724–730. [CrossRef] [PubMed]
27. Adkin, A.L.; Frank, J.S.; Carpenter, M.G.; Peysar, G.W. Fear of Falling Modifies Anticipatory Postural Control. *Exp. Brain Res.* **2002**, *143*, 160–170. [CrossRef]

28. Mochizuki, G.; Sibley, K.M.; Esposito, J.G.; Camilleri, J.M.; McIlroy, W.E. Cortical Responses Associated with the Preparation and Reaction to Full-Body Perturbations to Upright Stability. *Clin. Neurophysiol.* **2008**, *119*, 1626–1637. [CrossRef]
29. Borghuis, A.J.; Lemmink, K.A.; Hof, A.L. Core Muscle Response Times and Postural Reactions in Soccer Players and Nonplayers. *Med. Sci. Sport. Exerc.* **2011**, *43*, 108–114. [CrossRef]
30. Kanekar, N.; Aruin, A.S. The Effect of Aging on Anticipatory Postural Control. *Exp. Brain Res.* **2014**, *232*, 1127–1136. [CrossRef]
31. Blenkinsop, G.M.; Pain, M.T.; Hiley, M.J. Evaluating Feedback Time Delay during Perturbed and Unperturbed Balance in Handstand. *Hum. Mov. Sci.* **2016**, *48*, 112–120. [CrossRef]
32. Claudino, R.; Dos Santos, M.J.; Mazo, G.Z. Delayed Compensatory Postural Adjustments After Lateral Perturbations Contribute to the Reduced Ability of Older Adults to Control Body Balance. *Mot. Control* **2017**, *21*, 425–442. [CrossRef]
33. Mohebbi, A.; Amiri, P.; Kearney, R.E. Identification of Human Balance Control Responses to Visual Inputs Using Virtual Reality. *J. Neurophysiol.* **2022**, *127*, 1159–1170. [CrossRef] [PubMed]
34. Maurer, C.; Peterka, R.J. A New Interpretation of Spontaneous Sway Measures Based on a Simple Model of Human Postural Control. *J. Neurophysiol.* **2005**, *93*, 189–200. [CrossRef] [PubMed]
35. Welch, T.D.; Ting, L.H. A Feedback Model Reproduces Muscle Activity during Human Postural Responses to Support-Surface Translations. *J. Neurophysiol.* **2008**, *99*, 1032–1038. [CrossRef]
36. Sovol, A.W.; Valles, K.D.B.; Riedel, S.A.; Harris, G.F. Bi-Planar Postural Stability Model: Fitting Model Parameters to Patient Data Automatically. In Proceedings of the 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology, Buenos Aires, Argentina, 31 August 2010–4 September 2010; IEEE: Piscataway, NJ, USA, 2010; pp. 3962–3965.
37. van Drunen, P.; Maaswinkel, E.; van der Helm, F.C.; van Dieen, J.H.; Happee, R. Identifying Intrinsic and Reflexive Contributions to Low-Back Stabilization. *J. Biomech.* **2013**, *46*, 1440–1446. [CrossRef]
38. Valles, K.D.B.; Udoekwere, U.I.; Long, J.T.; Schneider, J.M.; Riedel, S.A.; Harris, G.F. A Bidirectional Model of Postural Sway Using Force Plate Data. *Crit. Rev. Biomed. Eng.* **2014**, *42*, 451–466. [CrossRef]
39. Yin, K.; Chen, J.; Xiang, K.; Pang, M.; Tang, B.; Li, J.; Yang, L. Artificial Human Balance Control by Calf Muscle Activation Modelling. *IEEE Access* **2020**, *8*, 86732–86744. [CrossRef]
40. Wang, H.; van den Bogert, A.J. Identification of Postural Controllers in Human Standing Balance. *J. Biomech. Eng.* **2021**, *143*, 041001. [CrossRef] [PubMed]
41. McKee, K.L.; Neale, M.C. Direct Estimation of the Parameters of a Delayed, Intermittent Activation Feedback Model of Postural Sway during Quiet Standing. *PLoS ONE* **2019**, *14*, e0222664. [CrossRef]
42. Nagy, D.J.; Bencsik, L.; Insperger, T. Experimental Estimation of Tactile Reaction Delay during Stick Balancing Using Cepstral Analysis. *Mech. Syst. Signal Process.* **2020**, *138*, 106554. [CrossRef]
43. Peterka, R.J.; Loughlin, P.J. Dynamic Regulation of Sensorimotor Integration in Human Postural Control. *J. Neurophysiol.* **2004**, *91*, 410–423. [CrossRef] [PubMed]
44. Basmajian, J.V. Electromyography—Dynamic Gross Anatomy: A Review. *Am. J. Anat.* **1980**, *159*, 245–260. [CrossRef] [PubMed]
45. Grin, L.; Frank, J.; Allum, J.H. The Effect of Voluntary Arm Abduction on Balance Recovery Following Multidirectional Stance Perturbations. *Exp. Brain Res.* **2007**, *178*, 62–78. [CrossRef] [PubMed]
46. Tokuno, C.D.; Carpenter, M.G.; Thorstensson, A.; Cresswell, A.G. The Influence of Natural Body Sway on Neuromuscular Responses to an Unpredictable Surface Translation. *Exp. Brain Res.* **2006**, *174*, 19–28. [CrossRef] [PubMed]
47. Weaver, T.B.; Hamilton, L.E.; Tokuno, C.D. Age-Related Changes in the Control of Perturbation-Evoked and Voluntary Arm Movements. *Clin. Neurophysiol.* **2012**, *123*, 2025–2033. [CrossRef] [PubMed]
48. Solnik, S.; Rider, P.; Steinweg, K.; DeVita, P.; Hortobágyi, T. Teager-Kaiser Energy Operator Signal Conditioning Improves EMG Onset Detection. *Eur. J. Appl. Physiol.* **2010**, *110*, 489–498. [CrossRef] [PubMed]
49. Santos, M.J.; Kanekar, N.; Aruin, A.S. The Role of Anticipatory Postural Adjustments in Compensatory Control of Posture: 1. Electromyographic Analysis. *J. Electromyogr. Kinesiol.* **2010**, *20*, 388–397. [CrossRef]
50. Crow, E.L.; Siddiqui, M.M. Robust Estimation of Location. *J. Am. Stat. Assoc.* **1967**, *62*, 353–389. [CrossRef]
51. Thongpanja, S.; Phinyomark, A.; Quaine, F.; Laurillau, Y.; Limsakul, C.; Phukpattaranont, P. Probability Density Functions of Stationary Surface EMG Signals in Noisy Environments. *IEEE Trans. Instrum. Meas.* **2016**, *65*, 1547–1557. [CrossRef]
52. Xu, L.; Peri, E.; Vullings, R.; Rabotti, C.; Van Dijk, J.P.; Mischi, M. Comparative Review of the Algorithms for Removal of Electrocardiographic Interference from Trunk Electromyography. *Sensors* **2020**, *20*, 4890. [CrossRef]
53. Parzen, E. On Estimation of a Probability Density Function and Mode. *Ann. Math. Stat.* **1962**, *33*, 1065–1076. [CrossRef]
54. Vasudevan, J.M.; Logan, A.; Shultz, R.; Koval, J.J.; Roh, E.Y.; Fredericson, M. Comparison of Muscle Onset Activation Sequences between a Golf or Tennis Swing and Common Training Exercises Using Surface Electromyography: A Pilot Study. *J. Sport. Med.* **2016**, *2016*, 3987486. [CrossRef] [PubMed]
55. Liebetrau, A.; Puta, C.; Anders, C.; de Lussanet, M.H.; Wagner, H. Influence of Delayed Muscle Reflexes on Spinal Stability: Model-Based Predictions Allow Alternative Interpretations of Experimental Data. *Hum. Mov. Sci.* **2013**, *32*, 954–970. [CrossRef] [PubMed]

sensors

MDPI

Article

A Denoising Method for Mining Cable PD Signal Based on Genetic Algorithm Optimization of VMD and Wavelet Threshold

Yanwen Wang [1], Peng Chen [1,*], Yongmei Zhao [2] and Yanying Sun [1]

[1] School of Mechanical, Electronic & Information Engineering, China University of Mining and Technology-Beijing, Beijing 100083, China
[2] CHN Energy Technology & Economics Research Institute Co., Ltd., Beijing 100083, China
* Correspondence: chenpeng@student.cumtb.edu.cn

Abstract: When the pulse current method is used for partial discharge (PD) monitoring of mining cables, the detected PD signals are seriously disturbed by the field noise, which are easily submerged in the noise and cannot be extracted. In order to realize the effective separation of the PD signal and the interference signal of the mining cable and improve the signal-to-noise ratio of the PD signal, a denoising method for the PD signal of the mining cable based on genetic algorithm optimization of variational mode decomposition (VMD) and wavelet threshold is proposed in this paper. Firstly, the genetic algorithm is used to optimize the VMD, and the optimal value of the number of modal components K and the quadratic penalty factor α is determined; secondly, the PD signal is decomposed by the VMD algorithm to obtain K intrinsic mode functions (IMF). Then, wavelet threshold denoising is applied to each IMF, and the denoised IMFs are reconstructed. Finally, the feasibility of the denoising method proposed in this paper is verified by simulation and experiment.

Keywords: PD denoising; VMD; wavelet threshold; genetic algorithm; mining cables

Citation: Wang, Y.; Chen, P.; Zhao, Y.; Sun, Y. A Denoising Method for Mining Cable PD Signal Based on Genetic Algorithm Optimization of VMD and Wavelet Threshold. *Sensors* **2022**, *22*, 9386. https://doi.org/10.3390/s22239386

Academic Editor: Manuel José Cabral dos Santos Reis

Received: 17 October 2022
Accepted: 28 November 2022
Published: 1 December 2022

1. Introduction

The online monitoring of partial discharge (PD) is considered an effective method for checking cable insulation defects and identifying potential faults. It has been widely used in the condition monitoring of power cables [1–3]. At present, the ultra-high frequency (UHF) method [4] and pulse current method [5] are commonly used for PD measurement. The center frequency of UHF is 500 MHz. According to different measurement methods, its bandwidth is more than ten MHz or even several GHz. The measurement frequency of the pulse current method is relatively low, usually a few kHz to a few hundred kHz. However, the UHF signal attenuation is serious or even cannot be measured during the transmission of the PD signal in the cable, so the effect of the UHF method in the PD monitoring of cables is not ideal. We believe that the pulse current method is more suitable for the PD monitoring of cables. The pulse current method is mainly subject to noise interference, and it is difficult to separate from the interference signal. To ensure that the PD signal is not distorted as much as possible and to improve the signal-to-noise ratio (SNR) of the PD signal, it is necessary to conduct more in-depth research on the denoising algorithm of the cable PD signal.

The PD signal has the characteristics of nonlinear, time series non-equilibrium, and wide frequency band distribution, so it is not easy to effectively denoise the PD signal by selecting its frequency band. Currently, the empirical mode decomposition (EMD) method and wavelet threshold method are the main denoising methods for PD [6–8]. The EMD method recursively detects the local maximum and minimum values in the signal, which is highly dependent on the extreme point search method. According to the energy rule, the intrinsic mode function (IMF) with a small order is directly discarded, resulting in the loss of some valuable signals. The wavelet threshold method can realize the signal

localization in both time and frequency domains simultaneously, which has good time-frequency analysis capability. However, the transient process of local projection will be lost due to the wavelet transform.

In order to solve the shortcomings of EMD and the wavelet threshold method in PD signal denoising, many scholars have made many improvements to these two methods. A partial discharge-based novel adaptive ensemble empirical mode decomposition (Novel Adaptive EEMD, NAEEMD) method is proposed by Tao Jin [9] for noise reduction. After using EEMD to decompose the PD signal, the method adaptively selects the intrinsic mode function for noise reduction reconstruction. Jeffery C. Chan [10] proposes a self-adaptive technique for partial discharge (PD) signal denoising with automatic threshold determination based on EEMD and mathematical morphology. On the basis of mathematical morphology, an automatic morphological thresholding (AMT) technique is developed to form upper and lower thresholds for automatically eliminating the residual noise while maintaining the PD signals. Ramy Hussein [11] proposes a wavelet-based denoising method with a new histogram-based threshold function and selection rule. The proposed threshold estimation technique obtains two different threshold values for each wavelet sub-band and uses a prodigious thresholding function that conserves the original signal energy. Jun Zhong [12] proposes a method without human intervention in choosing the threshold parameters or the decomposition layer numbers.

In 2014, Konstantin Dragomiretskiy proposed Variational Mode Decomposition (VMD) based on EMD [13]. Compared with EMD, VMD has fewer decomposition layers and rigorous mathematical theory, which improves its robustness against noise interference [14–16]. The VMD method can retain the transient PD process relatively completely, but its ability to suppress noise is weak. In conclusion, for the PD signal, which is more seriously affected by the field interference signal, the filtering effects of the above methods are less satisfactory when used alone.

After analyzing the advantages of VMD and wavelet threshold, a combined denoising method based on VMD and wavelet threshold is proposed. This method combines the advantages of VMD's ability to adaptively adjust the center frequency of each mode and the excellent time-frequency analysis ability of wavelet threshold method while avoiding the disadvantages of VMD's weak noise suppression ability and the transient process of wavelet threshold loss. At the same time, we use genetic algorithm (GA) to optimize the number of modal components and quadratic penalty factor in VMD and determine the two input parameters that can achieve the optimal decomposition of VMD. The structure of this paper is as follows: Section 2 introduces the basic principles of VMD, GA and wavelet threshold method; Section 3 describes the specific process of the proposed denoising algorithm and the process of optimizing the parameters by GA; Section 4 shows the simulation verification results and the comparison of the denoising effect between the proposed method and several different methods. In Section 5, the experimental signal denoising results are described in detail. The conclusion of this paper is given in Section 6.

2. Methods and Principles

2.1. VMD Method

VMD can adaptively and non-recursively decompose the input signal into multiple IMFs with specific sparse properties. Each IMF has a corresponding center frequency. During the decomposition process, each mode is continuously evaluated to optimize the distribution of each IMF and its center frequency.

The VMD algorithm is actually a solution process for a variational problem. It decomposes the original signal into K IMFs $u_k(t)$, so that the sum of the estimated bandwidths of

each IMF is minimized, and then, the corresponding constrained variation model can be expressed as:

$$\begin{aligned} &\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_{k=1}^{K} \left\| \partial_t [[(\delta(t) + \tfrac{j}{\pi t}) * u_k(t)] e^{-j\omega_k t}] \right\|_2^2 \right\} \\ &s.t. \sum_{k=1}^{K} u_k = f \end{aligned} \tag{1}$$

where $\{u_k\}$ represents the set of all IMFs, $\{\omega_k\}$ represents the set of center frequencies corresponding to the IMF, $(\delta(t) + \frac{j}{\pi t}) * u_k(t)$ is the unilateral frequency spectrum of each eigenmode, obtained by computing its analytic signal through the Hilbert transform, and f is the original signal.

To transform the above constrained variational problem into an unconstrained variational problem, the quadratic penalty factor α and the Lagrangian multiplier $\lambda(t)$ are introduced, and the extended Lagrangian expression is:

$$L(\{u_k\},\{\omega_k\},\lambda) := \alpha \sum_{k=1}^{K} \left\| \partial_t [[(\delta(t) + \tfrac{j}{\pi t}) * u_k(t)] e^{-j\omega_k t}] \right\|_2^2 + \left\| f(t) - \sum_{k=1}^{K} u_k(t) \right\|_2^2 + \left\langle \lambda(t), f(t) - \sum_{k=1}^{K} u_k(t) \right\rangle \tag{2}$$

The alternate direction method of multipliers (ADMM) is used to solve the variational problem, and the optimal solution of the above function is obtained by iteratively updating ${u_k}^{n+1}$, ${\omega_k}^{n+1}$ and λ^{n+1}. The value problem of ${u_k}^{n+1}$ can be expressed as:

$$u_k^{n+1} = \underset{u_k \in X}{\operatorname{argmin}} \{ \alpha \left\| \partial_t [[(\delta(t) + \tfrac{j}{\pi t}) * u_k(t)] e^{-j\omega_k t}] \right\|_2^2 + \left\| f(t) - \sum_{i=1, i\neq k}^{K} u_i(t) + \frac{\lambda(t)}{2} \right\|_2^2 \} \tag{3}$$

where ω_k and $u_{i\neq k}$ represent the latest available update value. Using the Parseval Fourier equidistant transform, the above equation can be transformed as:

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i=1, i\neq k}^{K} \hat{u}_i(\omega) + \frac{\overset{\wedge}{\lambda}(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2} \tag{4}$$

According to the same solution process as u_k, the solution of the quadratic optimization problem of the center frequency is shown in Formula (5):

$$\omega_k^{n+1} = \frac{\int_0^\infty \omega \left| \hat{u}_k(\omega) \right|^2 \mathrm{d}\omega}{\int_0^\infty \left| \hat{u}_k(\omega) \right|^2 \mathrm{d}\omega} \tag{5}$$

where $\hat{u}_k^{n+1}(\omega)$ is equivalent to the Wiener filtering result of the current residual $\hat{f}(\omega) - \sum_{i=1, i\neq k}^{K} \hat{u}_i(\omega)$; and ω_k^{n+1} is the center of gravity of the power spectrum of the modal function.

The VMD algorithm is continuously updated in the frequency domain, and then, the inverse Fourier transform is performed to obtain the time domain result. The specific process of the VMD algorithm can be described as follows:

Step 1: Given the number of modal decompositions K and the penalty factor α, initialize $\{\hat{u}_k^1\}$, $\{\overset{\wedge}{\omega}_k^1\}$, $\{\overset{\wedge}{\lambda}^1\}$ and n;

Step 2: Update u_k and ω_k in the frequency domain according to Formulas (4) and (5);

Step 3: Update λ, and its update formula is as follows:

$$\overset{\wedge}{\lambda}^{n+1}(\omega) = \overset{\wedge}{\lambda}^{n}(\omega) + \tau(\hat{f}(\omega) - \sum_{K} \hat{u}_k^{n+1}(\omega)) \tag{6}$$

Step 4: Given the discrimination accuracy ε, when it is satisfied $\frac{\sum_k \left\| \hat{u}_k^{n+1} - \hat{u}_k^n \right\|_2^2}{\left\| u_k^n \right\|_2^2} < \varepsilon$, stop the iteration and output the IMF $\{u_k\}$.

2.2. Genetic Algorithm

The genetic algorithm is an optimization algorithm that simulates the natural selection and genetic evolution of organisms [17–19]. It usually includes three genetic operators: selection, crossover and mutation. The genetic algorithm is an iterative process; each cycle is a generation. In the operation, the inheritance is terminated after a specified number of generations, and then, the optimal chromosome is found among all generations. The genetic algorithm optimization is performed again if the optimal solution is not found. When using the genetic algorithm to solve the optimization problem, it mainly needs to go through six steps: encoding, initial population generation, fitness value evaluation, selection, crossover, and mutation, so that the population evolves into a new generation of a better adaptive population. The specific process of the genetic algorithm is shown in Figure 1.

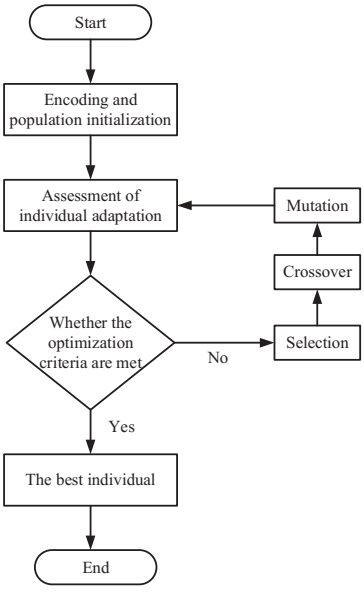


Figure 1. Genetic algorithm flow.

2.3. Wavelet Threshold Denoising

The basic idea of the wavelet threshold denoising method is that a noisy signal can be expressed as the superposition of the original signal and the noise obeying the Gaussian distribution [20,21]. Since the wavelet transform is linear, the wavelet coefficients of the original signal and the noise can be obtained, respectively, after the noisy signal undergoes discrete wavelet transform. Based on the fact that the useful signal and the noise have different statistical properties after the wavelet transform, the original signal's wavelet coefficients are larger and more significant than the noise's. Therefore, an appropriate threshold λ is found as a criterion for judging whether the decomposed signal is discarded or not. When the decomposition coefficient is less than the threshold λ, it is considered that the decomposition coefficient is mainly caused by noise, and the corresponding decomposition signal should be discarded; when the decomposition coefficient is greater than the threshold λ, it is considered that the decomposition coefficient is mainly caused

by the signal, and the corresponding decomposition signal is processed. Then, wavelet reconstruction is performed to obtain the denoised signal. The specific process of wavelet threshold denoising is as follows:

Step 1: Select the appropriate wavelet basis function and decomposition level and perform wavelet decomposition on the noisy signal.

Step 2: Select an appropriate threshold to properly process the wavelet coefficients. When the decomposed wavelet coefficients are smaller than the selected threshold, it is considered that the wavelet coefficients are mainly caused by noise and should be set to zero. When the wavelet coefficients are greater than the selected threshold, it is believed that the wavelet coefficients are mainly due to the signal.

Step 3: Perform inverse wavelet transform on the processed wavelet coefficients to obtain a denoising result.

Wavelet transform is a new transform analysis method, which transforms the function $f(t)$ under the wavelet basis, and its expression is:

$$WT_f(a,\tau) = [f(t), \psi_{a,\tau}(t)] = \frac{1}{\sqrt{a}} \int_R f(t)\psi^*(\frac{t-\tau}{a})dt \tag{7}$$

where $\psi(t)$ is the wavelet basis function; a is the expansion and contraction amount; and τ is the translation amount. It can be seen from the above formula that the wavelet transform is actually the integral transform of the function, $WT_f(a,\tau)$ represents the wavelet coefficient after the wavelet, and the expression of the inverse transform can be expressed as:

$$f(t) = \frac{1}{c_\phi} \int_0^{+\infty} \frac{da}{a^2} \int_{-\infty}^{+\infty} WT_f(u,\tau) \frac{1}{\sqrt{a}} \psi(\frac{t-\tau}{a})dt \tag{8}$$

When thresholding the wavelet coefficients, there are usually two methods: hard thresholding and soft thresholding. Hard thresholding is to keep larger coefficients and zero out smaller coefficients, as shown in Formula (9):

$$\hat{W}_{j,k} = \begin{cases} W_{j,k} & \left|W_{j,k}\right| \geq Thr \\ 0 & \left|W_{j,k}\right| < Thr \end{cases} \tag{9}$$

Soft thresholding is to set the smaller wavelet coefficients to zero and the larger coefficients to shrink toward zero, as shown in Formula (10):

$$\hat{W}_{j,k} = \begin{cases} \mathrm{sgn}(W_{j,k}) * (\left|W_{j,k}\right| - Thr) & \left|W_{j,k}\right| \geq Thr \\ 0 & \left|W_{j,k}\right| < Thr \end{cases} \tag{10}$$

where $W_{j,k}$ represents the wavelet coefficient; Thr represents the threshold.

3. PD Signal Denoising Based on Genetic Algorithm Optimization of VMD and Wavelet Threshold

3.1. PD Signal Denoising Process

The PD signal has a wide frequency band, and the main frequency is not apparent. It can be seen from the above basic principles that the VMD method can adaptively decompose the PD signal into multiple eigenmodes with center frequencies. It can not only extract the PD signal from the interference but also preserve the transient process of the PD signal as much as possible. According to the basic principles of VMD, genetic algorithm and wavelet threshold, we propose a PD denoising method based on the genetic algorithm optimization of VMD and wavelet threshold. Firstly, the two input parameters of VMD are optimized by genetic algorithm, and the optimal parameter value that can make VMD achieve the best decomposition effect are obtained. Then, K IMFs are decomposed by

VMD; that is, $\{u_1, u_2, \ldots, u_k\}$, wavelet threshold denoising is applied to each IMF to obtain the denoised components of each IMF. Finally, we perform signal reconstruction on the denoised components of all IMFs. The specific denoising process is shown in Figure 2.

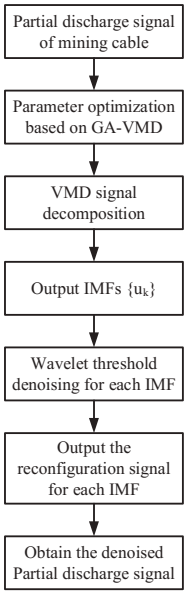


Figure 2. Denoising flow chart of PD signal.

3.2. Optimization of VMD Parameters Based on Genetic Algorithm

From the theory of VMD, it is known that VMD needs to pre-set the number of modal components K when decomposing the signal. The VMD decomposition results obviously differ with different settings of the number of modal components. It has been found that the quadratic penalty factor α in the VMD method also has a large impact on the VMD decomposition results. However, the number of modal components K and the quadratic penalty factor α need to be manually set in advance, and the randomness and uncertainty of the artificial setting will inevitably affect the correctness of the VMD decomposition result. How to choose the appropriate two input parameters is the premise and key to accurately decompose the signal by VMD.

Since genetic algorithm is a direct search optimization method generated by evolution theory and genetic mechanism, it has good global probability search ability. Therefore, this paper uses genetic algorithm to optimize the two input parameters K and α of VMD and obtain the optimal value. Input parameters. When the genetic algorithm searches for the input parameters of the VMD method, an adaptation function needs to be defined in step 3. The information entropy can well evaluate the sparse characteristic of the signal. The size of the information entropy reflects the uncertainty of the signal. The larger the entropy value, the greater the uncertainty of the signal. The entropy value of e_j (e_j is the signal sequence after demodulation and decomposition) is the envelope entropy, which can reflect the sparse characteristic of the original signal. The envelope entropy E_e of the zero mean signal $x(j)$ ($j = 1, 2, \ldots, N$) can be expressed as:

$$\begin{cases} E_e = -\sum_{j=1}^{N} e_j \lg e_j \\ e_j = a(j) / \sum_{j=1}^{N} a(j) \end{cases} \tag{11}$$

where e_j is the normalized form of $a(j)$; and $a(j)$ is the envelope signal of signal $x(j)$ after Hilbert transform.

In order to search for the global optimal u_k component combination, we take the local minimum envelope entropy value as the fitness value in the whole parameter optimization process, and we take the minimization of the local minimum envelope entropy value as the final parameter optimization goal.

4. Simulation Verification and Analysis

4.1. Simulation Results

In the PD monitoring site, various random noise disturbances will be generated in electrical systems such as analog circuits, photoelectric conversion, analog/digital conversion, and communication lines, which are expressed in the form of Gaussian white noise signals. The characteristic of white Gaussian noise is that its amplitude is Gaussian distribution N (0, 1), and its power spectral density is uniformly distributed. In order to simulate the pulse signal obtained in the field, the original PD pulse signal is superimposed with Gaussian white noise to obtain a noisy signal. The one-dimensional signal model with noise can be expressed as: $d_i = f_i + \varepsilon z_i$, $i = 1, 2, \ldots, N$. Among them, d_i is the noisy signal, f_i is the "pure" PD pulse signal, z_i is Gaussian white noise, ε is the noise level, and N is the signal length. The PD signal is weak, the signals measured on-site has many noise components, and the SNR of the signal is very low. Taking a noisy signal with a signal-to-noise ratio of −2.67 dB as an example, its waveform and spectrum are shown in Figure 3. It can be seen that the PD signal is submerged in noise, which has a wide frequency range and is randomly distributed in the frequency band of the PD signal. The spectral characteristics of the PD signal in the noisy signal are not prominent.

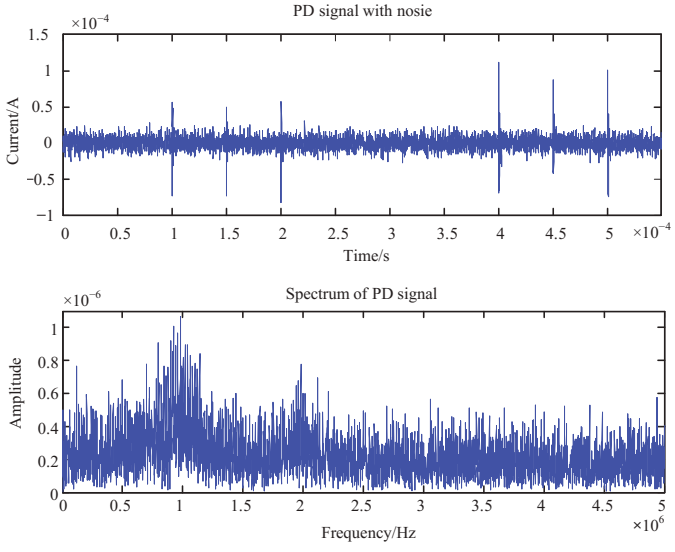


Figure 3. The waveform of the PD signal with noise and its spectrum.

Figure 4 shows the results of the genetic algorithm for the optimal search of VMD input parameters, which reflects the plot of the local minimal envelope entropy values of the simulated signals at different genetic generations. The minimum value of the local minimal envelope entropy value 0.1307 appears in the 6th generation, and the optimal input parameter $(K, \alpha) = (5, 847)$ is obtained by the search. Therefore, the number of modal components K in the VMD method is set to 5, and the quadratic penalty factor α is 847 to decompose the simulated signal.

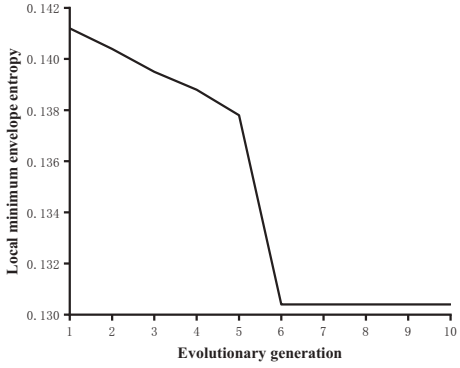


Figure 4. Local minimal envelope entropy values of the simulated signal at different genetic generations.

Figure 5 shows the decomposed eigenmodal components and their corresponding spectra when K is taken as 5. It can be seen that the peak of the spectrum of mode u_2 coincides with the peak of the spectrum of the noise-containing signal, and the peak of the spectrum of mode u_3 coincides with the peak of the noise-containing signal near 2 MHz. This shows that the VMD decomposition of the noise-bearing signal is the best at this time.

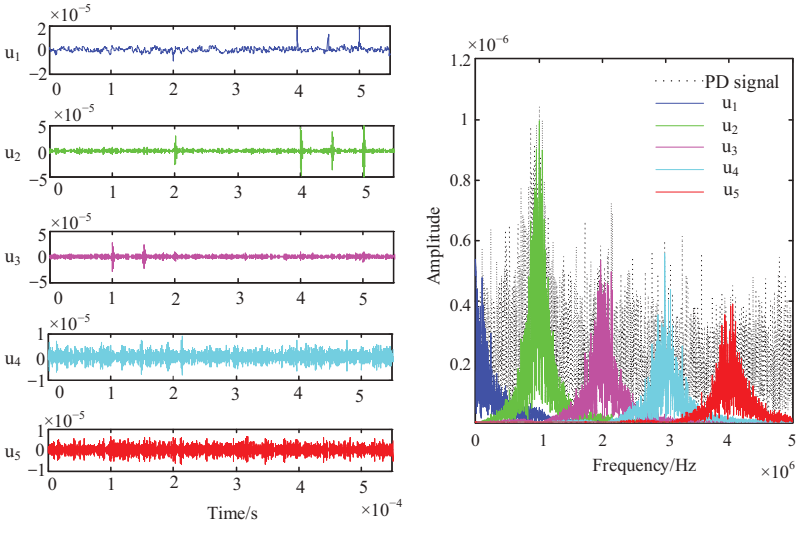


Figure 5. The decomposed eigenmodal components and their corresponding spectra, when $K = 5$.

After determining the values of K and α, we perform VMD on the noisy signal. Each modal component u_k obtained after decomposition still contains obvious noise interference, so it is necessary to perform wavelet threshold denoising on each mode component u_k separately to achieve better denoising effect. After our careful analysis, the db.4 wavelet is selected as the wavelet base in this paper, the decomposition scale is three layers, and the threshold value is calculated using the fixed threshold estimation method. In order to ensure the smoothness of the signal, a soft threshold function is selected for processing. As shown in Figure 6, the waveforms of each eigenmode u_k are shown on the left, and the corresponding waveforms c_k obtained after wavelet thresholding for each eigenmode are shown on the right.

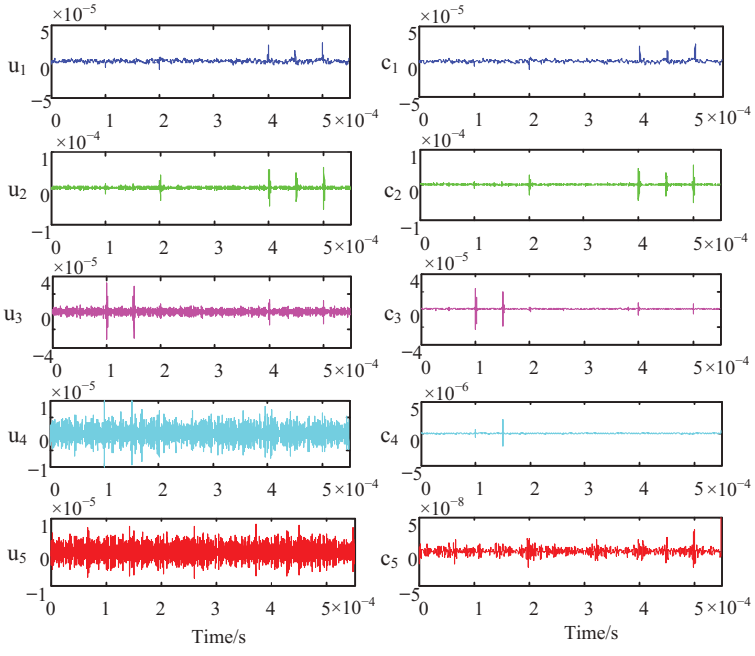


Figure 6. Intrinsic mode functions and wavelet threshold denoising signals.

It can be seen from Figure 6 that noise is significantly suppressed in the reconstructed signal ck obtained by the wavelet decomposition of each eigenmode component u_k. In particular, the PD components in u_4 and u_5 are completely submerged in the noise, and after the wavelet threshold decomposition, the PD components in c_4 and c_5 are obvious. VMD is characterized by the ability to decompose a broadband signal into a signal consisting of multiple narrowbands. Therefore, some scholars have achieved the separation of low-frequency mixed signals or low-frequency noise-laden signals by the selective rounding of IMFs through VMD. However, the characteristic of the PD signal is that its frequency band is extremely wide and it is difficult to find a fixed dominant frequency, so the abandonment of IMFs will lead to the loss of useful signal information. Through the secondary processing of wavelet threshold denoising, the noise interference is effectively removed, and the useful signals in each IMFs are preserved to a great extent.

4.2. Comparison of Different Methods

The signal c_k processed by wavelet thresholding is reconstructed to obtain the reconstructed signal based on VMD and wavelet threshold. In order to compare the denoising ability of the PD denoising method proposed in this paper, the method is compared with several current mainstream PD denoising methods. Figure 7 shows the reconstructed signal waveform of the noisy PD signal processed by methods such as VMD, wavelet threshold, EMD autocorrelation, and VMD and wavelet threshold. Among them, the EMD autocorrelation method is a denoising method improved by the EMD method, and its denoising effect is better than that of the EMD method. It can be seen that the reconstruction method based on VMD and wavelet threshold is better than the other three methods, and the transient part of the PD signal is well preserved. The denoising effect of the signal processed by the wavelet threshold is also significant, but the transient part of the PD signal is missing more seriously.

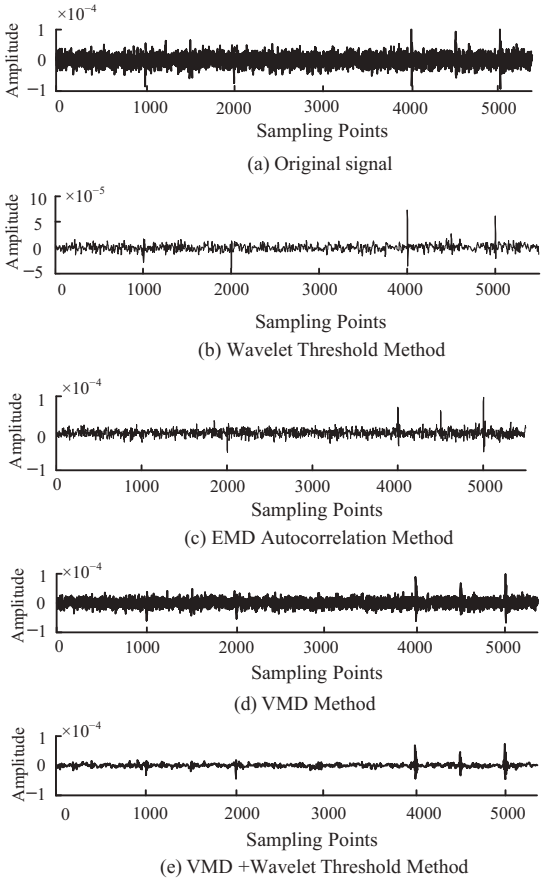


Figure 7. Comparison of four methods of denoising effect (original SNR is −2.673 dB).

Although the denoising ability of various methods can be visualized from Figure 7, in order to make further quantitative comparisons, we denoise the PD signals with different SNR values. The correlation coefficient R and root mean square error RMSE are also introduced as the basis for judging the denoising ability. The detailed calculation formulas of SNR, R and RMSE are shown in Formulas (12)–(14).

$$\mathrm{SNR} = 10\log 10\left(\sum_{i=1}^{N} {x_i}^2 / \sum_{i=1}^{N} \left(x_i - x_i'\right)^2\right) \tag{12}$$

$$\mathrm{R} = \frac{\sum_{i=1}^{N} \left(x_i - \overline{x_i}\right)\left(x_i' - \overline{x_i'}\right)}{\sqrt{\sum_{i=1}^{N} \left(x_i - \overline{x_i}\right)^2 \cdot \sum_{i=1}^{N} \left(x_i' - \overline{x_i'}\right)^2}} \tag{13}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \left(x_i - x_i'\right)^2} \tag{14}$$

As can be seen from Table 1, the denoising effect of VMD and the wavelet threshold method is significantly better than that of the VMD method, EMD autocorrelation method and wavelet threshold method, especially for signals with smaller SNR values.

Table 1. Comparison of four ways to suppress the noise effect.

PD Signal SNR	VMD			VMD + Wavelet Threshold Method			EMD Autocorrelation Method			Wavelet Threshold Method		
	SNR	R	RMSE/$\times 10^{-6}$	SNR	R	RMSE/$\times 10^{-6}$	SNR	R	RMSE/$\times 10^{-6}$	SNR	R	RMSE/$\times 10^{-6}$
−7	−4.29	0.47	9.66	2.79	0.72	4.28	−1.99	0.45	7.41	−0.01	0.50	5.91
−3	0.10	0.64	5.83	5.60	0.86	3.10	1.54	0.62	4.94	2.93	0.71	4.21
0.1	2.43	0.75	4.46	7.01	0.90	2.63	3.70	0.77	3.85	5.15	0.83	3.26
1	3.21	0.78	4.08	7.71	0.92	2.43	4.19	0.79	3.64	5.76	0.86	3.04
4	5.97	0.87	2.97	8.95	0.95	2.11	6.78	0.89	2.70	8.56	0.93	2.20
9	9.33	0.94	2.01	10.6	0.97	1.74	9.78	0.95	1.91	12.06	0.97	1.47

5. Experimental Signal Analysis

In order to further verify the denoising effect of the denoising method proposed in this paper on the measured signal, 2.5 kV DC voltage is applied to the 6 kV power cable in the laboratory, the PD voltage signal is measured by the detection impedance method, and the signal data obtained from the experiment are processed by Matlab. Figure 8a shows the waveform of the measured PD signal in the laboratory, and it can be seen that the PD signal is almost drowned in the noise interference. Figure 8b–e are the signals after denoising the experimental signals using the VMD method, wavelet threshold method, EMD autocorrelation method and the method proposed in this paper, respectively.

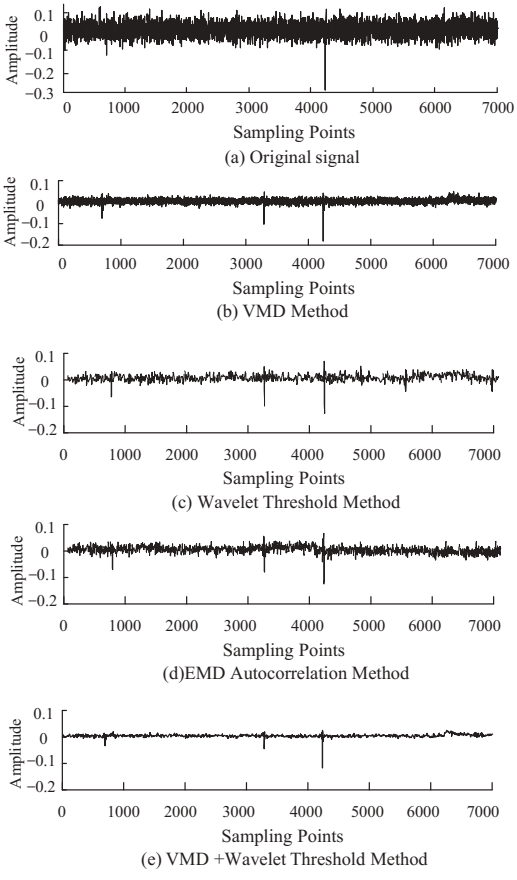


Figure 8. Analysis of PD signal obtained in laboratory.

It can be seen from Figure 8 that the denoising effect of the VMD and wavelet threshold reconstruction method is the best, the PD signal after denoising has no oscillation phenomenon, and the noise is significantly suppressed. Although the wavelet threshold method also plays an obvious role in suppressing the noise, it also causes a large amount of loss in the transient process of the PD signal, which is not conducive to the analysis of the PD signal later. The EMD autocorrelation method and VMD method have limited effects on noise suppression, especially the single VMD method has no obvious effect on noise suppression, and there is still a large number of noise signals. Since the calculation of SNR, R and RMSE requires the original "pure" signal, and the original "pure" signal of the experimental signal is unknown, the Noise Rejection Ratio (NRR) before and after signal denoising is introduced here to measure the denoising effect. NRR characterizes the prominence of the effective signal after denoising. Table 2 shows the NRR calculation results of the two methods.

$$\mathrm{NRR} = 10\log 10({\sigma_1}^2 - {\sigma_2}^2) \tag{15}$$

where ${\sigma_1}^2$ and ${\sigma_2}^2$ represent the variance of the signal before and after denoising, respectively.

Table 2. Comparison of NRR calculation results.

Denoising Method	NRR/dB
VMD method	3.7901
EMD Autocorrelation	4.2705
Wavelet Threshold	4.6294
VMD + Wavelet Threshold	5.3603

From the calculation results of the NRR of each method in Table 2, it can be seen more intuitively that the NRR of the VMD and wavelet threshold method is the highest, and the NRR of the single VMD method is the lowest. Through the analysis of the NRR of each method, the analysis of the denoising effect of each method in this paper can be supported from another perspective. Combining the results of Figure 8 and Table 2 further proves that the denoising effect of the method proposed in this paper is significantly better than the other three methods.

6. Conclusions

The pulse current method is an effective means for monitoring the PD of mining cables. However, when collecting PD signals, due to the influence of on-site working conditions, the noise interference is large, and the cable PD signals cannot be effectively extracted. In order to improve the SNR of the PD signal, this paper proposes a denoising method of mining a cable PD signal based on the optimization of VMD and wavelet threshold. Considering that the number of modal components K and the quadratic penalty factor α have a great influence on the results of VMD, this paper introduces a genetic algorithm to determine the optimal values of these two parameters, so that VMD can achieve the best results. Meanwhile, according to the respective characteristics of VMD and the wavelet threshold method, these two methods are effectively combined to further improve the denoising effect of a cable local discharge signal. The main conclusions are as follows.

(1) In this paper, by introducing the genetic algorithm, the minimization of the local minimal envelope entropy value is taken as the optimization goal of VMD parameters. Then, the optimal values of VMD parameters are obtained, which avoids the situation that the VMD denoising ability is insufficient due to the artificial setting of the parameter value.
(2) The combined denoising algorithm proposed in this paper combines the advantages of VMD's ability to adaptively adjust the center frequency of each mode and the

excellent time-frequency analysis capability of wavelet threshold. It also avoids the disadvantages of VMD's weak noise suppression capability and the loss of transient processes by wavelet threshold. Through the experimental comparison, it is found that the method has a more excellent denoising ability, and the filtering performance is better for PD signals with lower SNR.

Author Contributions: Conceptualization, Y.W. and P.C.; methodology, P.C.; simulation, P.C.; validation, Y.S. and Y.Z.; writing—original draft preparation, P.C.; writing—review and editing, Y.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Van Jaarsveldt, H.; Gouws, R. Condition monitoring of medium voltage electrical cables by means of partial discharge measurements. *S. Afr. Inst. Electr. Eng.* **2014**, *105*, 136–145. [CrossRef]
2. Rosle, N.; Muhamad, N.A.; Rohani, M.N.K.H.; Jamil, M.K.M. Partial Discharges Classification Methods in XLPE Cable: A Review. *IEEE Access* **2021**, *9*, 133258–133273. [CrossRef]
3. Lu, L.; Zhou, K.; Zhu, G.; Chen, B.; Yana, X. Partial Discharge Signal Denoising with Recursive Continuous S-Shaped Algorithm in Cables. *IEEE Trans. Dielectr. Electr. Insul.* **2021**, *28*, 1802–1809. [CrossRef]
4. Tenbohlen, S.; Denissov, D.; Hoek, S.; Markalous, S.M. Partial discharge measurement in the ultra high frequency (UHF) range. *IEEE Trans. Dielectr. Electr. Insul.* **2008**, *15*, 1544–1552. [CrossRef]
5. Hu, X.; Siew, W.H.; Judd, M.D.; Reid, A.J.; Sheng, B. Modeling of High-Frequency Current Transformer Based Partial Discharge Detection in High-Voltage Cables. *IEEE Trans. Power Deliv.* **2019**, *34*, 1549–1556. [CrossRef]
6. Kopsinis, Y.; McLaughlin, S. Development of EMD-Based Denoising Methods Inspired by Wavelet Thresholding. *IEEE Trans. Signal Process.* **2009**, *57*, 1351–1362. [CrossRef]
7. Boudraa, A.; Cexus, J. EMD-Based Signal Filtering. *IEEE Trans. Instrum. Meas.* **2007**, *56*, 2196–2202. [CrossRef]
8. Huimin, C.; Ruimei, Z.; Yanli, H. Improved Threshold Denoising Method Based on Wavelet Transform. *Phys. Procedia* **2012**, *33*, 1354–1359. [CrossRef]
9. Jin, T.; Li, Q.; Mohamed, M.A. A Novel Adaptive EEMD Method for Switchgear Partial Discharge Signal Denoising. *IEEE Access* **2019**, *7*, 58139–58147. [CrossRef]
10. Jeffery, C.C.; Hui, M.; Tapan, K.; Chandima, E. Self-adaptive partial discharge signal de-noising based on ensemble empirical mode decomposition and automatic morphological thresholding. *IEEE Trans. Dielectr. Electr. Insul.* **2014**, *21*, 294–303.
11. Ramy, H.; Khaled, B.S.; Ayman, H.E. Wavelet Transform with Histogram-Based Threshold Estimation for Online Partial Discharge Signal Denoising. *IEEE Trans. Instrum. Meas.* **2015**, *64*, 3601–3614.
12. Zhong, J.; Bi, X.; Shu, Q.; Zhang, D.; Li, X. An Improved Wavelet Spectrum Segmentation Algorithm Based on Spectral Kurtogram for Denoising Partial Discharge Signals. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–8. [CrossRef]
13. Dragomiretskiy, K.; Zosso, D. Variational Mode Decomposition. *IEEE Trans. Signal Process.* **2014**, *62*, 531–544. [CrossRef]
14. Wang, Q.; Wang, L.; Yu, H.; Wang, D.; Nandi, A.K. Utilizing SVD and VMD for Denoising Non-Stationary Signals of Roller Bearings. *Sensors* **2022**, *22*, 195. [CrossRef] [PubMed]
15. Tang, J.; Zhou, S.; Pan, C. A Denoising Algorithm for Partial Discharge Measurement Based on the Combination of Wavelet Threshold and Total Variation Theory. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 3428–3441. [CrossRef]
16. Long, J.; Wang, X.; Dai, D.; Tian, M.; Zhu, G.; Zhang, J. Denoising of UHF PD signals based on optimised VMD and wavelet transform. *IET Sci. Meas. Technol.* **2017**, *11*, 753–760. [CrossRef]
17. Katoch, S.; Chauhan, S.S.; Kumar, V. A review on genetic algorithm: Past, present, and future. *Multimed. Tools Appl.* **2021**, *80*, 8091–8126. [CrossRef]
18. Ding, S.; Su, C.; Yu, J. An optimizing BP neural network algorithm based on genetic algorithm. *Artif. Intell. Rev.* **2011**, *36*, 153–162. [CrossRef]
19. Langazane, S.N.; Saha, A.K. Effects of Particle Swarm Optimization and Genetic Algorithm Control Parameters on Overcurrent Relay Selectivity and Speed. *IEEE Access* **2022**, *10*, 4550–4567. [CrossRef]
20. He, C.; Xing, J.; Li, J.; Yang, Q.; Wang, R. A New Wavelet Threshold Determination Method Considering Interscale Correlation in Signal Denoising. *Math. Probl. Eng.* **2015**, *2015*, 280251. [CrossRef]
21. Lu, Y.-J.; Lin, H.; Dong, Y.; Zhang, Y.-S. A New Wavelet Threshold Function and Denoising Application. *Math. Probl. Eng.* **2016**, *2016*, 3195492.

sensors

MDPI

Article

Analysis of Physiological Responses during Pain Induction

Raquel Sebastião [1,*], Ana Bento [2] and Susana Brás [1]

[1] IEETA, DETI, LASI, University of Aveiro, 3810-193 Aveiro, Portugal
[2] DFis, University of Aveiro, 3810-193 Aveiro, Portugal
* Correspondence: raquel.sebastiao@ua.pt

Abstract: Pain is a complex phenomenon that arises from the interaction of multiple neuroanatomic and neurochemical systems with several cognitive and affective processes. Nowadays, the assessment of pain intensity still relies on the use of self-reports. However, recent research has shown a connection between the perception of pain and exacerbated stress response in the Autonomic Nervous System. As a result, there has been an increasing analysis of the use of autonomic reactivity with the objective to assess pain. In the present study, the methods include pre-processing, feature extraction, and feature analysis. For the purpose of understanding and characterizing physiological responses of pain, different physiological signals were, simultaneously, recorded while a pain-inducing protocol was performed. The obtained results, for the electrocardiogram (ECG), showed a statistically significant increase in the heart rate, during the painful period compared to non-painful periods. Additionally, heart rate variability features demonstrated a decrease in the Parasympathetic Nervous System influence. The features from the electromyogram (EMG) showed an increase in power and contraction force of the muscle during the pain induction task. Lastly, the electrodermal activity (EDA) showed an adjustment of the sudomotor activity, implying an increase in the Sympathetic Nervous System activity during the experience of pain.

Keywords: autonomic nervous system (ANS); cold pressor task (CPT); pain induction; pain assessment; physiological signals; signal processing; signal analysis

Citation: Sebastião, R.; Bento, A.; Brás, S. Analysis of Physiological Responses during Pain Induction. *Sensors* **2022**, *22*, 9276. https://doi.org/10.3390/s22239276

Academic Editor: Juan Pablo Martínez

Received: 15 October 2022
Accepted: 25 November 2022
Published: 29 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Pain is a complex biopsychosocial phenomenon caused by damage or potential damage in the tissues and serves a vital protective function. The International Association for the Study of Pain revised the definition of pain as: "an unpleasant sensory and emotional experience associated with, or resembling that associated with, actual or potential tissue damage" [1]. This revision also specified that (a): pain is always a personal experience that is influenced by several factors; (b) pain cannot be inferred solely from activity in the sensory neurons; (c) the concept of pain is learnt throughout an individual life experience; (d) a person's report of an experience as pain should be respected; (e) pain serves an adaptive role but can also cause effects on individual well-being; and (f) the inability to communicate does not exclude an individual's capacity to feel pain [1].

In almost every clinical practice, especially in neurological and musculoskeletal problems [2], pain is a common symptom and an accurate assessment is critical to ensure a safe and effective management of pain. Currently, the most used standard method to assess pain is based on self-reports, both in clinical and experimental settings. Despite subjective, self-reports being generally easy to obtain, requiring practically little to no equipment, allowing for comprehensive information collection, and exhibiting typically good reliability [2], these instruments rely on the ability of the individual to process external information and communicate a response, which may not always be feasible. Moreover, the use of self-reports from the patient to assess pain may be hazarded by the age, cognitive condition, and verbal communication capabilities of the patient. In such cases, pain needs to be assessed by a healthcare provider, which is a more complex and time-consuming

task, and can be highly challenging due to individual differences in pain expression and behavior and physiological changes not always being specific to pain.

It is thought that pain exacerbates the autonomic response to stress, a rationale supported by evidence showing a neuroanatomical overlap between nociceptive and autonomic pathways [3]. For example, studies have shown that the application of pain stimuli induces significant heart rate acceleration. Therefore, there has been a growing interest in the use of autonomic reactivity as an objective marker of pain, and several studies have investigated physiologic variables for this purpose under pain-induced conditions [2,4–6].

The Cold Pressor Task (CPT) is a pain-inducing method that requires individuals to immerse one hand (or forearm) in cold water for as long as they can tolerate it or during a fixed period of time. The main advantages of this method rely on its portability, minimal training to use, and few risks. The primary disadvantage of the CPT is the significant methodological divergences in its implementation and in the measurement of pain outcomes, crippling the comparison of results from different studies [4,5,7–9]. There is increasing information linking the feeling of pain with the Autonomic Nervous System (ANS). Therefore, several studies have investigated and recorded the alterations of the ANS with the use of CPT.

Goals and Organization

As pain is mostly assessed through the use of subjective instruments, such as self-reports relying on a pain-scale, the main goal of the present study was to study and characterize physiological responses when experiencing pain. For that, a pain-inducing protocol was implemented on forty-five healthy-volunteer participants, and several physiological signals (electrocardiogram—ECG; two electromyograms—EMG, electrodermal activity—EDA) were recorded, while, simultaneously, pain was induced through the exposure to cold stimuli. Thus, with this protocol, we analyze the physiological responses of pain and assess the pain perception through self-reports based on a numerical rating scale (NRS). This work is organized as follows: Section 2 presents related works with respect to the ANS reactions associated with induced pain. Section 3 presents the study protocol, data collection, and methodology for data analysis, while Section 4 presents the obtained results. Finally, before presenting the conclusions of our study and identifying further research in Section 6, Section 5 discusses the obtained results regarding the physiological characterization of pain.

2. Background

There is increasing information linking the feeling of pain with the ANS. Therefore, several studies have investigated and recorded the alterations of the ANS when participants are subjected to pain-inducing stimuli.

Concerning cold stimuli to induce pain, the work [4] quantifies the changes in skin impedance, Heart Rate (HR), and facial skin temperature when healthy volunteers were subjected to acute pain through a CPT (with water at 0 °C). A total of 19 participants were included in the study. The results showed an increase in all the parameters calculated during the CPT, in comparison to those calculated during the baseline. However, only the skin conductance increase was statistically significant. One possible justification for the minor variation of HR during both conditions (baseline and CPT) may stem from anxiety felt by some participants before the pain-inducing task.

Ref. [5] analyzed the relation between efferent sympathetic nervous system activity to skeletal muscle (MNSA) and pain sensation during localized skin cooling. Ten subjects took part in the study, immersing their right hand in different temperature water baths for three minutes each. The levels of temperature in the bath range from warm (28 °C and 21 °C—non-pain inducing) and mid-level (14 °C) to cold (7 °C and 0 °C—pain-inducing). The participants went in order from the warmest to coldest temperature, with a ten-minute interval between the recovery three-minute period of the last water tank and the three-minute baseline of the next. While the study was being performed, the MNSA, Blood

Pressure (BP), HR, and breathing were continuously recorded. The observations of this study demonstrated that there was no evident influence on MNSA when the participants were subjected to non-painful skin cooling. During the hand immersion in ice water, there was a progressive rise in MSNA as skin temperature started to decrease. However, there was a more significant peak increase in the MSNA signal during the 0 °C immersion compared to the 7 °C. Regarding HR, there was a significant rise during the initial phase of the 0 °C, which was expected. Even so, the HR consistently increased in less painful water temperatures, although on a smaller scale. As for the BP, there were no significant changes during the study.

Aiming at studying the relationship between HR and multidimensional aspects of pain (intensity and unpleasantness) in healthy individuals, 39 healthy volunteers were subjected to hot water (47 °C) hand immersion test for two minutes, while ECG and EDA were being recorded [10]. Participants also had to rate their perceived pain every 15 s using a Visual Analogue Scale (VAS). The HR, Heart Rate Variability (HRV) parameters, and Skin Conductance Level (SCL) were calculated. The study showed a steady rise in pain intensity and unpleasantness, as well as in HR, with the progression of immersion time. These results seemed to indicate a rise in sympathetic activity and a drop in parasympathetic activity, which is in agreement with the usual body reaction to a noxious stimulus. Regarding pain perception and HR, there seems to exist a greater correlation between HR and pain unpleasantness than with pain intensity, suggesting that pain-related autonomic responses are functionally related to the affective dimension of the experience. However, the correlation between pain perception and HR indicated a vast difference between genders, with men presenting much greater values.

In addition, through the use of hot stimulation, Ref. [11] assessed if there was a relation between sudomotor activity and heat pain perception. To that end, a thermal stimulus protocol, using a Peltier type contact thermode, was applied to 22 healthy participants while the EDA was being recorded. The participants also reported their subjective perception of pain through VAS. During the procedure, the baseline temperature of the thermode started at 31.5 °C. Three different types of stimuli were tested on all the participants on three different days. The results indicated a positive correlation between changes in sudomotor activity and pain perception as the mean EDA level and sympathetic skin response were higher in the pain phases. This was especially verified in quicker temperature slopes. Since both features are considered reliable indices of emotion, it is conceivable that the increase in sudomotor activity is also related to an emotional component. After the end of the pain-inducing protocol, the mean EDA decreased, indicating a drop in sympathetic outflow when an event responsible for an emotional response is over.

Ref. [2] studied the alterations in the HR, skin conductance, and VAS ratings in response to noxious stimulus created by calibrated heat stimulus of different intensities, which range from warm to pain-inducing. The data were analyzed from two different perspectives: the correlation between the autonomic response and pain intensity in subjects separately (subject analysis) and the correlation between the average pain intensity and the autonomic responses to the same temperature in all individuals (group analysis). The results demonstrated that an increase in pain intensity generated an increase in both HR and skin conductance. The subject analysis revealed a higher correlation with skin conductance, leading to a belief that this metric is more sensitive to changes in perception. However, the magnitude increases of the skin conductance did not significantly correlate with the magnitude of pain intensity, suggesting that this measure alone does not predict the absolute level of pain reported by the subject. The opposite was true for HR, as it did not reliably predict verbal responses to pain on a subject basis but did on the group level. These differences suggest that, although HR is affected by pain perception, it is a very noisy measure.

With a protocol of several tests to assess the autonomic function and considering patients with chronic neck or shoulder pain and control participants, Ref. [12] studied the differences in responses of muscle blood flow, muscle activity, HRV, and BP. The protocol

consisted of an initial 15-min rest period and three different tests with a 5-min rest period in between: the hand grip test (HGT), the cold pressor task (CPT), and the deep breathing test (DBT). During the rest period, patients showed lower parasympathetic activity compared to the control group. Blood flow in the trapezius muscle during HGT and CPT was also lower in patients than in the control group. This result may be the consequence of increased sympathetic activity leading to a change in blood flow due to an imbalance between vasoconstriction and dilation in the affected muscle. Finally, it was observed that trapezius muscle activity in patients was highest during the rest period after static contraction, which seems to indicate an inability on the part of patients to relax properly after static work.

Ref. [13] evaluated the changes in the ANS in patients with fibromyalgia through CPT. A total of 38 women participated in this study, of which 23 were patients with fibromyalgia. At the beginning and end of CPT, the pain was assessed with a numerical scale, and a thermographic recording of the forearm was measured. The physiological measurements considered included BP and pulse rate. It was observed that participants with fibromyalgia had a lower resistance to the stimulus of cold water. These observations may thus be related to the abnormal functioning of the ANS and, therefore, abnormal perception of pain and/or suffering from ischemia more rapidly.

Considering 13 physiological parameters derived from the HR, breath rate (BR), galvanic skin response (GSR), and facial surface electromyogram, the authors of [14] proposed artificial neural network classifiers to distinguish between no, mild, and moderate/severe acute pain. A group of 30 healthy volunteers was subjected to thermal and electrical pain stimulation, and pain was self-reported using VAS. The results show that HR, GSR, and BR were better correlated to pain intensity variations than facial muscle activities. The authors also concluded that the use of multiple physiological parameters for pain classification was revealed to be advantageous, especially in the classification of mild pain category since data from this category overlapped greatly with the other two categories.

Ref. [15] goes beyond the analysis of machine learning recognition models for pain assessment based on physiological and behavioral data. It also proposes a framework for feature extraction methods that allows a fair comparison of the performances of feature extraction and feature learning approaches. The authors concluded that simple feature engineering approaches, relying on features extracted from the signals based on expert knowledge, lead to better performances than deep learning approaches and that more complex deep learning architectures do not necessarily outperform simpler ones. However, although comparing five different approaches evaluated on two databases, the major drawback of this work relies on the use of the EDA signal only. Thus, further research should be endeavored by including other physiological data and by considering data fusion approaches to increase the performance of the pain classification models.

3. Materials and Methods

This section describes the protocol for data collection and presents the methods applied for analyzing the body response during the induction of pain through cold pain stimuli implemented as a CPT. The different methods used to analyze the data were implemented in Matlab R2021a (MATLAB R2021a and Simulink R2021a) [16].

3.1. Data Collection

Aiming to study the physiological changes that pain provokes, 45 participants were subjected to a pain-inducing protocol (CPT), while, simultaneously, physiological signals, namely ECG, EMG, and EDA, were being collected. This study was approved by the Ethics and Deontological Council of the University of Aveiro (number 09-CED/2019).

All the participants were recruited from the local community, they were healthy, did not suffer from any disease that causes chronic pain, did not present any mental illness or neurological disorder, and, lastly, could comprehend and answer to self-report measures. As explained before, we studied a total of 45 participants, 27 male and 18 female, with ages between 21 and 59 (33 ± 11 years old).

To perform the CPT, two specially designed tanks were used. These were produced to be able to sustain the water at the desired temperature. The physiological data were collected with the Biosignalsplux® Explorer tool kit, with a sampling frequency of 1000 Hz. A total of four sensors were used to record the signals, two EMG sensors, one ECG sensor, and one EDA sensor. The ECG was collected with a triode configuration: two electrodes were placed on the right and left side of the participant's ribcage, and a reference electrode was placed above the pelvic bone (as shown in Figure 1A). The EMG sensors, with a bipolar configuration, were placed in the trapezius and triceps muscles of the non-dominant arm (as observed in Figure 1B). Since there was no built-in reference electrode, one, serving for both EMG signals, was placed in the clavicle (Figure 1A). The EDA sensor, which also had a bipolar configuration, was collected on the dominant hand (as indicated in Figure 1C). Additionally, to mark the different epochs, a handheld switch directly connected to the hub was used. After collection, the raw ECG, EDA, and EMG signals were converted to microvolts (mV), according to the information provided in the Biosignalsplux sensor datasheets (https://support.pluxbiosignals.com/wp-content/uploads/2021/10/biosignalsplux-Electrocardiography-ECG-Datasheet.pdf; https://support.pluxbiosignals.com/wp-content/uploads/2021/10/biosignalsplux-Electromyography-EMG-Datasheet.pdf; https://support.pluxbiosignals.com/wp-content/uploads/2021/11/Electrodermal_Activity_EDA_Datasheet.pdf). Lastly, the BP was measured at three different moments during the study, with the resource of an upper arm blood pressure monitor that was placed on the bare upper dominant arm of the participant. Participants also had to self-report their level of pain at different moments, using a 0–10 level NRS. With zero score standing for no pain, one to three scores for light pain, four to six scores for moderate pain, seven to nine scores for severe pain, and, finally, a 10 score for the worst pain imaginable.

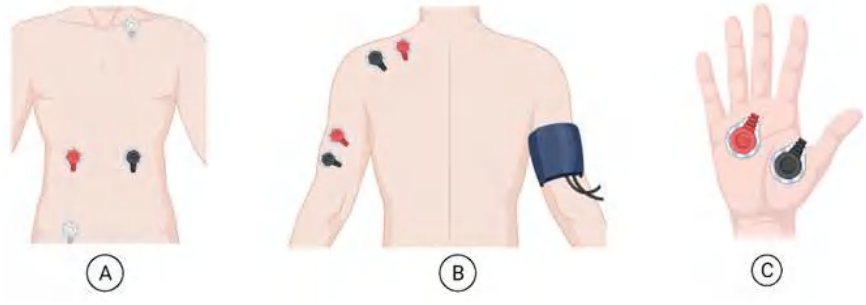


Figure 1. Illustration of the placement sites for the different electrodes (**A**) ECG electrodes plus reference electrode of the EMG; (**B**) EMG electrodes on the trapezius and triceps muscle plus BP monitor; and (**C**) EDA electrodes.

All the information regarding the study was given to the participants, and the respective informed consent was obtained. At the beginning of the procedure, the participants had to respond to the instrument for data collection regarding their age, gender, and health status, thus ensuring that they complied with the inclusion criteria. That same data collection sheet was later used to fill out their pain level.

The protocol started with a five-minute baseline recording, where the participant had to be seated, at a comfortable position, with their arm close to their body, trying to avoid movements. Afterwards, participants were asked to immerse the non-dominant hand and forearm inside the warm water tank (with temperature 37 °C ± 1 °C) for two minutes, to ensure that all the participants started the CPT with similar skin temperatures. Before the end of this task, the level of pain, with an NRS, was assessed. Afterward, for the induction of pain, the participants immersed the arm into the cold-water tank (with temperature 7 °C ± 1 °C) and the CPT started. If the participant was unable to withstand the CPT for the whole two minutes, they could withdraw their hand from the

cold tank. In this case, the participant was advised to notify their wish to remove the arm from the tank and, before doing so, to report their current pain level and the level of the maximum pain experienced during the CPT. If the participant was able to withstand the entire CPT, the current and maximum pain levels were reported at the two-minute mark. Right after removing the arm, the participants reported again the pain level, and the BP was measured. The participant transferred the arm back to the warm water tank for two minutes of immersion. Next, the hand and forearm were dried, and, while seated in a comfortable position, a five-minute rest period, similar to the initial baseline, commenced. At the three-minute point during this rest period (around five minutes after the end of the CPT), they were asked to give their current pain level and to report the maximum level of pain they felt in retrospect.

The scheme of the implemented protocol is shown in Figure 2.

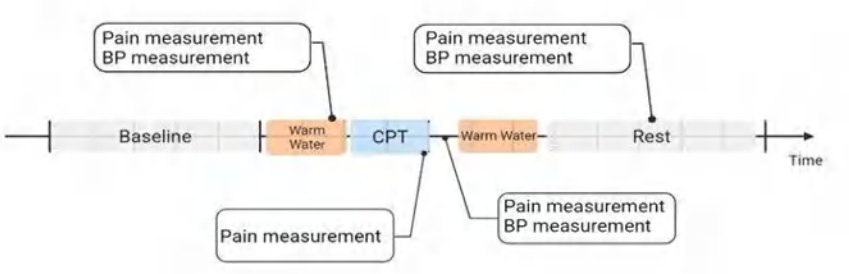


Figure 2. Representation of the different steps of the experimental procedure, with further indication of the moments where pain was self-reported and the BP measured. ECG, EDA, and EMG were continuously recorded throughout the study.

3.2. Data Analysis

After acquiring the raw ECG, EMG, and EDA signals, they had to be pre-processed. The ECG was filtered considering the frequencies between 0.5 Hz and 40 Hz. The EMG was filtered to remove the interference of the powerline, and high-passed at 10 Hz. Lastly, the EDA signals were low-pass filtered at 10 Hz.

After the pre-processing step, the data were normalized according to the baseline epoch, which corresponds to the first five minutes of this study. The feature extraction was performed using the Neurokit2 (https://neuropsychology.github.io/NeuroKit/) in Python. After the data were processed, it was divided into epochs according to the pressing of the triggers. The five epochs created are the five-minute baseline recording (Baseline), the first two-minute recordings of the hand and arm in the warm water tank (WarmWater1), the CPT recording, the two-minute recordings of the warm water tank for the second time (WarmWater2), and, finally, the last five minute rest (Rest).

Afterward, statistical analysis was performed to investigate differences in the extracted features in several epochs. As all of the features failed to be normally distributed, the differences between the five different epochs were evaluated with the non-parametric Friedmann. When a significant difference was found between the five epochs, the Wilcoxon signed-rank test, with Bonferroni correction, was performed to evaluate which epochs were significantly different from each other.

4. Results

Six of the 45 original volunteers had to be taken out of the study, as the participants did not fulfill all the protocol. As such, a total of 39 individuals were used in this study.

The felt pain was assessed through self-report at four moments. On the first pain evaluation, at the end of the WarmWater1, no participant reported pain (NRS = 0). On the second assessment, at the end of the CPT, the average value for the pain of the participants reported using the 0–10 level NRS, at that exact moment, was 6.85 ± 2.23.

After removing the arm from the water, the level of pain decreased around a 1.5 score, with participants reporting a mean of 5.37 $\pm$ 2.55. At the final assessment, participants reported their current level of pain and tried to recall their maximum. For the current level of pain, only three participants reported a low level of pain (NRS = 1). As with respect to the recall of the maximum level of pain, participants reported 7.37 $\pm$ 2.19. In general, women reported higher levels of pain when compared to men.

4.1. ECG Processing and Analysis

Regarding ECG features, the HR was computed and the maximum and minimum values of each ECG cycle were calculated, R peaks and S peaks, respectively. Afterward, for each epoch, the averages of those were computed.

With respect to the HRV features, and due to the different lengths of the epochs and the short term of the CPT epoch, only the following features were considered: RMSSD, pNN50 (time-domain features), and SampEn (nonlinear feature). The description of the used features is presented in Table 1.

Table 1. Description of the extracted ECG features.

Feature	Description
Mean HR	Number of beats per minute (mean)
R peaks	Maximum value of the ECG cycles (upward deflections)
S peaks	Minimum value of the ECG cycles (downward deflections)
RMSSD	Root Mean Square of Successive Differences between normal heartbeats. It is a reflection of the beat-to-beat difference in the HR and is used to estimate the alteration of the HRV caused by the vagus nerve.
pNN50	Percentage of successive RR intervals that are greater than 50 ms, associated with the Parasympathetic Nervous System (PNS) activity
SampEn	Measures the regularity and complexity of a given signal. Smaller values indicate a regular and predictable signal

Figure 3 represents the results for the normalized mean HR. It is clear that the most prominent boxplot is the CPT, being the epoch with the higher HR values, showing a response to the stress caused by the pain. Observing the matrix statistical results, there is a statistically significant difference between the mean HR during the CPT and from the remaining epochs. From the obtained results, there appears to be no difference between the Baseline and the WarmWater1, and between the Baseline and the Rest. However, the same was not verified with respect to the Baseline and the second tank of warm water, which may be the reflex of the pain induced during the CPT.

Figure 4 regards the maximum value of the ECG cycles, which correspond to the R-peaks, showing a median value increase for the normalized R-peak amplitude from the Baseline to the WarmWater1, with little variation of the dispersion. This increase is about 7.7% from the Baseline to the WarmWater1. However, this is followed by a decrease of 2.55% during the CPT. The median, rises, once again, reaching its peak with an increase of about 6% during the WarmWater2. The amplitude returns to near its original value during the Rest period. Although slight, there seems to be a reaction when the participants placed their hands on the water. However, there is no significant difference between the non-pain-inducing and pain-inducing water temperatures on the maximum amplitude of the ECG cycles. The statistical analysis corroborates this, as it did not show any inter-epochs significant differences, with the exception of the WarmWater2 for the Baseline and the Rest, the epochs with the highest and lowest amplitude values, respectively. These results suggest that the maximum ECG amplitude is not a suitable feature to examine the presence of pain in an individual when subjected to CPT.

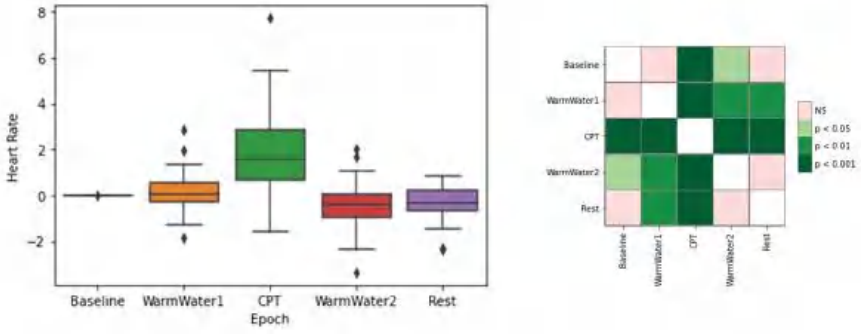


Figure 3. Boxplot of mean HR values for each epoch (**left**) and respective *p*-values between different epochs, with Bonferroni correction (**right**). The ◇ stands for outliers.

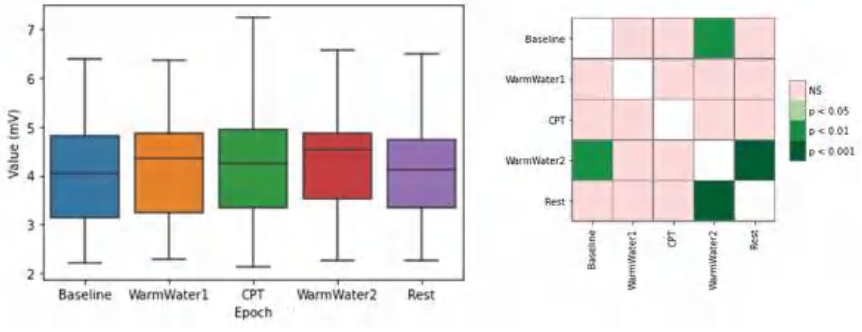


Figure 4. Boxplot for the mean maximum ECG cycle values (**left**) and respective matrix of calculated *p*-values between different epochs, with Bonferroni correction, for the (**right**).

Another ECG feature studied was the minimum value of the ECG cycles, which corresponds to the S-peak. Figure 5 shows a decrease of the median value from the Baseline to the WarmWater1, followed by a decrease from this epoch to the CPT. After the pain-inducing procedure, the minimum amplitude of the ECG cycles gradually increased. The statistical analysis for this feature shows that the CPT had significant differences from all the other epochs, being statistically more significant with the Baseline and Rest periods. There were no significant differences between the Baseline and Rest. Finally, regarding WarmWater1 and WarmWater2, there was, also, no significant difference between them. Nevertheless, both had statistically significant differences from the other groups.

Figure 6 shows the RMSSD results. Looking at the graph, the epoch with the lowest values is the CPT. As for the other epochs, the RMSSD values are higher. However, the Baseline and, especially, the WarmWater1 appear, in general, to have slightly lower levels when compared to the Rest and WarmWater2 epochs. Finally, analyzing the *p*-values obtained by the Wilcoxon test, there is only a significant difference between the CPT and the WarmWater1 and between the CPT and the following epochs.

Figure 7 displays the pNN50 results. In accordance with the findings of the RMSSD, the epoch with the lowest pNN50 values was the CPT. In this epoch, the participant with the highest pNN50 had less than 40% of their heartbeats longer than 50 ms. Overall, the median values in each epoch seem to be similar. Even so, the epoch with the lowest median was the CPT (7.7%), with a 0.9% difference when compared to the Baseline and 2.4% compared with the WarmWater1 and Rest, while the WarmWater2 was the epoch with the highest median pNN50 (11.4%). There seems to be a consistent positive skewness on the boxplots, which means that the values of the upper quartile are more dispersed. This may be due to natural differences between the participants. Along with the protocol, there is a general

increase of values from the Baseline to the WarmWater1, followed by a decrease during the CPT and a subsequent rise during the WarmWater2 and Rest epochs, indicating a recovery after the CPT. Unlike the previous features, there was no significant statistical difference shown between the epochs.

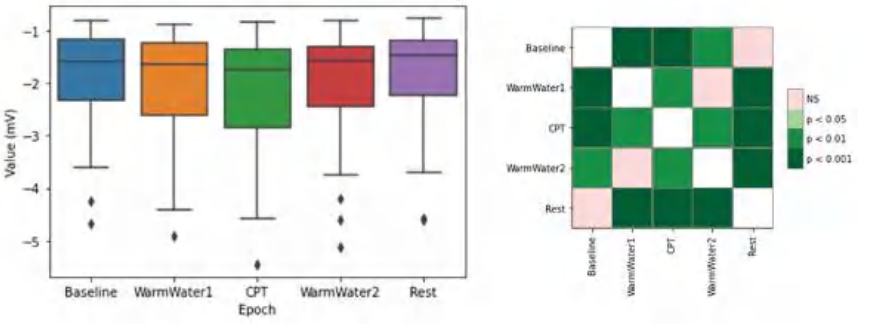


Figure 5. Boxplot for the mean minimum ECG cycles (**left**) and respective matrix of calculated *p*-values between different epochs, with Bonferroni correction, for the (**right**). The ◇ stands for outliers.

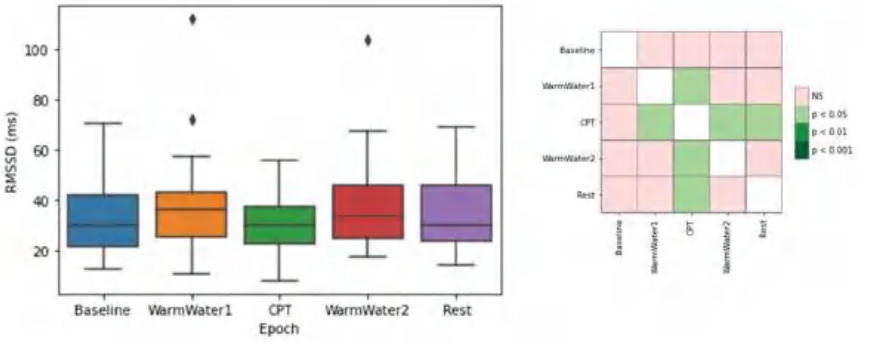


Figure 6. Boxplot of RMSSD values for each epoch (**left**) and respective *p*-values between different epochs, with Bonferroni correction (**right**). The ◇ stands for outliers.

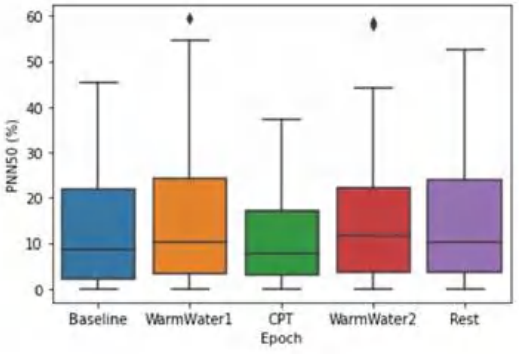


Figure 7. Boxplot of pNN50 values for each epoch. The ◇ stands for outliers.

Finally, the regularity and complexity of each epoch are presented in Figure 8, through the SampEn values. The epoch with the lowest value was the CPT. Another interesting observation is the results in the WarmWater2, which had generally higher values and a noticeable increase in the median value, which implies less predictability. Looking at the

Baseline and WarmWater1, both have equal median values (1.45). However, the values showed greater dispersion on the latter, which denotes greater behavioral differences in participants when compared to the former epoch. Lastly, the Rest epoch had a similar mean value to the two initial epochs and smaller dispersion, suggesting that, overall, the participants were able to recover after the CPT. The statistical analysis (Figure 8—right) only indicates a statistically significant difference between the CPT and the remaining epochs, with the exception of the Rest.

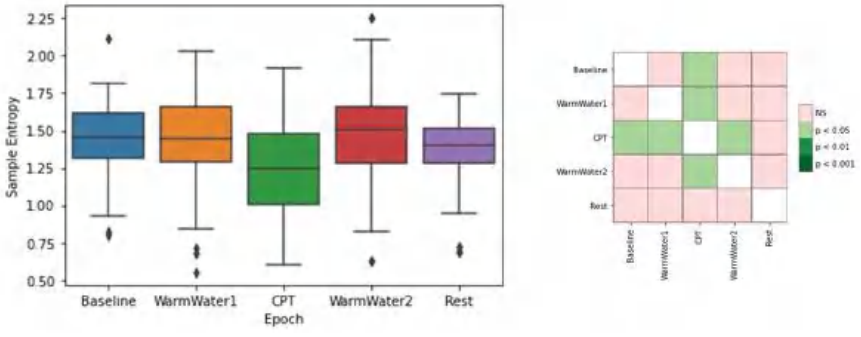


Figure 8. Boxplot of SampEn values for each epoch (**left**) and respective *p*-values between different epochs, with Bonferroni correction (**right**). The ⋄ stands for outliers.

4.2. EMG Processing and Analysis

With respect to the EMG signal, the features described in Table 2 were analyzed.

Table 2. Description of the extracted EMG features.

Features	Description
RMSE and RMSA	Root Mean Square for Electromyogram and Amplitude, respectively. Related to the constant force and not-fatiguing contractions of the muscles.
VAR	Variance. It allows for expressing the power of the EMG signal.

Figure 9 represents the Root Mean Square (RMS) of the electromyogram (RMSE), which is usually associated with the force a muscle exerts. It is clear that there is a progression in the RMSE values for the trapezius muscle from the Baseline, where the muscle was at rest, until the CPT when the participant has its forearm placed in cold water, experiencing pain, which was then followed by a steady decrease as the participant returned to rest. This suggests that the increase in the not fatiguing muscle contraction during CPT was, presumably, to endure the pain. Focusing on the boxplots, another interesting observation is near to no dispersion of the values during the non-painful epochs indicating a stable behavior from all the participants during these periods. The greater dispersion during the CPT may be due to different individual responses in reaction to pain. Concerning this muscle, the statistical analysis further demonstrates a significantly different behavior during the painful stimuli (CPT) in comparison with all the remaining epochs, especially with the Baseline and Rest epochs. On the contrary, the RMSE values for the triceps muscle did not indicate any type of evolution during the study, except for an increase in the Rest epoch. As for the statistical analysis, there were no significant differences between the values on the CPT and the other epochs, except for the Rest. These results are putting forward that the RMSE is not an adequate feature to study the triceps contraction force, and may be hypothesized that the trapezius is a better muscle to study the effects of pain on the body caused by the CPT.

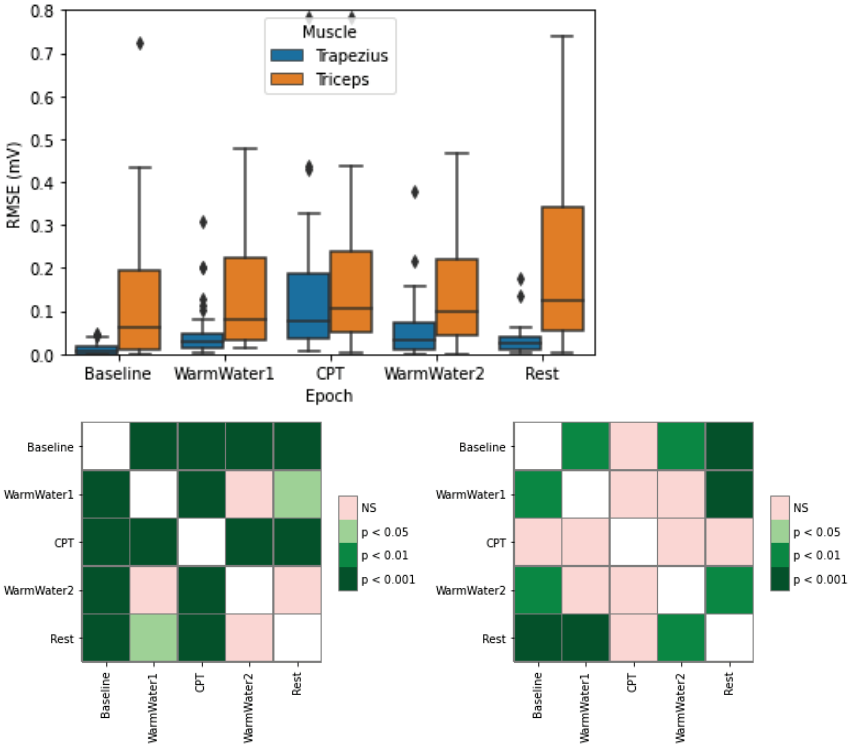


Figure 9. Boxplot of RMSE for the trapezius and triceps muscle (**left**) with respective statistical analysis with *p*-values, with Bonferroni correction, for the trapezius (**bottom left**) and triceps (**bottom right**). The ⋄ stands for outliers.

Figure 10 represents the RMS of the amplitude (RMSA). The results show a general increase in the values from the baseline to the CPT on both muscles. This implies that there was an increase in the contraction force from the Baseline, where both muscles were relaxed, to the CPT, where they were subjected to the cold-painful stimulus. On both muscles, the values decreased during the WarmWater2 and Rest, evidencing that the muscle was able to recover. Regarding the trapezius, the results present very little dispersion between the values during the non-painful epochs. For the triceps, the same is true for the Baseline and Rest epochs.

Concerning the results of the CPT, it is visible that the triceps had, overall, higher RMSA values. Nevertheless, its median value is not only lower than the median of the trapezius, but it is also more similar to the values obtained in WarmWater1 and WarmWater2, seemingly indicating that there was a great dispersion in results among the participants, with half of them not showing a significant reaction in the presence of the cold-painful stimulus, while the opposite was true for the remaining half of the participants. As for the trapezius muscle, a greater dispersion is also observable, especially from the median upwards. There is also observable greater dispersion, especially from the median upwards, and compared with the remaining epochs, presents a higher median.

For the trapezius, there are significant differences between the CPT and all the other epochs, supporting a change in the behavior of the trapezius contraction during the painful stimulus. It is also noticeable differences between the Rest and Warmwater1. The RMSA of the triceps also seems to validate that assumption, as there are significant differences between the CPT and the Baseline and between the CPT and the Rest.

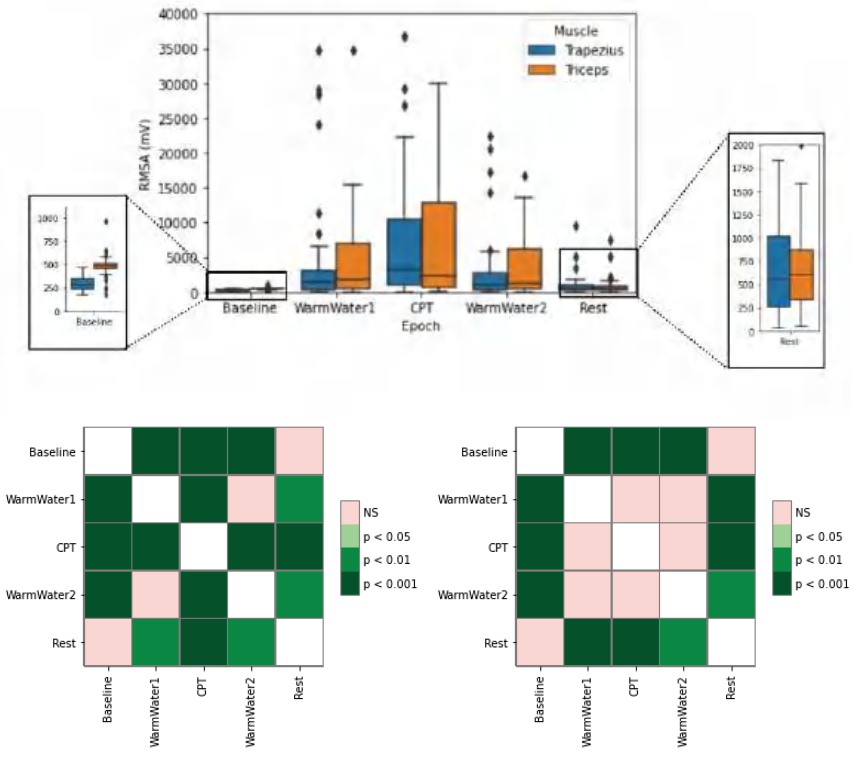


Figure 10. Boxplot of RMSA for the trapezius and triceps muscle (**top**) with respective statistical analysis with *p*-values, with Bonferroni correction, for the trapezius (**bottom left**) and triceps (**bottom right**). The ◇ stands for outliers.

Figure 11 presents the variance (VAR) results, of both muscles, for the different epochs (the values for the Baseline are close to one due to the normalization). The WarmWater1 shows an almost identical behavior between the two muscles. Even so, the trapezius does appear to have slightly higher values (both with and without considering the outliers) and a median marginally greater value. For the cold water, the data show a reaction to pain, with an increase in the variance. This was especially visible in the trapezius, which had overall considerable higher values, both in the maximum values (excluding outliers) and in the median value (18.06 mV), which is almost double the median of the triceps (9.95 mV). Although there are changes in the variance of the EMG from both muscles, the trapezius showed a more acute reaction to pain than the triceps. After the CPT, there was a decrease, showing a recovery from the pain. The variance is slightly lower relative to the WarmWater1, meaning that, before the beginning of the painful stimulus, the participants applied more power onto their muscles, giving further evidence that participants demonstrated apprehension at the beginning of the CPT. Nonetheless, there are some observed outliers, especially on the trapezius, whose values are closely similar to the ones observed on the CPT. This demonstrates that not all participant's musculoskeletal systems could recover immediately after the painful stimulus. During the Rest period, the variance roughly returned to values near one, similar to the Baseline, with a slightly higher level of dispersion observed on the trapezius. The statistical analysis for the trapezius shows, as in previous EMG features, very significant differences between the CPT and all the remaining epochs. As for the other epochs, only WarmWater1 and WarmWater2 showed no significant differences from each other. For the triceps, there is a significant

difference between the CPT and the other epochs, except for the WarmWater2, which may indicate a slightly longer recovery period of this muscle.

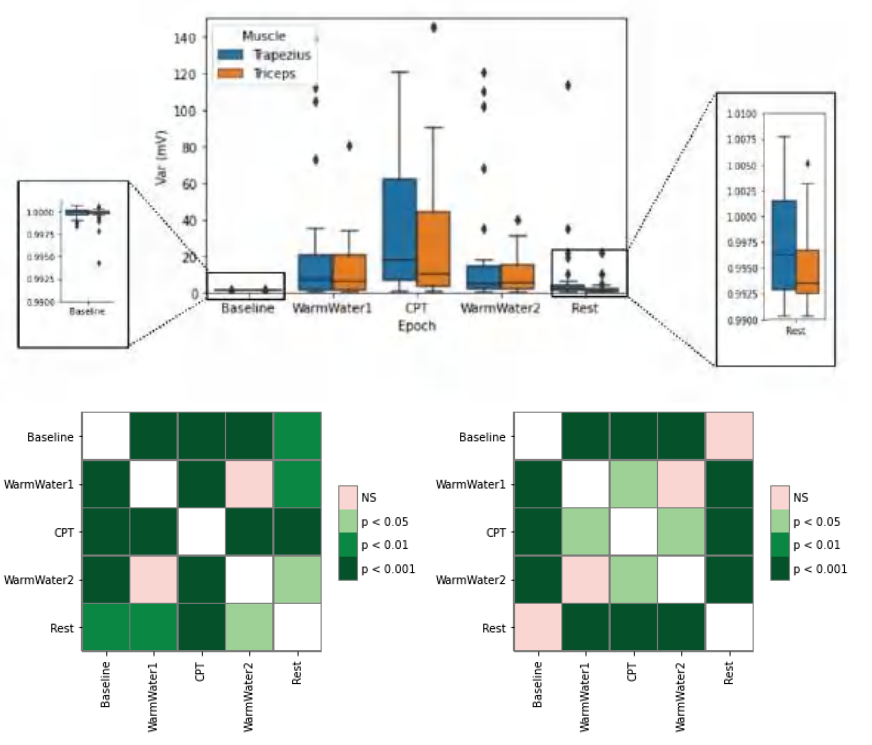


Figure 11. Boxplot of VAR for the trapezius and triceps muscle (**top**) with respective statistical analysis, with *p*-values, with Bonferroni correction, for the trapezius (**bottom left**) and triceps (**bottom right**). The ◇ stands for outliers.

4.3. EDA Processing and Analysis

With respect to the EDA signal, the number of SCR (Skin Conductance Response) peaks for each epoch was added up to identify the epoch with larger sympathetic activation. Furthermore, EDA indexes of the sympathetic nervous system (EDASymp) were also calculated based on the findings of [17], who argue that dynamics of the sympathetic component of the EDA signal are represented in the frequency band of 0.045–0.25 Hz.

Table 3 presents a brief description of these features.

Table 3. Description of the extracted EDA features.

Features	Description
SCR peaks	The number of SCR peaks (cumulative calculation of SCR peaks for each participant and epoch averaged and normalized by time in seconds).
EDASymp	Indexes of the sympathetic nervous system for the frequency band of 0.045–0.25 Hz [17].

Figure 12 (left) represents the mean sum of SCR peaks that occurred on a given epoch. Since the epochs varied greatly in length time, the results had to be normalized by time (in seconds). SCR peaks arise from a response to a stimulus. Considering the bar chart and the statistical analysis, it is apparent that the values reached their highest point during the CPT and exhibited statistically significant differences from all the other epochs. These results

suggest that the EDA signals of the participants were sensitive to the pain induced by the CPT. As for the remaining epochs, they registered, as expected, fewer peaks, with values pre and post-CPT being relatively similar and with no significant statistical differences between them. The Baseline and WarmWater1, nevertheless, have slightly higher values, which seems to be consistent with what was already hypothesized, regarding the general anticipation of the participants before the CPT.

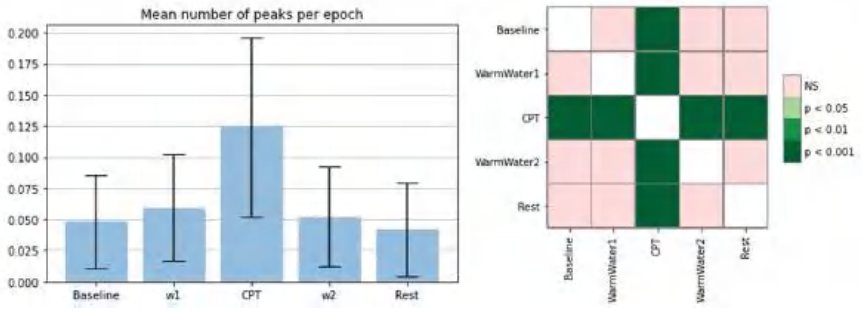


Figure 12. Bar chart of SCR peaks per second with corresponding standard deviation (**left**); *p*-values between different epochs, with Bonferroni correction (**right**).

The results for the EDASymp are presented in Figure 13. The results show that, overall, although with great levels of dispersion, the epoch with the highest values is the CPT, which seems to support the evidence that the sympathetic outflow increased when the participants were induced into pain.

As for the statistical analysis, it shows that the only significant statistical differences were between the CPT and both Baseline and WarmWater2 epochs. This sustains that the CPT did induce some changes in the sudomotor activity of the participants, in comparison with their initial state.

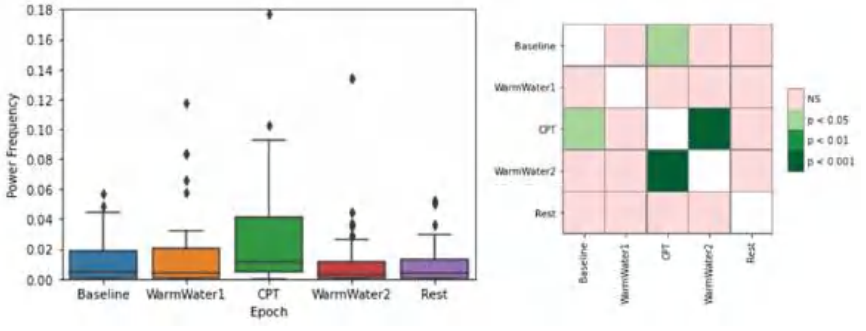


Figure 13. Boxplot of EDASymp values (**left**) and respective *p*-values between different epochs, with Bonferroni correction (**right**). The ◇ stands for outliers.

Figure 14 represents the evaluation of the systolic and diastolic BP values throughout the study. From the first measurement (before the CPT) to the second measurement (right after the CPT), the systolic and diastolic BP had an increase of 6% and 7%, respectively. In the third BP measurement, five minutes after the ending of the CPT, both systolic and diastolic values returned to their original values. This shows that the participants were able to recover to their initial state, which is supported by the results of the statistical analysis, revealing no significant differences between the first and third BP measurements and significant differences between both the CPT and previous measurements and the CPT and after measurements.

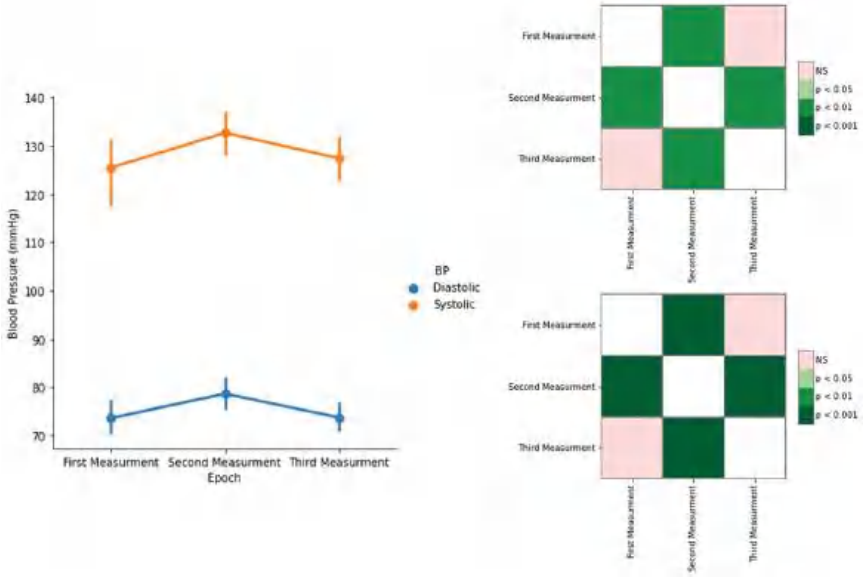


Figure 14. Systolic and diastolic BP values (mean ± level of confidence) (**left**). Statistical differences between the BP measurements and *p*-values, with Bonferroni correction, for systolic (**upper right**) and diastolic (**bottom right**).

5. Discussion

For the study of induced pain, the different physiological systems analyzed seemed to respond in accordance with what was hypothesized. On the ECG, the results on the meanHR were similar to what was described in previous literature [2–4,14,18], and is most likely a result of the increased sympathetic outflow on the body. As for the RMSSD, an HRV metric, which is an estimation of the vagally mediated changes, demonstrated a decrease in the parasympathetic outflow to the cardiovascular system during the pain-inducing task. The RMSSD also showed lower values for the two epochs before the CPT, when compared to those that preceded it. This may be a result of a higher level of anxiety felt by the participants before being subjected to pain. The statistical analysis also corroborates this conjecture, as there were no significant differences observed between the Baseline and CPT. The SampEn, which measures the regularity of the signal, demonstrated that the pain induced by the CPT caused a reaction in the participants that lead to a more consistent heartbeat pattern in their cardiac system. The WarmWater2 presented higher values, which implies less predictability and may be attributed to the recovery time that the body needed to return to the initial state by decreasing its HR. Lastly, the mean amplitude of S-peaks showed a progressive decrease in values until the CPT, followed by a progressive increase in the latter epochs. The statistical analysis for this feature also endorsed that there was a response in the S-waves of the ECG to the pain. In general, the results for the ECG features show that the cardiac system seems to react to the cold-painful stimulus. However, it must be pointed out that the RMSSD, the only time-domain HRV metric analyzed, can only contribute to the observation of the PNS, which means that it is not possible to conclude if the reaction observed on the ECG signal is due to the activation of the Sympathetic Nervous System (SNS) or, simply, due to the suppression of the PNS.

The [19] concluded that, when in a state of stress, due to the decrease in parasympathetic activity and increase in sympathetic, the energy will move from the cardiac system into the muscle. An overall analysis of the results for the EMG signal discloses that there was a reaction on part of the musculoskeletal system to the cold pain stimulus. As the RMSA, which is the representation of the non-fatiguing force, both muscles showed greater

median values and large dispersion for the CPT epoch—thus corroborating the premise of reaction on the ANS due to the presence of pain, more evident for the trapezius. The statistical analysis shows that the Baseline presents significant differences with all the remaining epochs, indicating that, during the Rest, the muscle did not recover to its original state. The VAR, a representation of the power, also conveys a response of the muscle to the induced pain, showing that the trapezius had a more acute reaction to the pain. For this study, in general, the trapezius seemed to be a consistently more stable source of information in comparison with the triceps. In addition, and according to study [19], our findings indicate an increase in the SNS reactivity, earlier in the procedure, which was further augmented during the CPT intensifying the activity recorded on the EMG.

Overall, the results for the EMG features validate the previous research in the area [15,19]: firstly, the RMSA, where both muscles showed a well-marked of value boxplots for the CPT epoch, thus corroborating the premise of reaction on the ANS due to the presence of pain. The statistical analysis shows that the Baseline presents significant differences with all the remaining epochs, indicating that, during the Rest, the muscle did not recover to the original state, whereas the triceps show differences between the baseline and the remaining epochs, except for the Rest period, indicating that it was able to recover.

For the EDA, it is noticeable that there was a response in the sudomotor activity of the participants when subjected to the painful stimulus. Comparing both time domain and frequency domain analysis, the former yielded better results, especially when taking into account the statistical results. Since the EDA features only measure the SNS, it can be concluded that, indeed, this system was activated during the pain induction task (CPT), in accordance with related works reported [4]. Finally, the BP shows statistically significant differences between consecutive measurements, which is in agreement with the literature [18,20,21].

Overall the results of this study are similar to what was reported in related works [2–4,14,15,18–21]. Nonetheless, this investigation goes further than previous literature, since it uses a higher number of physiological data, and, thus, a deep analysis of the effects of pain in the body.

6. Conclusions and Further Research

A new data collection protocol for induced pain is described and evaluated in the present work. For that, four different signals (ECG, EMG, EDA, and BP) were collected. The major innovation in this protocol was the use of a wider range of signals, which allowed for a broader analysis of the ANS reactivity on the various body systems.

Under this study, a deep evaluation of physiological data was performed, and, thus, a more concrete analysis of the effects of pain in the body was provided.

From the ECG, a significant increase in the HR and a decrease in the PNS activity were observed, based on the HRV metrics calculated, as a result of the cold-painful stimulus. Furthermore, the ECG also suffered a change in its amplitude, which was particularly noticeable in the S-wave evaluation.

The EMG, recorded both on the trapezius and triceps muscles, also showed changes brought on by the pain-inducing protocol—mainly an increase in amplitude during the CPT, in comparison with the other resting periods. Moreover, the results on the trapezius muscles seemed to indicate that the stabilization of the values after an initial increase was crucial to withstanding the painful stimulus for the participants who completed the CPT.

Both time and frequency domain features on the EDA demonstrated an increase in the values during the CPT and hence an increase in the SNS.

Finally, the BP shows statistically significant differences between consecutive measurements, which is in accordance with the literature.

Since this study uses a variety of physiological signals, future work should be concerned with the study of the signals interrelation in the process of pain and devoted to multimodal classification providing further reliable measurements of pain. Moreover, a setback in this study is the short length of time recordings of the CTP, which hindered the

study of the majority of HRV metrics. Thus, a protocol with an increased length of the CPT would allow the investigation into the influence of the SNS on the cardiovascular system.

Author Contributions: Experimental protocol conceptualization, R.S.; data acquisition: A.B.; methodology, R.S. and S.B.; data analysis, A.B.; validation, R.S. and S.B.; visualization, A.B.; supervision, R.S. and S.B.; writing—original draft preparation, R.S. and A.B.; writing—review and editing, R.S., A.B. and S.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by national funds through FCT—Fundação para a Ciência e a Tecnologia, I.P., under the Scientific Employment Stimulus—Individual Call—CEECIND/03986/2018, and is also supported by the FCT through national funds, within IEETA/UA R&D unit (UIDB/00127/2020). This work is also funded by national funds, the European Regional Development Fund, FSE through COMPETE2020, through FCT, in the scope of the framework contract foreseen in the numbers 4, 5, and 6 of the article 23, of the Decree-Law 57/2016, of 29 August, changed by Law 57/2017, of 19 July.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics and Deontological Council of the University of Aveiro (number 09-CED/2019).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data are protected by the GDPR and cannot be publicly available.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Raja, S.; Carr, D.; Cohen, M.; Finnerup, N.; Flor, H.; Gibson, S.; Keefe, F.; Mogil, J.; Ringkamp, M.; Sluka, K.; et al. The revised International Association for the Study of Pain definition of pain: Concepts, challenges, and compromises. *Pain* **2020**, *161*, 1976–1982. [CrossRef] [PubMed]
2. Loggia, M.L.; Napadow, V. Multi-parameter autonomic-based pain assessment: More is more? *Pain* **2012**, *153*, 1779–1780. [CrossRef] [PubMed]
3. Cowen, R.; Stasiowska, M.K.; Laycock, H.; Bantel, C. Assessing pain objectively: The use of physiological markers. *Anaesthesia* **2015**, *70*, 828–847. [CrossRef] [PubMed]
4. Hampf, G. Influence of cold pain in the hand on skin impedance, heart rate and skin temperature. *Physiol. Behav.* **1990**, *47*, 217–218. [CrossRef] [PubMed]
5. Kregel, K.C.; Seals, D.R.; Callister, R. Sympathetic nervous system activity during skin cooling in humans: relationship to stimulus intensity and pain sensation. *J. Physiol.* **1992**, *454*, 359–371. [CrossRef] [PubMed]
6. Younger, J.; Mccue, R.; Mackey, S. Pain Outcomes: A Brief Review of Instruments and Techniques. *Curr. Pain Headache Rep.* **2009**, *13*, 39–43. [CrossRef] [PubMed]
7. Birnie, K.A.; Caes, L.; Wilson, A.C.; Williams, S.E.; Chambers, C.T. A practical guide and perspectives on the use of experimental pain modalities with children and adolescents. *Pain Manag.* **2014**, *4*, 97–111. [CrossRef] [PubMed]
8. McCaul, K.D.; Monson, N.; Maki, R.H. Does distraction reduce pain-produced distress among college students? *Health Psychol.* **1992**, *11*, 210–217. [CrossRef] [PubMed]
9. Myers, C.; Robinson, M.; Riley, J.; Sheffield, D. Sex, Gender, and Blood Pressure: Contributions to Experimental Pain Report. *Psychosom. Med.* **2001**, *63*, 545–550. [CrossRef] [PubMed]
10. Tousignant-Laflamme, Y.; Rainville, P.; Marchand, S. Establishing a Link Between Heart Rate and Pain in Healthy Subjects: A Gender Effect. *J. Pain Off. J. Am. Pain Soc.* **2005**, *6*, 341–347. [CrossRef] [PubMed]
11. Schestatsky, P.; Valls-Solé, J.; Costa, J.; León, L.; Veciana, M.; Chaves, M.L. Skin autonomic reactivity to thermoalgesic stimuli. *Clin. Auton. Res.* **2007**, *17*, 349–355. [CrossRef] [PubMed]
12. Hallman, D.; Lindberg, L.G.; Arnetz, B.; Lyskov, E. Effects of static contraction and cold stimulation on cardiovascular autonomic indices, trapezius blood flow and muscle activity in chronic neck–shoulder pain. *Eur. J. Appl. Physiol.* **2011**, *111*, 1725–1735. [CrossRef] [PubMed]
13. Brusselmans, G.; Nogueira, H.; De Schamphelaere, E.; Devulder, J.; Crombez, G. Skin Temperature during Cold Pressor Test in Fibromyalgia: An Evaluation of the Autonomic Nervous System? *Acta Anaesthesiol. Belg.* **2015**, *66*, 19–27.
14. Jiang, M.; Rosio, R.; Syrjälä, E.; Anzanpour, A.; Terävä, V.; Rahmani, A.M.; Salanterä, S.; Aantaa, R.; Hagelberg, N.; Liljeberg, P. Acute pain intensity monitoring with the classification of multiple physiological parameters. *J. Clin. Monit. Comput.* **2019**, *33*, 493–507. [CrossRef] [PubMed]
15. Gouverneur, P.; Li, F.; Adamczyk, W.M.; Szikszay, T.M.; Luedtke, K.; Grzegorzek, M. Comparison of Feature Extraction Methods for Physiological Signals for Heat-Based Pain Recognition. *Sensors* **2021**, *21*, 4838. [CrossRef] [PubMed]
16. MATLAB. *Version 9.10.0.1684407 (R2021a)*; The MathWorks Inc.: Natick, MA, USA, 2021.

17. Posada-Quintero, H.F.; Florian, J.P.; Orjuela-Cañón, A.D.; Aljama-Corrales, T.; Charleston-Villalobos, S.; Chon, K.H. Power Spectral Density Analysis of Electrodermal Activity for Sympathetic Function Assessment. *Ann. Biomed. Eng.* **2016**, *44*, 3124–3135. [CrossRef] [PubMed]
18. Streff, A.; Kuehl, L.K.; Michaux, G.; Anton, F. Differential physiological effects during tonic painful hand immersion tests using hot and ice water. *Eur. J. Pain* **2010**, *14*, 266–272. [CrossRef] [PubMed]
19. Wijsman, J.; Grundlehner, B.; Penders, J.; Hermens, H. Trapezius Muscle EMG as Predictor of Mental Stress. *ACM Trans. Embed. Comput. Syst.* **2013**, *12*, 1–20. [CrossRef]
20. Lovallo, W. The Cold Pressor Test and Autonomic Function: A Review and Integration. *Psychophysiology* **1975**, *12*, 268–282. [CrossRef] [PubMed]
21. Weise, F.; Laude, D.; Girard, A.; Zitoun, P.; Siché, J.P.; Elghozi, J.L. Effects of the cold pressor test on short-term fluctuations of finger arterial blood pressure and heart rate in normal subjects. *Clin. Auton. Res.* **1993**, *3*, 303–310. [CrossRef] [PubMed]

sensors

MDPI

Article

Evaluating the Window Size's Role in Automatic EEG Epilepsy Detection

Vasileios Christou [1,*], Andreas Miltiadous [1], Ioannis Tsoulos [1], Evaggelos Karvounis [1], Katerina D. Tzimourta [1,2], Markos G. Tsipouras [2], Nikolaos Anastasopoulos [3], Alexandros T. Tzallas [1] and Nikolaos Giannakeas [1,*]

[1] Department of Informatics and Telecommunications, University of Ioannina, 47100 Arta, Greece
[2] Department of Electrical and Computer Engineering, Faculty of Engineering, University of Western Macedonia, 50100 Kozani, Greece
[3] Department of Electrical and Computer Engineering, University of Patras, 26504 Rio, Greece
* Correspondence: bchristou1@gmail.com (V.C.); giannakeas@uoi.gr (N.G.)

Abstract: Electroencephalography is one of the most commonly used methods for extracting information about the brain's condition and can be used for diagnosing epilepsy. The EEG signal's wave shape contains vital information about the brain's state, which can be challenging to analyse and interpret by a human observer. Moreover, the characteristic waveforms of epilepsy (sharp waves, spikes) can occur randomly through time. Considering all the above reasons, automatic EEG signal extraction and analysis using computers can significantly impact the successful diagnosis of epilepsy. This research explores the impact of different window sizes on EEG signals' classification accuracy using four machine learning classifiers. The machine learning methods included a neural network with ten hidden nodes trained using three different training algorithms and the k-nearest neighbours classifier. The neural network training methods included the Broyden–Fletcher–Goldfarb–Shanno algorithm, the multistart method for global optimization problems, and a genetic algorithm. The current research utilized the University of Bonn dataset containing EEG data, divided into epochs having 50% overlap and window lengths ranging from 1 to 24 s. Then, statistical and spectral features were extracted and used to train the above four classifiers. The outcome from the above experiments showed that large window sizes with a length of about 21 s could positively impact the classification accuracy between the compared methods.

Keywords: EEG; seizure detection; window size; neural network; genetic algorithm; k-nearest neighbours

Citation: Christou, V.; Miltiadous, A.; Tsoulos, I.; Karvounis, E.; Tzimourta, K.D.; Tsipouras, M.G.; Anastasopoulos, N.; Tzallas, A.T.; Giannakeas, N. Evaluating the Window Size's Role in Automatic EEG Epilepsy Detection. *Sensors* **2022**, *22*, 9233. https://doi.org/10.3390/s22239233

Academic Editor: Manuel José Cabral dos Santos Reis

Received: 18 October 2022
Accepted: 24 November 2022
Published: 27 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Epilepsy is the most common condition affecting the central nervous system, where 80% of the patients are citizens from developing or middle-income countries [1]. Besides the young population, it can also occur in the elderly population (people over 65 years old) [2]. Epilepsy has a severe economic impact in terms of healthcare needs. It causes premature deaths and can lead to lost work productivity. Considering all the above reasons, it is an essential topic in the biomedical sciences [1,3].

Epilepsy is a chronic brain disease characterized by seizures affecting all age groups. It causes recurrent seizures, ranging from one episode per year to several episodes per day. There is a distinction between epilepsy and seizures since not all seizures are epileptic fits. The main characteristic of epilepsy is that it is responsible for triggering unprovoked recurrent seizures caused by chronic abnormal bursts of electrical discharges in the brain [4]. This process is called "epileptogenesis" and makes epilepsy highly unpredictable. Other types of seizure disorders can be activated by various causes, which can be measured, including stroke, tumours, and other space-occupying lesions. Secondary or symptomatic epilepsy is epilepsy caused due to an underlying abnormality of the structure of the brain

and is the type of epilepsy where preventive measures can be applied according to various causes. It can be noted that more than 60% of the cases lack a definitive cause. This epilepsy type is called primary or idiopathic epilepsy and is not preventable but can be treated using antiepileptic medicines [3,5].

The occurrence of epileptic seizures is due to a malfunction in the brain, which triggers a sudden excessive electrical discharge in a group of cells in the brain's cerebral cortex. This malfunction causes motor function abnormalities, resulting in tonic–clonic muscle spasms. The vast and abrupt energy surge triggered by the brain's neurons is the cause of epileptic seizures, which show differences in their properties. Seizures range from a few seconds to severe, generalized, and prolonged convulsions, leading to dangerous and life-threatening situations. Seizures' characteristics depend on the specific brain region involved, the extent of the abnormal electrical discharge and its spread [3,6].

The limited knowledge regarding the human brain creates a challenge in understanding the properties of a brain with epilepsy. The disease's temporary symptoms include mindfulness loss, minor (almost undetectable) abnormalities in movement, mild muscle twitching, and abnormalities in visual, auditory, and gustatory senses and mood. The epileptic seizures start and finish unexpectedly, without involving interference from the external environment, and it is possible to remain unnoticed. For this reason, detecting and measuring epileptic seizures is a challenging task [3,7].

Seizure occurrence is not always connected to epilepsy since, statistically, 10% of the world population will have one seizure during their life [3]. These nonepileptic seizure types can be caused by chemical imbalances. If two or more seizures occur without a specific reason, it may have been caused due to epilepsy. In case of epileptic seizures, the patient can start receiving antiepileptic medicines to improve their safety and quality of life. The unpredictable nature of epileptic seizures can be a severe life-threatening cause (e.g., if they are triggered while driving a car or swimming). The most common method for diagnosing epilepsy is an electroencephalogram (EEG) signal analysis. EEG signals reflect the brain's electrical activity at a given timestamp [3].

An EEG can record the electrical brain activity using a series of electrodes placed on the patient's scalp. Brain abnormalities that are not related to epilepsy can be analysed by studying EEG signals. Soikkeli et al. [8] investigated the generalized slowing of the EEG in patients with Parkinson's disease. Wieser et al. [9] studied Creutzfeldt–Jakob disease using EEG signals while Neto et al. [10] conducted a regularized linear discriminant analysis of EEG features taken from patients with dementia [3]. Overall, EEG has been used for the detection and quantification of many neurological diseases [11] or conditions [12] or cognitive states such as stress induction [13,14], thus becoming a significant tool for neurologists.

The study of epileptic seizures analyses EEG signals received before and during the seizures, which contain patterns that differentiate them from those recorded in a nonepileptic person. The identification of epileptic seizures is made by observing the EEG data. For this reason, an EEG signal analysis approach which provides information regarding the brain's condition must be applied [3].

This paper explores the impact of the window size on classifying epileptic short-term EEG signals using four machine learning methods. The machine learning methods used were a single-layer neural network (SLNN) with ten hidden nodes, trained using three different training algorithms, and the k-nearest neighbours (K-NN) classifier [15]. The neural network training methods were the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm [16], the multistart algorithm for global optimization problems proposed by [17], and the modified genetic algorithm (GA) proposed by Tsoulos [18].

This paper is structured into six main sections, starting with an "Introduction", which explains the significance of epilepsy, the importance of EEG for its diagnosis, and includes a short description of the research's motivation. The "Related Work" section contains existing work regarding automated methods for diagnosing epilepsy. The "Methods" section presents four machine learning methods for exploring the window size's effect on classifying epileptic short-term EEG signals. The "Results" section analyses the four

machine learning algorithms' results presented above using different window types applied to the University of Bonn epilepsy database [19]. The following two sections contain the "Discussion" and "Conclusion". Finally, the "Methods" section describes each machine learning method used to explore the window size effect on classifying epileptic short-term EEG signals.

2. Related Work

Existing seizure detection works include the method proposed by Naghsh-Nilchi and Aghashahi [20]. The proposed approach was based on two eigensystem pseudospectral estimation methods: eigenvector and multiple signal classification for time-domain EEG signal pseudospectrum estimation. The pseudospectrum was partitioned into sub-bands, each having a smaller frequency. Then, a feature extraction stage was applied to produce the input to a multilayer perceptron (MLP). The MLP classified the input vectors into three classes: normal, interictal and ictal. Tzallas et al. [21] compared various time–frequency (t-f) analysis methods for categorizing epileptic seizures EEG segments. A three-stage analysis was utilized, starting with the t-f analysis and a power spectrum density (PSD) calculation from each EEG segment. The next stage involved the extraction of a feature set by measuring the signal segment fractional energy on specific t-f windows. In contrast, the third stage was the categorization (normal and epileptic) of the EEG segment using artificial neural networks (ANNs). Martinez-del Rincon et al. [22] used an EEG analysis system for automatic epilepsy seizure detection that could exploit EEG data's underlying nonlinear nature. Hassan and Subasi [23] addresses the automated seizure detection problem using single-channel EEG signals. The EEG signal segments were initially decomposed using the complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) signal processing model. The training and testing data were formed by extracting six spectral moments from the CEEMDAN mode functions, which were entered as inputs to the linear programming boosting (LPBoost) classifier. Juarez-Guerra et al. [24] used a wavelet analysis system for identifying epilepsy seizures from EEG signals. The proposed system utilized the discrete wavelet transform (DWT) and the maximal overlap discrete wavelet transform (MODWT) for extracting a feature set. This set was entered as input to an ANN, which performed the classification task. Hossain et al. [25] used a CNN for feature learning from raw EEG data to detect seizures on an open-access EEG epilepsy dataset from the Boston Children's Hospital [26]. The proposed model extracted spectral and temporal features from EEG epilepsy data and utilized them to learn the overall structure of a seizure that was less sensitive to variations. Nicolaou and Georgiou [27] explored the use of permutation entropy (PE) as a feature for automatic epilepsy seizure detection. Their method utilized a support vector machine (SVM) for the binary classification task and was based upon the observation that the PE dropped during a seizure. Shoeb and Guttag [28] presented a method utilizing an SVM to construct patient-specific classifiers that could use EEG signals from patients' scalps to detect the onset of epileptic seizures. Guo et al. [29] proposed an EEG-based method for automatic epileptic seizure detection, which utilized the approximate entropy features derived from the multiwavelet transform. These features were introduced as input data to an ANN for classifying the EEG signals as epileptic or nonepileptic. Subasi [30] decomposed EEG signals into their frequency sub-bands using a wavelet transform. Then, these sub-bands were introduced as input to an ANN for classification into two categories (epileptic and nonepileptic). Moreover, this research developed and compared classifiers based on feedforward error backpropagation ANNs and dynamic wavelet networks. The comparison was made to test their accuracy in EEG signals classification. Ghosh-Dastidar et al. [31] combined the mixed-band wavelet-chaos methodology [32,33] with a principal component analysis (PCA)-enhanced cosine radial basis function neural network classifier for classifying EEG signals into three categories (healthy, ictal, and interictal). Guo et al. [34] proposed a method for automatic epileptic seizure detection. This method utilized line length features based on a wavelet transform multiresolution decomposition and introduced them as input to an ANN for classifying

the EEG signals into two categories (healthy or epileptic). Hassan et al. [35] proposed an automated epilepsy diagnosis system based on a tuneable-Q factor wavelet transform and bootstrap aggregating. Finally, the general-purpose method proposed by Tsoulos et al. [36] utilized genetic programming to create ANNs. The proposed method could infer the ANN's architecture and estimate the optimal number of neurons for each given problem.

3. Materials and Methods

This research studied the four machine learning methods that are analysed in the Methods section for exploring window size's effect on classifying epileptic short-term EEG signals.

The well-established epileptic database from the University of Bonn was used for the evaluation, since it is the most used database from the published databases. The Bonn database consists of 5 groups of recordings namely Z-O-N-F-S. The Z and O datasets consist of EEG recordings of healthy, nonepileptic participants with closed and open eyes, respectively. The N, F, and S subsets include intracranial EEG recordings acquired from five epileptic patients, during presurgical examination. Specifically, the N subset includes parts of interictal recordings originating from the epileptic zone of the opposite hemisphere, while the O subset includes parts of EEG recordings obtained from the epileptic zone. The S subset includes 100 intracranial EEG recordings, obtained from the epileptogenic zone during epileptic activity. The epileptogenic zone was the hippocampus and no further patient data were provided.

For the classification task, all 5 subsets of the Bonn database were used, for a 5-class Z-O-N-F-S problem. Each group consisted of 100 single-channel recordings with 23.6 s duration and all recordings were used for the training and testing. Before the experiment, a low-pass FIR filter at 40 Hz was applied to all recordings, and then the recordings were split into datasets of different time window lengths. The examined window lengths were 1–24 s (24 s being in fact 23.6 s).

For each examined window length, a set of extracted univariate and spectral features were calculated to create a feature vector. Specifically, the following time-domain features were extracted: mean, median, variance. Moreover, a fast Fourier transform was employed to transform the signal into the frequency domain and the spectrum amplitude of four EEG bands was calculated. The EEG bands were:

- Alpha band (8–12 Hz)
- Beta band (12–25 Hz)
- Theta band (4–8 Hz)
- Delta band (1–4 Hz)

The following subsections analyse the machine learning methodologies that were tested for the classification of the 5-class problem and the evaluation of the time window length. Particularly, Sections 3.1–3.3 analyse the optimization techniques used to optimize the hyperparameters of a 10-layer multilayer perceptron neural network. Section 3.4 analyses the last classification methodology, k-nearest neighbours (kNN).

3.1. The BFGS Method

The BFGS algorithm is a quasi-Newton approach utilizing a new updating formula which has become very popular and has been subjected to numerous modifications. Quasi-Newton methods are used to solve unconstrained optimization problems [16,37–41].

An unconstrained optimization problem can be described by using Equation (1):

$$\min_{x \in \mathbb{R}^n} f(x) \tag{1}$$

In this formula, $\mathbb{R}^n$ denotes an n-dimensional Euclidean space while $f : \mathbb{R}^n \to \mathbb{R}$ is continuously twice differentiable. The update formula of BFGS is defined in Equation (2) where s_k and y_k are the step vectors, and g is used to denote the gradient for Equation (1).

$$\begin{array}{ccc} s_k & \stackrel{def}{=} & x_{k+1} - x_k \\ y_k & = & g_{k+1} - g_k \end{array} \tag{2}$$

The BFGS method is considered the best among all quasi-Newton based methods. The updating formula for BFGS takes the form shown in Equation (3).

$$B_{k+1} = B_k + \frac{y_k y_k^T}{y_k^T s_k} - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} \tag{3}$$

In this formula, the B_k symbol denotes the Hessian approximation at x_k, and the matrix B_{k+1} is generated by (3) to satisfy the following secant formula:

$$B_{k+1} s_k = y_k \tag{4}$$

The above secant formula is considered an approximation of the Newton relation. The secant can be fulfilled if $s_k^T y_k > 0$, which is called the curvature condition and ensures that the BFGS updating matrix shown in (3) is positive definite [16]. Unconstrained optimization problems are solved using an iterative procedure. Equation (5) defines the iterative formula for quasi-Newton methods.

$$x_{k+1} = x_k + a_k d_k \tag{5}$$

In this formula, the term a_k defines the step size while d_k defines the search direction. The step must be a positive number in order $f(x)$ to be able to reduce sufficiently, while both a_k and d_k must be chosen carefully for an efficient search line. The step size can be calculated by using various formulas divided into two main categories (exact or inexact line search). An ideal choice would be the exact line choice defined by the formula $a_k = \arg\min(f(x_k + a_k d_k)), a > 0$ but it is computationally intensive to define this value. The reason behind this problem is that it requires a large number of evaluations for the objective function f and its gradient g. The inexact line search has a number of formulas proposed by different researchers, including the formulas of Armijo [42], Wolfe [43,44], and Goldstein [45] with the first one being the easiest one to implement. The Armijo search line formula is defined in (6).

$$f(x_k) - f(x_k + a_k d_k) \geq -\sigma a_k g_k^T d_k \tag{6}$$

Given $s > 0, \lambda \in (0,1), \sigma \in (0,1)$ and $a_i = \max\{s, s\lambda, s\lambda^2, \dots\}$ such that $k = 0, 1, 2, 3, \dots$, the reduction in f should be proportional to both the step size and directional derivative $g_k^T d_k$ [16].

The search directions are important for determining the f value, and the quasi-Newton methods can be defined using the following equation.

$$d_k = -B_k^{-1} g_k \tag{7}$$

In this formula, B_k is a nonsingular symmetric approximation matrix of the Hessian defined in (3). The initial matrix B_0 is an identity matrix updated by an update formula. When d_1 is defined from the above formula and B_k is a positive definite matrix, then $d_k^T = -g_k^T B_k^{-1} g_k < 0$, which makes d_k a descent direction. Algorithm 1 describes the iterative process of the BFGS algorithm [16].

Algorithm 1 : The BFGS Algorithm

1: Having a starting point x_0 and $B_0 = I_n$. Set the values for s, β, and σ.
2: End if $\|g(x_{k+1})\| < 10^{-6}$.
3: Calculate the search direction using Formula (7).
4: Calculate the difference $s_k = x_{k+1} - x_k$ and $y_k + g_{k+1} - g_k$.
5: Update B_k by (3) in order to obtain B_{k+1}.
6: $k = k + 1$.
7: Go to step 2.

The current research uses the BFGS variant proposed by Powell [46]. The main advantage of Powell's methodology is that the step along the search direction is not restricted by constraints having small residuals, which significantly increases efficiency, specifically the nearly degenerate constraints.

3.2. The Multistart Method

The multistart method described in Algorithm 2 is a two-phase stochastic black-box global optimization approach consisting of a global and a local phase. In black-box optimization problems, no known structure can be used, and the problem can be formulated by minimizing, for example, a continuous function f over a compact set $S \subseteq \mathbb{R}^n$. Due to the nature of stochastic problems where the outcome is random, it is particularly suitable for black-box optimization problems. Another characteristic of these approaches is that they require little to no assumptions about the optimization problem. On the other hand, they can only provide a probabilistic convergence guarantee in the best-case scenario [47].

In the first phase of a two-phase method, many randomly sampled points in the feasible region are used to evaluate the function. In the second phase, a local search procedure is applied to each sample point mentioned above, yielding various local optima. Amongst all local optima, the best one forms the resulting estimation of the global optimum [17,47].

Algorithm 2 : The Multistart Algorithm

1: $i = 0$ and $X^* = .$
2: Take a random sample x from S.
3: Start a deterministic local search process at x and conclude at a local minimum x^*.
4: Check if a new minimum is found.
5: $x^* \notin X^*$ then
6: $\quad i \leftarrow i + 1$.
7: $\quad x_i^* = x^*$.
8: $\quad X^* \leftarrow X^* \cup \{x_i^*\}$.
9: end.
10: If ending criteria have been met, terminate the process.
11: Go to step 2.

3.3. The Modified GA Method

GAs are global optimization methods based on Charles Darwin's theory of natural evolution. A GA begins with a pool of candidate solutions, which are the artificial equivalent of chromosomes in biological organisms. Then, these chromosomes are evolved in an iterative process using the selection, crossover, and mutation genetic operations. The process is continued until the termination criterion is reached, or the algorithm converges to the best chromosome, which can be the optimal or a suboptimal solution of the problem [18].

The real-coded GA proposed by Kaelo and Ali [48] can be seen in Algorithm 3. In this algorithm, the problem is to find the global minimum of the following unconstrained optimization problem.

$$\text{minimize } f(x) \text{ subject to } x \in \Omega \tag{8}$$

where $f(x) : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ is a continuous real-valued function and x is an n-dimensional continuous variable vector. The term Ω denotes a box or other region which is easy to sample. The x_{opt} point is the global minimizer of f if $f_{opt} = f(x_{opt}) \leq f(x), \forall x \in \Omega$. At each iteration of the algorithm (generation), the candidate points set S is updated which new chromosomes (offspring) created by the reproduction process (crossover and mutation) of the algorithm [18,48].

Algorithm 3 : The Real-Coded GA

1: Create N random points in Ω from the uniform distribution.
2: Store the points in set S.
3: $iter = 0$.
4: Evaluate each chromosome using its function value.
5: If the termination criteria are achieved, stop the GA.
6: Select $m \leq N$ parents from S.
7: Create m offspring using the selected parent chromosomes of the previous step.
8: Mutate the offspring with probability p_m.
9: Remove the m worst chromosomes and replace them with the offspring.
10: Create a trial point $\tilde{x}$. If $f(\tilde{x}) \leq f(x_h)$ where x_h is the current worst point in S, then replace x_h with $\tilde{x}$.
11: $iter = iter + 1$.
12: Go to step 4.

The real-coded GA starts by creating the initial population in the first two lines, followed by the initialization of the generation counter. The following step evaluates the population. In step 5, the GA checks if the termination criteria have been achieved and terminates the algorithm. The termination is done when $|f_h - f_1| \leq e$ or the maximum number of iterations has been reached. The term f_h denotes the function value of the most optimal chromosome in the population, while f_h denotes the function value of the least optimal chromosome in the population. If the termination criteria have not been achieved, the evolution process continues. In step 6, the selection of two parent chromosomes $x = (x_1, x_2, \ldots, x_n)$ and $y = (y_1, y_2, \ldots, y_n)$ for the reproduction process is done using the tournament selection [49] mechanism. Step 7, creates the offspring using the equations shown in (9)

$$\begin{aligned} \tilde{x_l} &= a_i x_i + (1 - a_i) y_i \\ \tilde{y_l} &= a_i y_i + (1 - a_i) x_i \end{aligned} \tag{9}$$

where $a_i \in [-0.5, 1.5]$ [50]. The mutation procedure in step 8 follows the formula depicted in (10).

$$x_i' = \begin{cases} x_i + \Delta(iter, r_i - x_i), & t = 0 \\ x_i - \Delta(iter, x_i - l_i), & t = 1 \end{cases} \tag{10}$$

In this formula, t is a random number taking the values 0 or 1, $iter$ is the current generation and $\Delta(iter, y) = y(1 - r^{(1 - \frac{iter}{ITERMAX})})$ with $r \in [0, 1]$ and $ITERMAX$ being the maximum allowed number of generations. Step 9 replaces the m worst chromosomes in the population with the offspring. Step 10 is the local technique that creates trial points to replace the least optimal points in the population. Using the following equation, this technique initially selects a random point y from S and creates a trial point $\tilde{x_i}$.

$$\tilde{x_i} = (1 + \gamma_i) x_{l,i} - \gamma_i y_i, \quad i = 1, \ldots, n \tag{11}$$

where $\gamma_i \in [-0.5, 0.5]$ and $x_{l,i}$ is the ith component of the most optimal chromosome x_l. The technique ends by replacing the least optimal point x_h in S with $\tilde{x}$, if $f(\tilde{x}) \leq f(x_h)$ [18,48].

The current paper used the modifications proposed by Tsoulos [18]. These modifications include a novel stopping rule, a new mutation operator, and a local search procedure application. This procedure is applied to the most optimal chromosome x_l every

K_{ls} generations, with K_{ls} being a constant that defines the frequency of the applied local search procedure.

3.4. The K-NN Classifier

The K-NN algorithm is one of the simplest and oldest classification algorithms [15]. It has a set containing n samples $D_n = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$, where $X_i \in \mathbb{R}^d$ are the vectors containing the features and $Y_i \in \{\omega_1, \omega_2, \ldots, \omega_M\}$ are the labels which correspond to each class. The K-NN algorithm categorizes a new input pattern x into the class of its nearest neighbour in the n training examples. The identification of the closest class is made using the Euclidean distance (although other distance metrics can be used) [51,52]. The K-NN method can be seen in Algorithm 4.

Algorithm 4 : The K-NN Algorithm

1: Classify (X, Y, x).
2: for $i = 1$ to n do
3: Calculate the Euclidean distance $d_E(X_i, x)$.
4: end.
5: Compute set I having the indices for the k smallest distances $d_E(X_i, x)$.
6: Return majority label for Y_i where $i \in I$.

4. Results

The current research investigated the role of the window size in epilepsy EEG signal analysis by running a series of experiments using the database from the University of Bonn [19]. The tests were performed using a 10-fold cross-validation and are visualized in Table 1 and Figure 1.

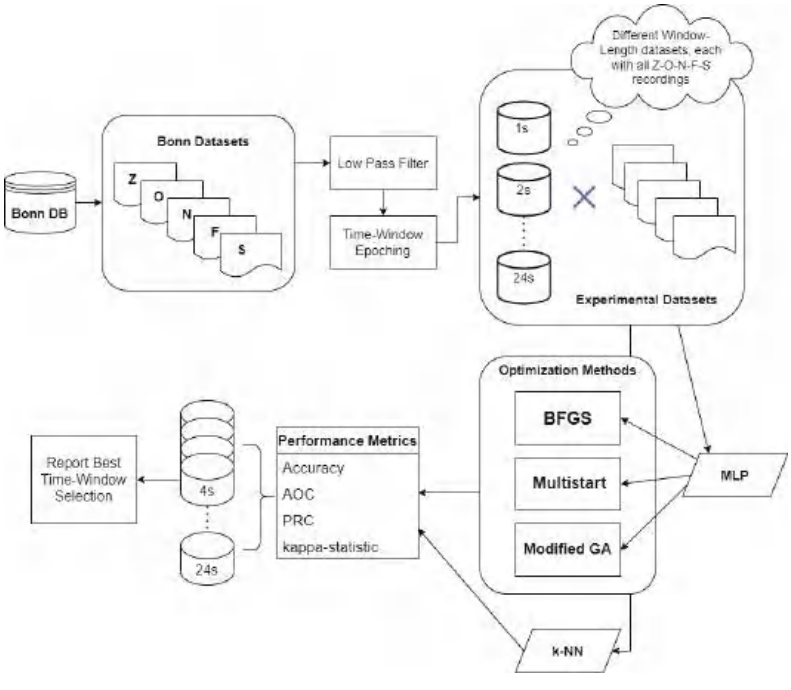


Figure 1. Flowchart of the proposed methodology.

All experiments were repeated 30 times with the window size ranging from 1 to 24 s. The number in each method's cell represents the average classification accuracy of

the test set for each window size (1–24 s). The accuracy for one fold was defined as the number of correctly classified instances divided by the total number of instances, as seen in Formula (12).

$$accuracy = \frac{correctly\ \ classified\ \ instances}{total\ \ number\ \ of\ \ instances} \quad (12)$$

The accuracy was calculated by estimating the average value over all folds and then calculating the average value over all experiment runs. The SLNN used for training in the Broyden–Fletcher–Goldfarb–Shanno (BFGS), multistart and modified genetic algorithm (GA) methods had ten hidden neurons, and in every iteration of the multistart approach, a BFGS method was used to optimize the weights. Finally, the k-nearest neighbours (K-NN) method with $K = 2$ was used.

In the experimental results depicted in Table 1, the bold fonts describe the time window that achieved the highest accuracy for each methodology.

Table 1. Experimental Results expressed in classification accuracy for the four algorithms employed regarding time windows ranging from 1 to 24 s. BFGS stands for Broyden–Fletcher–Goldfarb–Shanno algorithm. GA stands for genetic algorithm, K-NN stands for k-nearest neighbours.

Epoch (s)	BFGS	Multistart	GA	K-NN
1 s	56.86%	57.68%	56.91%	68.9%
2 s	65.06%	65.56%	65.06%	75.14%
3 s	69.7%	69.57%	69.01%	76.66%
4 s	72.62%	70.53%	70.06%	76.99%
5 s	75.69%	73.46%	71.96%	77.89%
6 s	74.63%	76.37%	75.44%	79.53%
7 s	74.76%	75.84%	74.43%	79.1%
8 s	76.06%	75.55%	74.95%	78.41%
9 s	76.25%	77.64%	76.5%	79.88%
10 s	76.96%	77.12%	76.38%	80.05%
11 s	76.42%	79.01%	77.2%	79.08%
12 s	76.55%	78.26%	77.06%	79.84%
13 s	77.04%	78.04%	76.05%	78.56%
14 s	77.81%	78.26%	77.13%	79.01%
15 s	79.75%	78.98%	78.41%	78.68%
16 s	77.35%	80.98%	78.59%	79.52%
17 s	77.7%	78.05%	77.82%	79.92%
18 s	78.5%	79.24%	78.10%	79.92%
19 s	80.7%	79.71%	78.47%	79.49%
20 s	**80.92%**	**81.59%**	80.78%	80.00%
21 s	**80.92%**	81.23%	**81.06%**	79.25%
22 s	80.04%	80.88%	81.00%	**81.17%**
23 s	80.69%	80.88%	80.89%	78.88%
24 s	80.25%	80.43%	79.98%	79.04%

It is seen that the window size dramatically impacted the accuracy values since when the window had a size of 20–21 s, the accuracy had its highest value and decreased when the window size gradually increased or decreased. The multistart method obtained the highest accuracy with a window size between 20 and 21 s (81.59%). Regarding the BFGS algorithm, the highest accuracy was achieved at with 20-s and 21-s time windows (80.92%), while the GA methodology achieved the highest accuracy when the time window was 21 s (81.06%). Finally, the K-NN algorithm achieved its best accuracy scores with a 22-s time window (81.17%).

Table 2 illustrates other standard evaluation measures for the K-NN algorithm, namely the area under the ROC, the area under the PRC, and the kappa statistic. The results of this table are in agreement with Table 1, with the 20–21-second time windows achieving the best performances at every evaluation metric.

Table 2. Area under the ROC, area under the PRC, and kappa statistic regarding the classification performance of the K-NN algorithm.

Epoch (s)	AOC	PRC	k-Stat
1 s	78.91%	48.6%	62.21%
2 s	79.89%	50.2%	68.74%
3 s	80.68%	50.1%	75.23%
4 s	86.44%	53.3%	71.95%
5 s	85.92%	56.8%	74.62%
6 s	85.45%	54.0%	76.38%
7 s	83.21%	58.1%	77.55%
8 s	87.21%	60.9%	77.19%
9 s	87.17%	61.8%	80.02%
10 s	86.57%	64.3%	78.84%
11 s	90.89%	64.2%	83.40%
12 s	90.49%	64.8%	82.32%
13 s	89.04%	68.1%	82.14%
14 s	88.88%	68.3%	82.85%
15 s	86.22%	70.4%	79.94%
16 s	85.45%	70.1%	80.15%
17 s	85.92%	73.6%	82.15%
18 s	84.70%	73.0%	84.29%
19 s	86.07%	74.7%	85.42%
20 s	92.22%	**78.5%**	**85.49**%
21 s	**92.51%**	76.5%	83.26%
22 s	88.70%	77.3%	82.44%
23 s	82.28%	75.7%	83.51%
24 s	88.37%	73.7%	80.00%

5. Discussion

The current article investigated the time window size's impact on EEG signal classification for epilepsy detection. The experimental part utilized three neural networks trained using three different algorithms (BFGS, multistart, modified GA) and the K-NN classifier. The experiments were repeated 30 times, and the average classification accuracy was reported.

The primary outcome from the experimental results summarized in Table 1 was that the window size in epilepsy EEG signals significantly impacted the classification accuracy of the compared methods. It was shown that for more accurate results, the window size must be between 20 and 21 s. Another significant outcome was the mixed results regarding the method which managed to get the best accuracy for each window size. There was no clear winning method for all window sizes, but the results varied when the window size changed.

An appropriate window length selection is crucial for machine learning methodologies on signal data (such as EEG). Too small time windows may fail to capture each condition's signal characteristics. For example, a very small time window in an epilepsy methodology may result in not being able to capture the complete seizure waveforms. On the other side, too large time windows may capture signal properties of two different situations (such as ictal state and interictal state), thus negatively affecting the classification performance. The proposed study can be utilized in future methodologies that propose a classification scheme for EEG epilepsy detection problems. Our study's resulting optimal window length agreed with another study proposed by Tzimourta et al. [53]. This study evaluated the optimal window length using different classification algorithms (naive Bayes, MLP, support vector machines, and decision trees) and found that 21-s windows achieved the best accuracy results. Moreover, our results suggested that the 20–21-s windows achieved the best performance. These findings agreed with Thangavel et al. [54], who classified epileptic signals using different features and examined different window lengths, concluding that the 20-s time window generated some of the best performance results.

However, some limitations regarding our methodology should be mentioned. One of them is the restricted length of the recordings, which did not allow exploring time windows larger than 24 s. To alleviate this limitation, a future extension of this methodology that incorporates longer EEG recordings from other publicly available databases should be performed. Furthermore, no wavelet transformations were used for the feature extraction step, as well as a limited number of machine learning algorithms were used (neural networks and K-NN), limiting the ability to generalize these findings to all automatic EEG epilepsy detection methodologies.

6. Conclusions

Epilepsy has attracted much attention from the research community because it can affect various people ranging from very young to the elderly. It can also have a serious economic impact on healthcare needs; it can cause premature deaths and lead to lost work productivity. Consequently, much scientific effort has been made to propose machine learning methodologies that perform automatic epilepsy detection from EEG signals. These methodologies commonly perform epoching of the time signals to produce the experiment's training and test set. Thus, the window size in the signal decomposition is significant for detecting subtle changes in the EEG recording. This study evaluated the optimal time window length for four classification algorithms: three neural networks trained using the BFGS, multi-start and modified GA methods and the K-NN approach. Time windows from 1 to 24 s were explored and examined regarding the classification accuracy of the four algorithms. The epoching of 20–21 s achieved the best classification performance.

Author Contributions: Conceptualization and methodology, I.T.; software, I.T. and N.A.; validation, E.K., A.T.T. and M.G.T.; investigation, K.D.T. and A.T.T.; data curation, M.G.T. and N.G.; writing—original draft preparation, V.C. and I.T.; writing—review and editing, V.C., A.M., I.T., E.K., K.D.T., M.G.T., N.A., A.T.T. and N.G.; visualization, A.T.T. and A.M.; supervision, I.T., M.G.T., A.T.T. and N.G. All authors have read and agreed to the published version of the manuscript.

Funding: We acknowledge support for this work from the project "Immersive Virtual, Augmented and Mixed Reality Center Of Epirus" (MIS 5047221), which is implemented under the Action "Reinforcement of the Research and Innovation Infrastructure".

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by Ethics Committee of University of Ioannina.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Written informed consent has been obtained from the patient(s) to publish this paper.

Data Availability Statement: The research utilizes the database from the University of Bonn [19].

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

EEG	electroencephalogram
K-NN	k-nearest neighbours
BFGS	Broyden–Fletcher–Goldfarb–Shanno
SLNN	single-layer neural network
GA	genetic algorithm
BCI	brain–computer interface
MLP	multilayer perceptron
t-f	time frequency
PSD	power spectrum density

ANNs	artificial neural networks
CEEMDAN	complete ensemble empirical mode decomposition with adaptive noise
LPBoost	linear programming boosting
DWT	discrete wavelet transform
MODWT	maximal overlap discrete wavelet transform
PE	permutation entropy
SVM	support vector machine
CSI	combined seizure index
PCA	principal component analysis

References

1. World Health Organization. *Epilepsy*; WHO: Geneva, Switzerland, 2020.
2. Ramsay, R.E.; Rowan, A.J.; Pryor, F.M. Special considerations in treating the elderly patient with epilepsy. *Neurology* **2004**, *62*, S24–S29. [CrossRef] [PubMed]
3. Acharya, U.R.; Sree, S.V.; Swapna, G.; Martis, R.J.; Suri, J.S. Automated EEG analysis of epilepsy: A review. *Knowl.-Based Syst.* **2013**, *45*, 147–165. [CrossRef]
4. Tzallas, A.T.; Tsipouras, M.G.; Tsalikakis, D.G.; Karvounis, E.C.; Astrakas, L.; Konitsiotis, S.; Tzaphlidou, M. Automated Epileptic Seizure Detection Methods: A Review Study. In *Epilepsy*; Stevanovic, D., Ed.; IntechOpen: Rijeka, Croatia, 2012; Chapter 4. [CrossRef]
5. Cross, D.J.; Cavazos, J.E. The role of sprouting and plasticity in epileptogenesis and behavior. In *Behavioral Aspects of Epilepsy*; DEMOS: New York, NY, USA, 2007.
6. Buck, D.; Baker, G.A.; Jacoby, A.; Smith, D.F.; Chadwick, D.W. Patients' experiences of injury as a result of epilepsy. *Epilepsia* **1997**, *38*, 439–444. [CrossRef] [PubMed]
7. Iasemidis, L.D.; Shiau, D.S.; Sackellares, J.C.; Pardalos, P.M.; Prasad, A. Dynamical resetting of the human brain at epileptic seizures: Application of nonlinear dynamics and global optimization techniques. *IEEE Trans. Biomed. Eng.* **2004**, *51*, 493–506. [CrossRef]
8. Soikkeli, R.; Partanen, J.; Soininen, H.; Pääkkönen, A.; Riekkinen Sr, P. Slowing of EEG in Parkinson's disease. *Electroencephalogr. Clin. Neurophysiol.* **1991**, *79*, 159–165. [CrossRef]
9. Wieser, H.G.; Schindler, K.; Zumsteg, D. EEG in Creutzfeldt–Jakob disease. *Clin. Neurophysiol.* **2006**, *117*, 935–951. [CrossRef]
10. Neto, E.; Biessmann, F.; Aurlien, H.; Nordby, H.; Eichele, T. Regularized linear discriminant analysis of EEG features in dementia patients. *Front. Aging Neurosci.* **2016**, *8*, 273. [CrossRef]
11. Miltiadous, A.; Tzimourta, K.D.; Giannakeas, N.; Tsipouras, M.G.; Afrantou, T.; Ioannidis, P.; Tzallas, A.T. Alzheimer's Disease and Frontotemporal Dementia: A Robust Classification Method of EEG Signals and a Comparison of Validation Methods. *Diagnostics* **2021**, *11*, 1437. [CrossRef]
12. Christodoulides, P.; Miltiadous, A.; Tzimourta, K.D.; Peschos, D.; Ntritsos, G.; Zakopoulou, V.; Giannakeas, N.; Astrakas, L.G.; Tsipouras, M.G.; Tsamis, K.I.; et al. Classification of EEG signals from young adults with dyslexia combining a Brain Computer Interface device and an Interactive Linguistic Software Tool. *Biomed. Signal Process. Control* **2022**, *76*, 103646. [CrossRef]
13. Aspiotis, V.; Miltiadous, A.; Kalafatakis, K.; Tzimourta, K.D.; Giannakeas, N.; Tsipouras, M.G.; Peschos, D.; Glavas, E.; Tzallas, A.T. Assessing Electroencephalography as a Stress Indicator: A VR High-Altitude Scenario Monitored through EEG and ECG. *Sensors* **2022**, *22*, 5792. [CrossRef]
14. Miltiadous, A.; Aspiotis, V.; Sakkas, K.; Giannakeas, N.; Glavas, E.; Tzallas, A.T. An experimental protocol for exploration of stress in an immersive VR scenario with EEG. In Proceedings of the 2022 7th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM), Ioannina, Greece, 23–25 September 2022; pp. 1–5. [CrossRef]
15. Fix, E.; Hodges, J. *Discriminatory Analysis, Nonparametric Discrimination: Consistency Properties*; Technical Report, TX, Tech. Rep. 4; USAF School of Aviation Medicine, Randolph Field: Dayton, OH, USA, 1951.
16. Hery, M.A.; Ibrahim, M.; June, L. BFGS method: A new search direction. *Sains Malays.* **2014**, *43*, 1591–1597.
17. Lagaris, I.E.; Tsoulos, I.G. Stopping rules for box-constrained stochastic global optimization. *Appl. Math. Comput.* **2008**, *197*, 622–632. [CrossRef]
18. Tsoulos, I.G. Modifications of real code genetic algorithm for global optimization. *Appl. Math. Comput.* **2008**, *203*, 598–607. [CrossRef]
19. Andrzejak, R.G.; Lehnertz, K.; Mormann, F.; Rieke, C.; David, P.; Elger, C.E. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Phys. Rev. E* **2001**, *64*, 061907. [CrossRef] [PubMed]
20. Naghsh-Nilchi, A.R.; Aghashahi, M. Epilepsy seizure detection using eigen-system spectral estimation and Multiple Layer Perceptron neural network. *Biomed. Signal Process. Control* **2010**, *5*, 147–157. [CrossRef]
21. Tzallas, A.T.; Tsipouras, M.G.; Fotiadis, D.I. Epileptic seizure detection in EEGs using time–frequency analysis. *IEEE Trans. Inf. Technol. Biomed.* **2009**, *13*, 703–710. [CrossRef]

22. Martinez-del Rincon, J.; Santofimia, M.J.; del Toro, X.; Barba, J.; Romero, F.; Navas, P.; Lopez, J.C. Non-linear classifiers applied to EEG analysis for epilepsy seizure detection. *Expert Syst. Appl.* **2017**, *86*, 99–112. [CrossRef]
23. Hassan, A.R.; Subasi, A. Automatic identification of epileptic seizures from EEG signals using linear programming boosting. *Comput. Methods Programs Biomed.* **2016**, *136*, 65–77. [CrossRef]
24. Juarez-Guerra, E.; Alarcon-Aquino, V.; Gomez-Gil, P. Epilepsy seizure detection in EEG signals using wavelet transforms and neural networks. In *New Trends in Networking, Computing, E-Learning, Systems Sciences, and Engineering*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 261–269.
25. Hossain, M.S.; Amin, S.U.; Alsulaiman, M.; Muhammad, G. Applying deep learning for epilepsy seizure detection and brain mapping visualization. *ACM Trans. Multimed. Comput. Commun. Appl. (TOMM)* **2019**, *15*, 1–17. [CrossRef]
26. Shoeb, A.H. Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment. Ph.D. Thesis, Massachusetts Institute of Technology, Boston, MA, USA, 2009.
27. Nicolaou, N.; Georgiou, J. Detection of epileptic electroencephalogram based on permutation entropy and support vector machines. *Expert Syst. Appl.* **2012**, *39*, 202–209. [CrossRef]
28. Shoeb, A.H.; Guttag, J.V. Application of machine learning to epileptic seizure detection. In Proceedings of the 27th International Conference on Machine Learning (ICML-10), Haifa, Israel, 21–24 June 2010; pp. 975–982.
29. Guo, L.; Rivero, D.; Pazos, A. Epileptic seizure detection using multiwavelet transform based approximate entropy and artificial neural networks. *J. Neurosci. Methods* **2010**, *193*, 156–163. [CrossRef] [PubMed]
30. Subasi, A. Epileptic seizure detection using dynamic wavelet network. *Expert Syst. Appl.* **2005**, *29*, 343–355. [CrossRef]
31. Ghosh-Dastidar, S.; Adeli, H.; Dadmehr, N. Principal component analysis-enhanced cosine radial basis function neural network for robust epilepsy and seizure detection. *IEEE Trans. Biomed. Eng.* **2008**, *55*, 512–518. [CrossRef] [PubMed]
32. Adeli, H.; Ghosh-Dastidar, S.; Dadmehr, N. A wavelet-chaos methodology for analysis of EEGs and EEG subbands to detect seizure and epilepsy. *IEEE Trans. Biomed. Eng.* **2007**, *54*, 205–211. [CrossRef]
33. Ghosh-Dastidar, S.; Adeli, H.; Dadmehr, N. Mixed-band wavelet-chaos-neural network methodology for epilepsy and epileptic seizure detection. *IEEE Trans. Biomed. Eng.* **2007**, *54*, 1545–1551. [CrossRef]
34. Guo, L.; Rivero, D.; Dorado, J.; Rabunal, J.R.; Pazos, A. Automatic epileptic seizure detection in EEGs based on line length feature and artificial neural networks. *J. Neurosci. Methods* **2010**, *191*, 101–109. [CrossRef]
35. Hassan, A.R.; Siuly, S.; Zhang, Y. Epileptic seizure detection in EEG signals using tunable-Q factor wavelet transform and bootstrap aggregating. *Comput. Methods Programs Biomed.* **2016**, *137*, 247–259. [CrossRef]
36. Tsoulos, I.G.; Gavrilis, D.; Glavas, E. Neural network construction using grammatical evolution. In Proceedings of the 5h IEEE International Symposium on Signal Processing and Information Technology, Athens, Greece, 18–21 December 2005; pp. 827–831.
37. Broyden, C.G. The convergence of a class of double-rank minimization algorithms 1. general considerations. *IMA J. Appl. Math.* **1970**, *6*, 76–90. [CrossRef]
38. Broyden, C.G. The convergence of a class of double-rank minimization algorithms: 2. The new algorithm. *IMA J. Appl. Math.* **1970**, *6*, 222–231. [CrossRef]
39. Fletcher, R. A new approach to variable metric algorithms. *Comput. J.* **1970**, *13*, 317–322. [CrossRef]
40. Goldfarb, D. A family of variable-metric methods derived by variational means. *Math. Comput.* **1970**, *24*, 23–26. [CrossRef]
41. Shanno, D.F. Conditioning of quasi-Newton methods for function minimization. *Math. Comput.* **1970**, *24*, 647–656. [CrossRef]
42. Armijo, L. Minimization of functions having Lipschitz continuous first partial derivatives. *Pac. J. Math.* **1966**, *16*, 1–3. [CrossRef]
43. Wolfe, P. Convergence conditions for ascent methods. *SIAM Rev.* **1969**, *11*, 226–235. [CrossRef]
44. Wolfe, P. Convergence conditions for ascent methods. II: Some corrections. *SIAM Rev.* **1971**, *13*, 185–188. [CrossRef]
45. Goldstein, A.A. On steepest descent. *J. Soc. Ind. Appl. Math. Ser. Control* **1965**, *3*, 147–151. [CrossRef]
46. Powell, M. A tolerant algorithm for linearly constrained optimization calculations. *Math. Program.* **1989**, *45*, 547–566. [CrossRef]
47. Pardalos, P.M.; Romeijn, H.E.; Tuy, H. Recent developments and trends in global optimization. *J. Comput. Appl. Math.* **2000**, *124*, 209–228. [CrossRef]
48. Kaelo, P.; Ali, M. Integrated crossover rules in real coded genetic algorithms. *Eur. J. Oper. Res.* **2007**, *176*, 60–76. [CrossRef]
49. Miller, B.L.; Goldberg, D.E. Genetic algorithms, tournament selection, and the effects of noise. *Complex Syst.* **1995**, *9*, 193–212.
50. Michalewicz, Z.; Hartley, S.J. Genetic algorithms+data structures=evolution programs. *Math. Intell.* **1996**, *18*, 71.
51. Wang, J.G.; Neskovic; Cooper. An adaptive nearest neighbor algorithm for classification. In Proceedings of the 2005 International Conference on Machine Learning and Cybernetics, Guangzhou, China, 18–21 August 2005; Volume 5, pp. 3069–3074.
52. Tay, B.; Hyun, J.K.; Oh, S. A machine learning approach for specification of spinal cord injuries using fractional anisotropy values obtained from diffusion tensor images. *Comput. Math. Methods Med.* **2014**, *2014*, 276589. [CrossRef] [PubMed]
53. Tzimourta, K.D.; Astrakas, L.G.; Gianni, A.M.; Tzallas, A.T.; Giannakeas, N.; Paliokas, I.; Tsalikakis, D.G.; Tsipouras, M.G. Evaluation of window size in classification of epileptic short-term EEG signals using a Brain Computer Interface software. *Eng. Technol. Appl. Sci.* **2018**, *8*, 3093–3097. [CrossRef]
54. Thangavel, P.; Thomas, J.; Sinha, N.; Peh, W.Y.; Yuvaraj, R.; Cash, S.S.; Chaudhari, R.; Karia, S.; Jin, J.; Rathakrishnan, R.; et al. Improving automated diagnosis of epilepsy from EEGs beyond IEDs. *J. Neural Eng.* **2022**, *19*, 066017. [CrossRef]

sensors

MDPI

Article

Learning-Based Image Damage Area Detection for Old Photo Recovery

Tien-Ying Kuo [1,*], Yu-Jen Wei [1], Po-Chyi Su [2] and Tzu-Hao Lin [1]

[1] Department of Electrical Engineering, National Taipei University of Technology, Taipei 10608, Taiwan
[2] Department of Computer Science and Information Engineering, National Central University, Taoyuan City 32001, Taiwan
* Correspondence: tykuo@ntut.edu.tw

Abstract: Most methods for repairing damaged old photos are manual or semi-automatic. With these methods, the damaged region must first be manually marked so that it can be repaired later either by hand or by an algorithm. However, damage marking is a time-consuming and labor-intensive process. Although there are a few fully automatic repair methods, they are in the style of end-to-end repairing, which means they provide no control over damaged area detection, potentially destroying or being unable to completely preserve valuable historical photos to the full degree. Therefore, this paper proposes a deep learning-based architecture for automatically detecting damaged areas of old photos. We designed a damage detection model to automatically and correctly mark damaged areas in photos, and this damage can be subsequently repaired using any existing inpainting methods. Our experimental results show that our proposed damage detection model can detect complex damaged areas in old photos automatically and effectively. The damage marking time is substantially reduced to less than 0.01 s per photo to speed up old photo recovery processing.

Keywords: deep learning; damage area detection; damaged old photo

Citation: Kuo, T.-Y.; Wei, Y.-J.; Su, P.-C.; Lin, T.-H. Learning-Based Image Damage Area Detection for Old Photo Recovery. *Sensors* **2022**, *22*, 8580. https://doi.org/10.3390/s22218580

Academic Editor: Manuel José Cabral dos Santos Reis

Received: 18 October 2022
Accepted: 4 November 2022
Published: 7 November 2022

1. Introduction

Old photos can often contain various levels of damage caused by human improper storage or environmental factors that deteriorate the integrity of photos. Fortunately, digital image processing technology can be applied to recover the content of these photos to its original state. The existing recovery methods for damaged old photos can be divided into non-automatic and automatic processes according to whether human intervention is required. The non-automatic methods can be further subdivided into manual and semi-automatic methods. Manual recovery is made through a variety of image editing tools, such as Photoshop or GIMP [1], to recover damaged photos based on user knowledge. The semi-automatic method manually marks the damaged areas on the photos and then applies the inpainting methods [2,3] to recover the contents of these locations. The mentioned works focused on the design of repair methods. For example, Li et al. [2] modified the confidence computation, strategy matching, and filling scheme to improve the inpainting method. Zhao et al. [3] proposed an inpainting model based on the generative adversarial network (GAN) and gated convolution [4]. With their methods, in addition to the damaged photos as the input, additional damage masks should be specified before inputting into the model. Non-automatic methods, while providing good recovery results, require physically marking the damaged areas in the photos, taking a lot of time and effort.

The automatic method does not have the aforementioned problems as it does not require any additional information in the process of restoring damaged old photos. Works [5,6] have used deep learning techniques to develop automatic methods that can be applied to a wider variety of photo content and types of damage. Wan et al. [5] designed a model based on the architecture of variational autoencoder (VAE) [7]. They used an encoder model to

first obtain the feature vectors representing the input photo in the latent space, then used the latent restoration network to remove the damage and noisy components embedded in the feature vectors, and finally the feature vectors were reverted back to the recovered photo. Liu et al. [6] designed two modules: latent class-attribute restoration (LCR) and dynamic condition-guided restoration (DCR). LCR first analyzes the four class attributes of smoothness, clarity, connectivity, and completeness in the photo to repair the global defects and then uses multiple DCRs in series to process the local defects to restore the details in the photo. Although the automatic method can reduce the processing time for restoring damaged old photos, the results generated are not satisfactory. For example, in [5,6], some textures and objects were removed from the recovered photos because they were mistakenly treated as noise or damage, and some undamaged areas in the photos were also modified, which is undoubtedly a problem for preserving the integrity of the photo content.

In order to improve these shortcomings, we propose a method by which to automatically detect damaged areas in old photos and use the detection results to guide inpainting methods to automatically recover the original content of these areas. In general, damaged area detection involves finding damaged areas in objects, such as steel structures [8], murals [9], photos [10,11], frescoes [12], and pavements [13–19], through algorithms. The methods for detecting damaged areas can be divided into traditional algorithms and deep learning algorithms depending on the development method.

Damage detection methods [9–12] were developed using traditional image processing techniques. Jaidilert et al. [9] used seeded region growing [20] and morphology to detect cracks. Bhuvaneswari et al. [10] combined a bilateral filter and Haar wavelet transform to detect scratch damage in images. The Hough transformation was used in [11] to detect line cracks in images. Cornelis et al. [12] believed that the luminance value of cracks is low, so the top-hat transformation of morphology was used to find cracks with a low luminance value. The damage detection methods mentioned above are not effective in detecting irregular damage areas and can only detect simple damage with limited accuracy, which may affect their subsequent repair performance.

In deep learning-based algorithms, although there are a few fully automatic repair methods [5,6] as mentioned previously, they are in the style of end-to-end repairing, which means that it is not easy to have control over the detection of damaged areas, potentially destroying or being unable to completely preserve valuable historical photos to the full degree. We note that although the image content is different between worn-out old photos and pavement crack images, the damage types are similar and both include mainly irregular cracks, so we review and discuss the related literature on pavement crack detection as well. König et al. [13] replaced standard convolutional blocks with residual blocks and added an attention gating mechanism to preserve spatial correlation in the feature map and suppress gradients in unrelated regions. Yang et al. [14] proposed a feature pyramid and hierarchical boosting network (FPHBN) to fuse features of different sizes. Lau et al. [15] used a pre-trained ResNet-34 to enhance the feature extraction capability of the network, while Liu et al. [16] used the dilated convolution approach to make the area of the receptive field wider.

It is mentioned in [17] that the ratio between cracked and non-cracked pavement is very imbalanced, often leading to poor network segmentation results and the failure of network training for crack detection, and a similar problem exists in our task. The solution to the imbalance between the cracked and non-cracked data can be adjusted by either the data set [15,17,18] or the loss function [15,16,19]. The dataset adjustment strategy breaks the picture into smaller blocks, such as 48×48, 64×64, or even multiple block sizes [15] for the training model, and then picks the proper ratio of cracked and non-cracked blocks for training to reduce the dataset imbalance problem. For example, Zhang et al. [17] used cracked blocks only as the training set for their crack-patch-only (CPO) supervised adversarial learning. Jenkins et al. [18] set the specific ratio between cracked and non-cracked blocks in the training set to place more weight on cracked blocks. As for the loss function, most works use binary cross-entropy (BCE) as a loss function

for semantic segmentation-like applications, but this function is weak in handling the imbalanced dataset issue. As a consequence, Lau et al. [15] replaced BCE functions with dice coefficients to evaluate the correctness of the detected areas. Liu et al. [16] further combined the BCE functions and the dice coefficients. Cheng et al. [19] applied distance weight to improve the original binary cross entropy. Existing deep learning-based road crack detection algorithms can work on more complex and diverse damage than can traditional algorithms. However, since there are many differences between the content of road images and old photos, it is not possible to use the road crack detection method directly, so we need to develop a method suitable for detecting damage in old photos.

To summarize the main contributions of our work, unlike other literature approaches where the content of some intact areas is changed during repair, our way of recovering damaged old photos ensures no alteration of intact areas during repair to preserve photo integrity and fidelity. Since the existing methods for detecting image damage are not satisfactory, in this paper an automatic damage detection method is proposed for the recovery of old damaged photos to save time and effort. The advantage of our work is that our detection result enables the possibility of combining any subsequent inpainting methods to repair the photo, which is not possible using existing automatic end-to-end repairing methods.

2. Proposed Method

Our recovery processing of damaged old photos is divided into two parts, as shown in Figure 1. In the detection model (M_D), the model input is an old damaged photo ($I_{damaged}$) and the model output is a damaged area mask ($Mask$). The $I_{damaged}$ and the $Mask$ are then exported to the inpainting method (M_R) to generate the repaired photo ($I_{Repaired}$), where the M_R can be any existing method.

$$Mask = M_D\left(I_{damaged}\right) \tag{1}$$

$$I_{Repaired} = M_R\left(Mask, I_{damaged}\right) \tag{2}$$

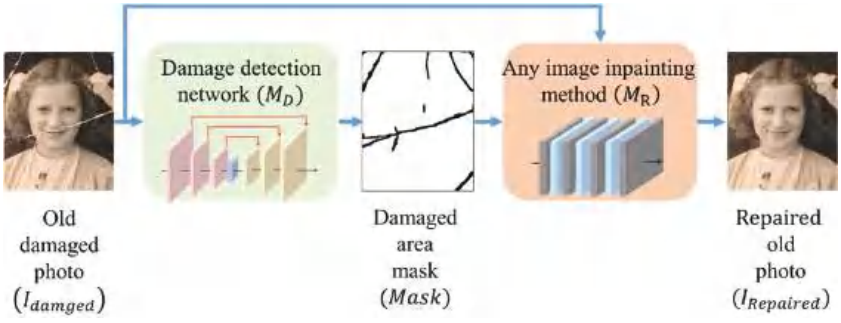


Figure 1. Flow chart of our architecture to automatically repair damaged old photos. By feeding an old damaged photo into our damage detection network, we can generate a damaged area mask. To restore the photo, the damaged photo and the mask are fed together into an arbitrary inpainting algorithm.

Figure 2 shows the architecture of our damaged detection model is derived from U-Net [21]. The first half is an encoder that extracts the image features, while the second half restores the image to its original size by up-sampling and uses the sigmoid function to find out the map of pixel damage probabilities. The advantage of using U-Net is its ability to capture features at different scales, which are important for old photo damage detection and allow the model to more accurately identify damage in different shapes and sizes. Another merit of U-Net is the ability to concatenate features of the encoder into the decoder, allowing the model to train without losing the features obtained in the shallow network.

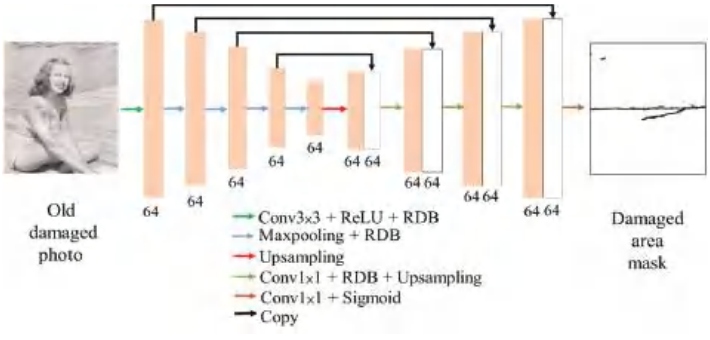


Figure 2. Architecture of the damage detection model.

In order to improve the ability to extract features, we replaced the original convolutional layers of U-Net with residual dense blocks (RDBs) [22]. This block is a combination of a residual block [23] and a dense block [24]. The residual block uses a skip connection to combine the input of the block with the output of the block, thus increasing the stability of the model training and the speed of convergence. The dense block continuously passes all the shallow features of the block to the deeper layers, thus making full use of the information from the shallow features. The RDB retains these advantages to improve the performance of the whole model. The original convolution layers at each scale used in U-Net would gradually lose its shallow feature information, but this problem was solved when we adopted RDB. In this way, it is possible to use more information from the area surrounding the damage for damage detection.

Since there is no open dataset of damaged old photos available for use, we collected photos from the Internet and marked the damaged areas in the images by ourselves. These photos consisted mainly of portraits, buildings, and natural scenery, with their sizes ranging from 129 $\times$ 317 to 797 $\times$ 1131 pixels. To generate ground truths, we manually marked the damaged areas of the collected photos using the image editing tool GIMP [1]. The transparency function of the GIMP layer feature makes marking damaged areas in photos easier and more precise. Figure 3 shows examples of photos from our collected dataset as well as the corresponding marked ground truth. We collected a total of 170 old damaged photos and manually labeled them, 123 of which were for the training set, 18 for the test set, and the remaining 29 for the validation set. On account of the limited number of photos in the data collection, the data augmentation technique was used to increase the dataset size via horizontal flipping and the 90-, 180-, and 270-degree rotation of photos.

(**a**) (**b**)

Figure 3. Dataset for damage detection: (**a**) old damaged photo; (**b**) corresponding marked ground truth.

Because there are more undamaged old photos on the Internet, in order to further extend the training dataset we collected and used these undamaged photos, along with a collection of damage-like textures, to synthesize artificial damaged photos. Compared to Figure 4a we can see some differences between the artificially damaged photo and the old real damaged photo. The real damaged area of an old photo is composed of complex

multitoned contents, not just simple good or bad, but our synthesized damaged photo only uses a single color to represent damage. We treated this difference as a type of damage to improve the generalizability of the model.

(**a**) (**b**)

Figure 4. Real damaged photos and damaged photos synthesized by texture mask: (**a**) real damaged photo; (**b**) our synthesized damaged photo.

The model parameters are initialized using the MSRA initialization method [25] in the experiments, and the optimizer is the Adam optimizer with β_1 =0.9, β_2 =0.999. The initial learning rate of the model is set to 0.0001, and every 1000 epochs are multiplied by 0.1 to train a total of 2000 epochs. The training patch size 48 × 48, which is commonly used in pavement crack detection, is less appropriate for our task. The main reason for this is that most cracked pavement images only have a black background and a few white cracks, whereas old cracked photos have more complex content, such as portraits, objects, buildings, and so on. Therefore, we partitioned the photos of the training set into patches of 100 × 100 pixels in size to account for more context to improve the performance, and in our experiment, larger patch sizes than this did not result in any additional performance gain. We also controlled the ratio of patches with damaged areas to patches without damage at 8:2 in training.

The loss function was balanced cross entropy. The main reason for employing balanced cross entropy was to compensate for the imbalance between intact and damaged areas. It modified the original binary cross entropy with the ratio of the two categories, giving more weight to the fewer damaged areas and less weight to the more numerous intact areas as shown in (3) where N is the total number of pixels in training blocks, α_i is the weight of the intact areas, y_i denotes whether the ith pixel belongs to the intact category in the ground truth, and $p(i)$ is the model's prediction of the probability that the ith pixel belongs to the intact areas.

$$L_{detection} = -\frac{1}{N}\sum_{i=1}^{N} \alpha_i \cdot y_i \cdot log(p(i)) + (1-\alpha_i) \cdot (1-y_i) \cdot log(1-p(i)), \quad (3)$$

3. Experiment Result

In our experiment, model training and testing were carried out on a computer equipped with an Intel i5-2400 CPU and an NVIDIA 2070 8GB GPU. To assess the model performance of damage detection, we adopted the evaluation methods commonly used in image segmentation and pavement crack segmentation, including precision, recall, F1-measure, and precision-recall curve (PR curve), as our evaluation metrics. Precision is the percentage of the results identified as damaged areas that are actually damaged. The percentage of true damaged areas detected is represented by recall. The F1 measure considers both precision and recall. Since the ground truth is created by manual marking and

each person has different damage marking criteria, we adopted the regional precision and recall proposed in [26], which considers the detection result correct as long as it is within five pixels of the manual marking results, to compensate for the ground truth credibility problem caused by manual marking errors.

3.1. Comparison of Various Modules

In this section, we first evaluate the performance of our damage detection model on old photos by testing the performance of U-Net barebones combined with various modules. We compared the results of our proposed method with three methods, including the original U-Net architecture, the U-Net architecture with a residual block module, and the U-Net architecture with a dense block module. Figure 5 and Table 1 show the results in terms of the PR curve, precision, recall, and F1-measure, which show that our proposed approach outperformed all other module combinations. Figure 6 depicts the visual outcome of using various modules to detect damage. It can be seen that our proposed method is capable of detecting more subtle damage as well as the damage border. The more complete the detection, particularly along the damage border, the more it can assist us in repairing damage without affecting the repair result.

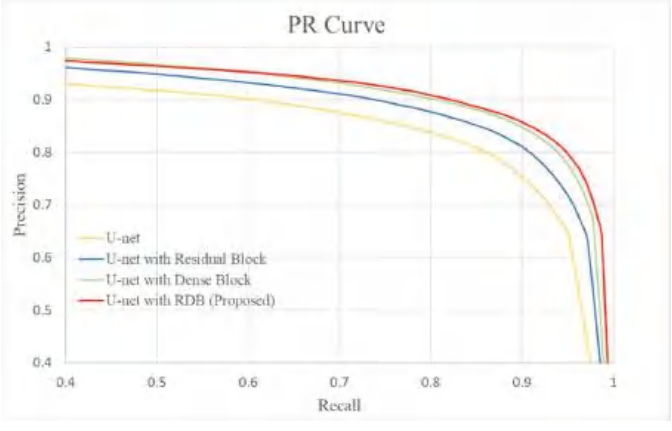


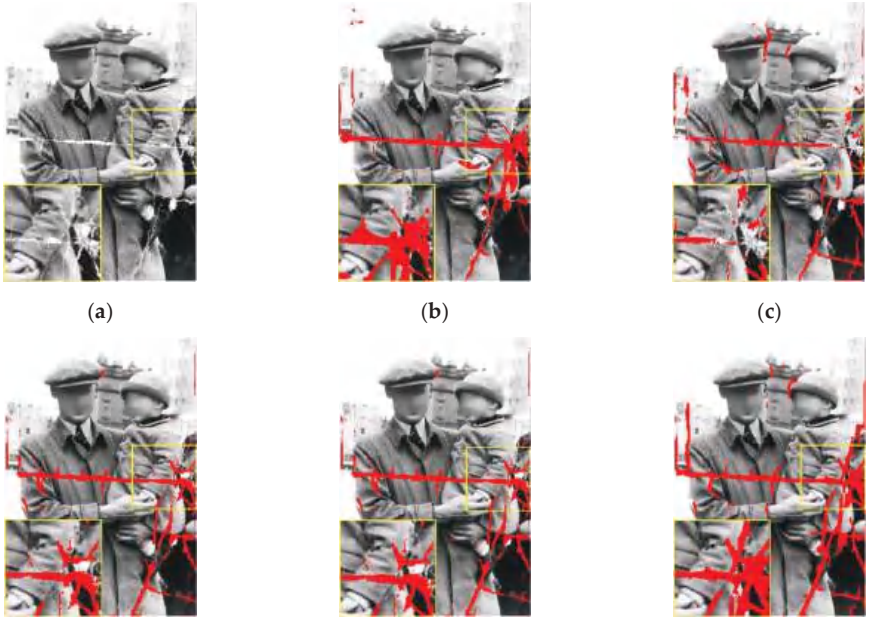
Figure 5. The PR curve of U-Net with various modules.

Table 1. The recall, precision, and F1 measure of different modules.

Structure	Recall	Precision	F1 Measure
U-Net	0.857	0.802	0.817
U-Net with residual block	0.876	0.833	0.846
U-Net with dense block	0.903	0.843	0.866
U-Net with RDB (proposed)	0.911	0.847	0.873

3.2. Comparison of Different Detection Methods

Next, we compare our method with other methods in the literature. We disassembled the damage detection part from the whole end-to-end work [5] and compared it to our method. Since there are so few existing deep learning-based damage detection methods for old photos, we also compared the results of pavement crack detection models [16,18,19] that have been retrained using our dataset to work on old photo damage detection. The results of the PR curve are shown in Figure 7. The best recall, precision, and F1 measure values for each method are shown in Table 2. The comparison results show that our detection effect is the best.



(**a**) (**b**) (**c**)

(**d**) (**e**) (**f**)

Figure 6. The detection results of different modules: (**a**) the old damaged photo; (**b**) labeled ground truth of damaged areas; (**c**) the detection result of U-Net; (**d**) the detection result of U-Net with residual block; (**e**) the detection result of U-Net with dense block; (**f**) our proposed detection result of U-Net with RDB.

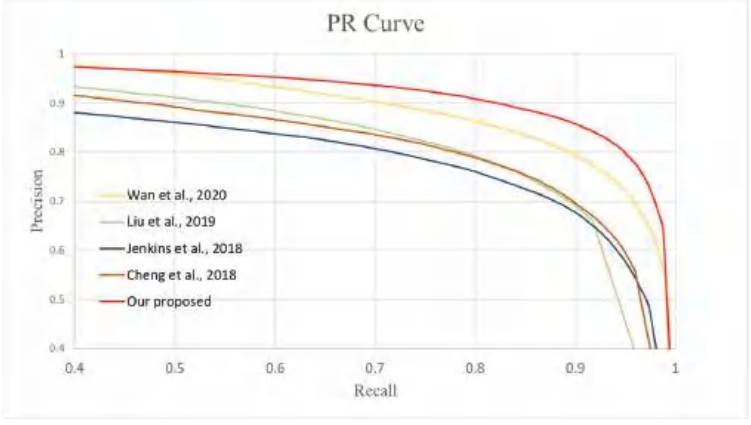
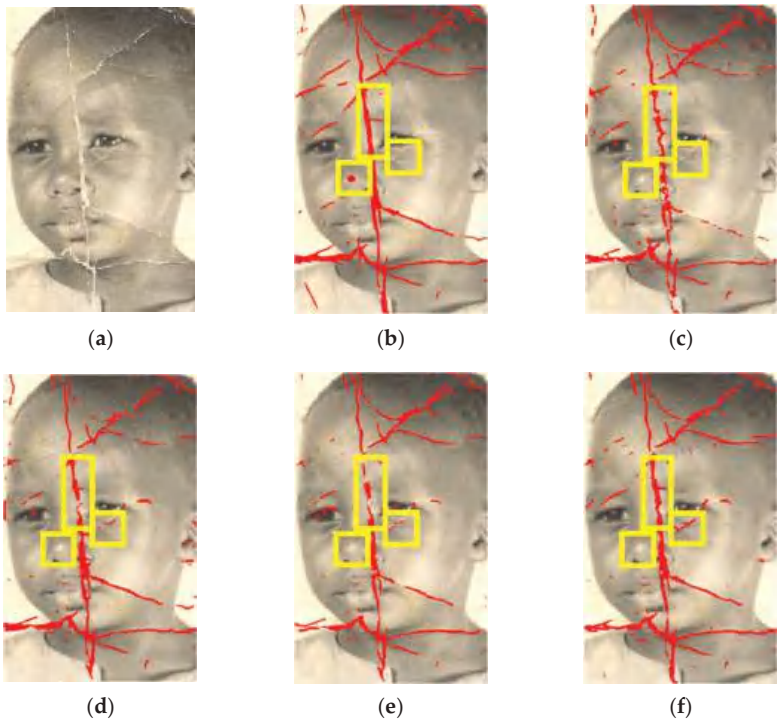


Figure 7. The PR curve of different methods, including [5,16,18,19], and our proposed method.

Table 2. The recall, precision, and F1 measure of different methods.

Method	Recall	Precision	F1 Measure
Wan et al. [5]	0.845	0.837	0.831
Liu et al. [16]	0.832	0.767	0.785
Jenkins et al. [18]	0.838	0.734	0.763
Cheng et al. [19]	0.839	0.763	0.784
Our proposed method	0.911	0.847	0.873

Figure 8 compares the visual results of the proposed method with those detected by other methods. It can be seen that our proposed method of detecting damage in the photo was more accurate, especially in the detection border denoted inside the yellow boxes. By contrast, the methods proposed by [16,18,19] failed to completely detect the damage in the image, and [5] often labeled undamaged areas as damage, such as around the tip of the nose in Figure 8.



(**a**) (**b**) (**c**)

(**d**) (**e**) (**f**)

Figure 8. The results of different detection methods, with yellow boxes indicating areas of performance difference: (**a**) the old damaged photo; (**b**) the result of Wan et al. [5]; (**c**) the result of Liu et al. [16]; (**d**) the result of Jenkins et al. [18]; (**e**) the result of Cheng et al. [19]; (**f**) the result of our proposed method.

As shown the Table 3, we also compared the number of parameters and computation speed with these methods [5,16,18,19] where the size of the test photos was 512 × 512. Jenkins et al. [18] and Cheng et al. [19] used the same model framework, but the model was trained using different strategies. Therefore, they have the same number of parameters and running time. Table 3 shows that both our detection models and those of [5] are fast as both lower to the scale of 10^{-3} s, but our model is much lighter as our number of parameters is only about one-sixteenth of all the other methods.

Table 3. Parameter and run time.

Method	Parameter	Computation Time (s)
Our proposed method	2.3 M	0.0084
Wan et al. [5]	37 M	0.0042
Liu et al. [16]	31.38 M	0.0122
Jenkins et al. [18]	33.24 M	0.0162
Cheng et al. [19]	33.24 M	0.0162

3.3. Combination with Inpainting Methods

Next, we present our results regarding practical application. We used [4,27,28] as the inpainting method in the subsequent process to repair actually damaged photos. The repair results using actually damaged photos are shown in Figures 9c–e and 10c–e, which demonstrate the results of our damage detection followed by different inpainting methods [4,27,28]. We can see in Figure 9c that Yu [27] failed to repair the cheeks and mouth in our detected area. Repair to damaged areas by gated convolution [4] is generally blurred as shown in Figure 9d. Figure 10c,d shows that deformation of the collar edge occurred after restoration. In general, the results of partial convolution [28] as shown in Figures 9e and 10e are more satisfactory compared to other inpainting methods [4,27]. This demonstrates that our architecture can be combined with any inpainting method, but we suggest that partial convolution [28] will achieve better results. In Figures 9b and 10b, we also compare our method with the end-to-end method [5], which integrates damage detection and repairs in one stage. Although [5] looks to have been effective in repairing the damaged areas, there are some color distortion problems with unfaithful tonal changes and a loss of texture in the image, such as in the cheeks as shown in Figure 9b. We can see that there are unrestored damaged areas and missing window frame details marked in the red box in Figure 10b. Thus, combining our architecture with the inpainting method [4,27,28] in contrast to [5] provides better results without affecting content in the undamaged regions in the recovery results.

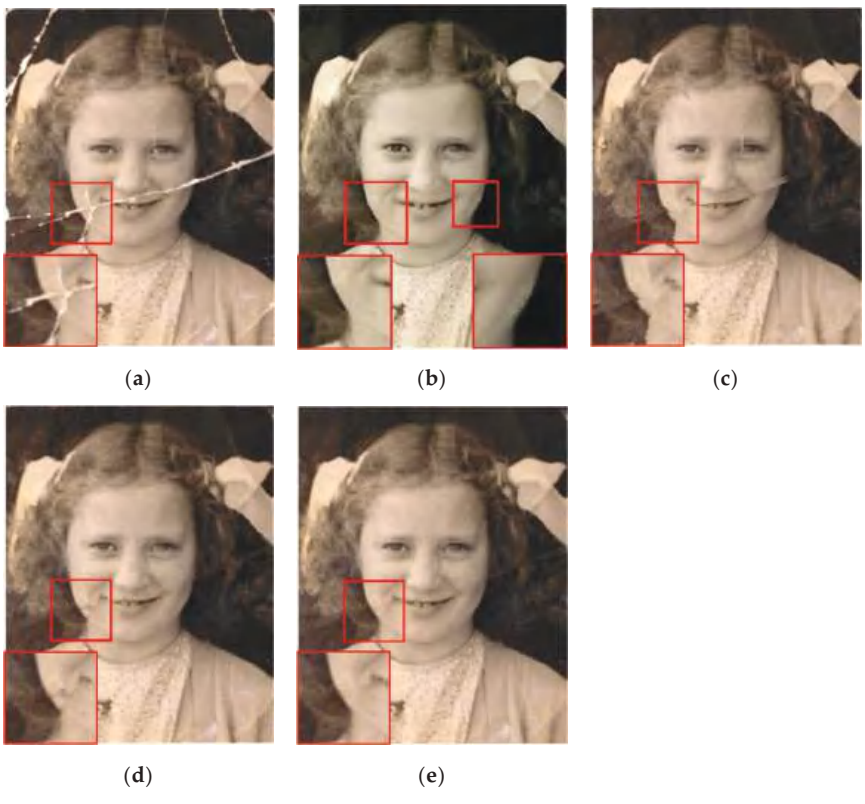
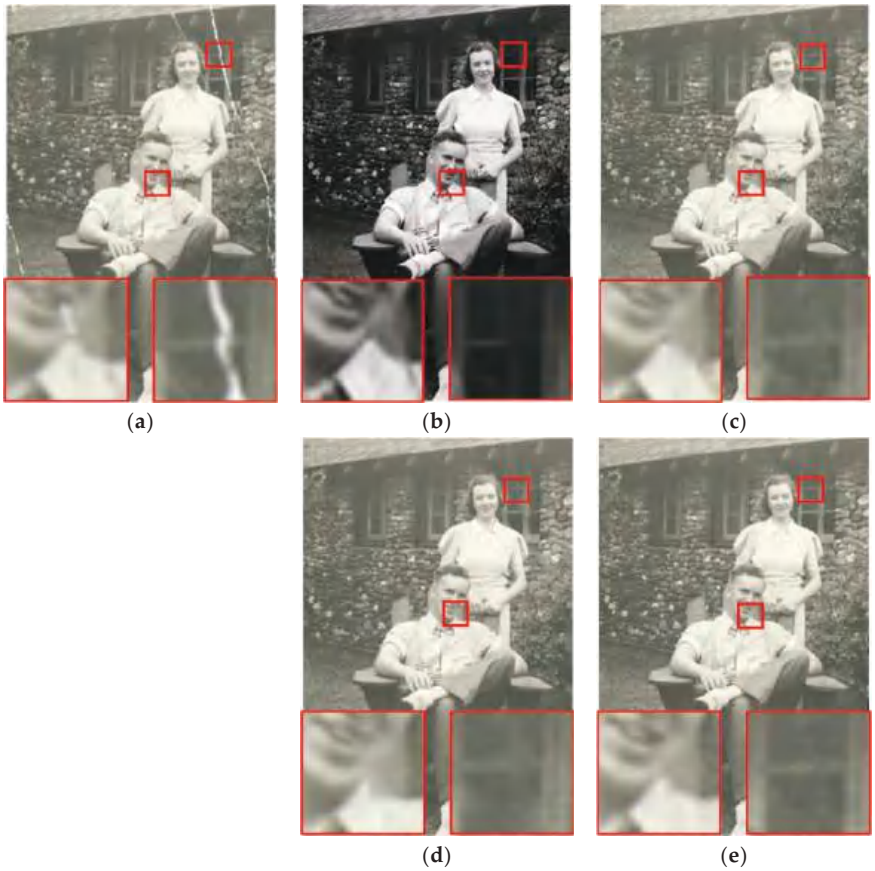


Figure 9. Results of different restoration methods on the damaged photo: (**a**) the old damaged photo; (**b**) the result of Wan et al. [5]; (**c**) the result of ours + Yu et al. [27]; (**d**) the result of ours + gated convolution [4]; (**e**) the result of ours + partial convolution [28].



(**a**) (**b**) (**c**) (**d**) (**e**)

Figure 10. Results for different restoration methods on the damaged photo: (**a**) the old damaged photo; (**b**) the result of Wan et al. [5]; (**c**) the result of ours + Yu et al. [27]; (**d**) the result of ours + gated convolution [4]; (**e**) the result of ours + partial convolution [28].

There will still be cases where our approach may fail. For example, if the model encounters a mixture of various complex damage, as shown in Figure 11, it becomes difficult to distinguish the damaged areas, resulting in partial detection and incomplete repair results. To deal with such a complex pattern of damage, future studies could investigate and apply the concept of directional clues in damage patterns [29–31] to aid in crack damage detection.

(**a**) (**b**) (**c**)

Figure 11. The case of failure detection: (**a**) damaged photo; (**b**) result of damage detection; (**c**) result of damage restoration.

4. Conclusions

Most restoration methods for damaged old photos require the manual marking of damaged areas for restoration, which is quite inefficient. Therefore, we proposed a damage detection model for old photos. Our method can detect damaged areas automatically without manual marking, which significantly reduces repair time. The detection results can be optionally screened and flexibly combined with any powerful inpainting method to fully automatically recover the content of the photos. We analyzed various block modules to design the detection model and found that the residual dense block (RDB), which combines the advantages of residual block and dense block, can effectively improve model detection capability. When compared to other detection algorithms, our method can detect damaged areas more accurately. We demonstrated the restoration of damaged old photos by combining our detection results with three different inpainting methods. In our restoration results, both the damaged and undamaged areas of the photos did not suffer from color tone changes, color distortion, or texture loss. Our method can better preserve the integrity of photos than can the existing end-to-end method, which alters the undamaged areas of photos.

Author Contributions: Conceptualization, T.-Y.K.; Data curation, Y.-J.W.; Investigation, Y.-J.W.; Methodology, T.-Y.K.; Resources, T.-Y.K.; Software, T.-H.L.; Supervision, T.-Y.K. and P.-C.S.; Validation, Y.-J.W. and T.-H.L.; Writing—original draft, Y.-J.W.; Writing—review & editing, T.-Y.K. and P.-C.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Ministry of Science and Technology under grant number MOST 111-2221-E-027-065-.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Some damaged source images were obtained from https://commons.wikimedia.org/wiki/Category:Damaged_photographs#/media/File:1945BunnyLakeTeeth.jpg under CC BY-SA 2.0 license, as well as from https://www.flickr.com/photos/simpleinsomnia/2529343285 4/in/photostream/ under CC BY 2.0 license.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Graphical Design Team. GIMP. Available online: https://www.gimp.org/ (accessed on 20 August 2022).
2. Li, B.; Qi, Y.; Shen, X. An Image Inpainting Method. In Proceedings of the Ninth International Conference on Computer Aided Design and Computer Graphics (CAD-CG'05), Hong Kong, China, 7–10 December 2005; p. 6.
3. Zhao, Y.; Po, L.-M.; Lin, T.; Wang, X.; Liu, K.; Zhang, Y.; Yu, W.-Y.; Xian, P.; Xiong, J. Legacy Photo Editing with Learned Noise prior. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Online, 5–9 January 2021; pp. 2103–2112.
4. Yu, J.; Lin, Z.; Yang, J.; Shen, X.; Lu, X.; Huang, T.S. Free-Form Image Inpainting with Gated Convolution. In Proceedings of the Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 4471–4480.
5. Wan, Z.; Zhang, B.; Chen, D.; Zhang, P.; Chen, D.; Liao, J.; Wen, F. Bringing Old Photos Back to Life. In Proceedings of the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Online, 14–19 June 2020; pp. 2747–2757.
6. Liu, J.; Chen, R.; An, S.; Zhang, H. CG-GAN: Class-Attribute Guided Generative Adversarial Network for Old Photo Restoration. In Proceedings of the Proceedings of the 29th ACM International Conference on Multimedia, Chengdu, China, 20–24 October 2021; pp. 5391–5399.
7. Kingma, D.P.; Welling, M. Auto-encoding variational bayes. *arXiv* **2013**, arXiv:1312.6114.
8. Dong, C.; Li, L.; Yan, J.; Zhang, Z.; Pan, H.; Catbas, F.N. Pixel-level fatigue crack segmentation in large-scale images of steel structures using an encoder–decoder network. *Sensors* **2021**, *21*, 4135. [CrossRef] [PubMed]
9. Jaidilert, S.; Farooque, G. Crack Detection and Images Inpainting Method for Thai Mural Painting Images. In Proceedings of the 2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC), Chongqing, China, 27–29 June 2018; pp. 143–148.
10. Bhuvaneswari, S.; Subashini, T. Automatic scratch detection and inpainting. In Proceedings of the 2015 IEEE 9th International Conference on Intelligent Systems and Control (ISCO), Coimbatore, India, 9–10 January 2015; pp. 1–6.

11. Ghosh, S.; Saha, R. A simple and robust algorithm for the detection of multidirectional scratch from digital images. In Proceedings of the 2015 Eighth International Conference on Advances in Pattern Recognition (ICAPR), Kolkata, India, 4–7 January 2015; pp. 1–6.
12. Cornelis, B.; Ružić, T.; Gezels, E.; Dooms, A.; Pižurica, A.; Platiša, L.; Cornelis, J.; Martens, M.; De Mey, M.; Daubechies, I. Crack detection and inpainting for virtual restoration of paintings: The case of the Ghent Altarpiece. *Signal Process.* **2013**, *93*, 605–619. [CrossRef]
13. König, J.; Jenkins, M.D.; Barrie, P.; Mannion, M.; Morison, G. A convolutional Neural Network for Pavement Surface Crack Segmentation Using Residual Connections and Attention Gating. In Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 22–25 September 2019; pp. 1460–1464.
14. Yang, F.; Zhang, L.; Yu, S.; Prokhorov, D.; Mei, X.; Ling, H. Feature pyramid and hierarchical boosting network for pavement crack detection. *IEEE Trans. Intell. Transp. Syst.* **2019**, *21*, 1525–1535. [CrossRef]
15. Lau, S.L.; Wang, X.; Xu, Y.; Chong, E.K. Automated Pavement Crack Segmentation Using Fully Convolutional U-Net with a Pretrained ResNet-34 Encoder. *arXiv* **2020**, arXiv:2001.01912.
16. Liu, W.; Huang, Y.; Li, Y.; Chen, Q. FPCNet: Fast pavement crack detection network based on encoder-decoder architecture. *arXiv* **2019**, arXiv:1907.02248.
17. Zhang, K.; Zhang, Y.; Cheng, H.-D. CrackGAN: A Labor-Light Crack Detection Approach Using Industrial Pavement Images Based on Generative Adversarial Learning. *arXiv* **2019**, arXiv:1909.08216.
18. Jenkins, M.D.; Carr, T.A.; Iglesias, M.I.; Buggy, T.; Morison, G. A Deep Convolutional Neural Network for Semantic Pixel-Wise Segmentation of Road and Pavement Surface Cracks. In Proceedings of the 2018 26th European Signal Processing Conference (EUSIPCO), Rome, Italy, 3–7 September 2018; pp. 2120–2124.
19. Cheng, J.; Xiong, W.; Chen, W.; Gu, Y.; Li, Y. Pixel-level Crack Detection using U-Net. In Proceedings of the TENCON 2018-2018 IEEE Region 10 Conference, Jeju, Korea, 28–31 October 2018; pp. 0462–0466.
20. Adams, R.; Bischof, L. Seeded region growing. *IEEE Trans. Pattern Anal. Mach. Intell.* **1994**, *16*, 641–647. [CrossRef]
21. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
22. Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.; Fu, Y. Residual dense network for image restoration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *43*, 2480–2495. [CrossRef] [PubMed]
23. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778.
24. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
25. Long, J.; Shelhamer, E.; Darrell, T. Fully Convolutional Networks for Semantic Segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
26. Shi, Y.; Cui, L.; Qi, Z.; Meng, F.; Chen, Z. Automatic road crack detection using random structured forests. *IEEE Trans. Intell. Transp. Syst.* **2016**, *17*, 3434–3445. [CrossRef]
27. Yu, J.; Lin, Z.; Yang, J.; Shen, X.; Lu, X.; Huang, T.S. Generative Image Inpainting with Contextual Attention. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 5505–5514.
28. Liu, G.; Reda, F.A.; Shih, K.J.; Wang, T.-C.; Tao, A.; Catanzaro, B. Image Inpainting for Irregular Holes using Partial Convolutions. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 85–100.
29. Liu, L.; Ouyang, W.; Wang, X.; Fieguth, P.; Chen, J.; Liu, X.; Pietikäinen, M. Deep learning for generic object detection: A survey. *Int. J. Comput. Vis.* **2020**, *128*, 261–318. [CrossRef]
30. Li, S.; Zhang, Z.; Li, B.; Li, C. Multiscale rotated bounding box-based deep learning method for detecting ship targets in remote sensing images. *Sensors* **2018**, *18*, 2702. [CrossRef] [PubMed]
31. Yang, X.; Yan, J.; Liao, W.; Yang, X.; Tang, J.; He, T. Scrdet++: Detecting small, cluttered and rotated objects via instance-level feature denoising and rotation loss smoothing. *IEEE Trans. Pattern Anal. Mach. Intell.* 2022; *early access*.

sensors

MDPI

Article

Non-Local Temporal Difference Network for Temporal Action Detection

Yilong He [1,2,†], Xiao Han [1,2,†], Yong Zhong [1,2,*] and Lishun Wang [1,2]

[1] Chengdu Institute of Computer Application, Chinese Academy of Sciences, Chengdu 610081, China
[2] School of Computer Science and Technology, University of Chinese Academy of Sciences, Beijing 100049, China
* Correspondence: zhongyong@casit.com.cn
† These authors contributed equally to this work.

Abstract: As an important part of video understanding, temporal action detection (TAD) has wide application scenarios. It aims to simultaneously predict the boundary position and class label of every action instance in an untrimmed video. Most of the existing temporal action detection methods adopt a stacked convolutional block strategy to model long temporal structures. However, most of the information between adjacent frames is redundant, and distant information is weakened after multiple convolution operations. In addition, the durations of action instances vary widely, making it difficult for single-scale modeling to fit complex video structures. To address this issue, we propose a non-local temporal difference network (NTD), including a chunk convolution (CC) module, a multiple temporal coordination (MTC) module, and a temporal difference (TD) module. The TD module adaptively enhances the motion information and boundary features with temporal attention weights. The CC module evenly divides the input sequence into N chunks, using multiple independent convolution blocks to simultaneously extract features from neighboring chunks. Therefore, it realizes the information delivered from distant frames while avoiding trapping into the local convolution. The MTC module designs a cascade residual architecture, which realizes the multiscale temporal feature aggregation without introducing additional parameters. The NTD achieves a state-of-the-art performance on two large-scale datasets, 36.2% mAP@avg and 71.6% mAP@0.5 on ActivityNet-v1.3 and THUMOS-14, respectively.

Keywords: temporal action detection; deep learning; convolutional neural networks; computer vision; video understanding

Citation: He, Y.; Han, X.; Zhong, Y.; Wang, L. Non-Local Temporal Difference Network for Temporal Action Detection. *Sensors* **2022**, *22*, 8396. https://doi.org/10.3390/s22218396

Academic Editor: Manuel José Cabral dos Santos Reis

Received: 15 September 2022
Accepted: 26 October 2022
Published: 1 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

With the wide application of image content understanding technology and the rapid growth of video data, video content understanding has attracted the attention of both industry and academia. It has application requirements in many scenarios, such as security surveillance, precision medicine, and video audits. One of the pressing needs is understanding human action within videos. Previous work tackled it as a pure action recognition task. In recent years, action recognition technology has made great achievements. However, action recognition requires trimmed videos with only action instances. In actual scenarios, most video data are unlabeled. Temporal action detection (TAD) can predict temporal boundaries (start/end) as well as action categories in an untrimmed video. Therefore, as an upstream task of video content analysis, TAD has become one of the bottlenecks that needs to be broken through.

In addition to the content information, the video also has contextual relevance, which requires modeling long-range temporal structures. The usual practice is to stack 1D convolutions. However, the original video has the following characteristics: (1) the duration of different actions varies widely; (2) information redundancy between neighboring frames; and (3) the information is weakened in the long-distance delivery. How to design an

efficient network to model long-range temporal relationships while taking into account the above problems is the key to further improve the performance of a TAD task. In recent years, many works have tried to solve these problems, but most of them only consider one or two aspects.

To model long-range temporal dependencies, the commonly used methods are the stacked 1D temporal convolutions [1–3] and transformer [4–6]. However, limited by the kernel size, the former method can only capture the local scope context information, neither can learn the relationship between frames with distant temporal intervals, and it cannot establish the relationship between instances. Due to the redundant information of adjacent frames in a video, this method is prone to fall into local traps. With the success of transformers in object detection [7,8] and NLP [9,10], part work has migrated the self-attention mechanism to the temporal action detection task. The attention mechanism can learn the relationship between each frame and other frames one-to-one, avoiding the distance limitation of a 1D convolution. Because the length of the action instances is much smaller than the length of the video, such methods not only perform a lot of invalid computations but may also introduce irrelevant information. Therefore, neither the global nor the local scope can effectively model complex temporal dependencies. To solve this problem, we propose a chunk convolution (CC). Specifically, each chunk consists of three independent, traditional 1D temporal convolutions with fixed intervals, which not only enlarges the temporal receptive field but also alleviates the redundancy problem. In addition, a multi-branch strategy is adopted, where each branch handles a specific redundancy rate to be compatible with redundancy rate changes.

Similar to object detection, temporal action detection also belongs to the category of visual detection, which is to locate and classify potential objects. Object detection aims to generate bounding boxes in an image (2D), while temporal action detection aims to predict the boundary locations of action instances in a temporal sequence (1D). Therefore, most of the current methods for processing a temporal multiscale are migrated from an image multiscale. Considering that it is difficult to find a specific receiving field that balances all scales, TAL-Net [11], A2net [12], and DCAN [13] borrow the idea of an anchor in object detection, which consists of K-convolution blocks with parallel structures. Each block has a different kernel size, corresponding to a different temporal receiver field. The responses of all blocks are fused to provide finer-grained features. Due to the unsatisfactory effect of a large-size convolution kernel, such methods are not scalable enough. Inspired by res2net [14] and Xception [15], we propose a cascade residual architecture to process the temporal multiscale issue. Specifically, the module consists of several parallel branches, and each branch contains two 1D convolutions with kernel sizes of 1 and 3, respectively. Except for the first branch, the output features of the former branch are added with the input features as the input of this branch. Each time features pass through a branch, its temporal receptive field will expand once. Finally, the features from all the branches are concatenated along the temporal dimension to aggregate the features with different temporal receptive fields.

Among all the features, motion information and boundary features undoubtedly play an important role in precise locating. Some work [16,17] uses an optical flow to represent the motion information. However, as the network deepens, the motion information will weaken over a long-range delivery. In order to address the issues, we propose a temporal difference (TD) module. Concretely, the temporal-level action confidences are firstly calculated across the full sequence, where the scores represent the attention weights. These weights are then used to produce motion-sensitive weights. Finally, we utilize multiplication between the original feature and motion-sensitive weights to enhance the discriminability of the features. In this way, the network has the ability to adaptively discover and enhance the features of motion-sensitive locations. We proposed an NTD network that achieves a new state-of-the-art performance on two large-scale datasets, ActivityNet-v1.3 [18] and THUMOS-14 [19].

2. Related Work

Temporal action detection aims to classify and localize action instances in an untrimmed video as precisely as possible. The existing approaches can be divided into three main types: anchor-based, anchor-free, and the bottom–up method.

Anchor-based methods rely on manually pre-defined K anchors with different scales. The early anchor mechanism is a window anchor. The S-CNN [20] uses sliding windows to generate multiple candidate regions and then uses a binary classification network to identify a possible action instance. The TURN [21] and CTAP [22] first generate multiscale candidate regions at each temporal position and then use temporal regression to refine boundary positions. The window mechanism can cover all action instances, thus avoiding missed detection. However, the disadvantages are also obvious, generating a large number of redundant regions and the boundaries are imprecise. Inspired by a faster-rcnn [23] in object detection, the R-C3D [24] predicts the relative offsets and corresponding classification scores of K-different scales at each temporal position. Considering that the duration of the action instance varies more dramatically than the target in an image, the TAL-Net [11] proposes to align the temporal receiving field of the anchor with the corresponding temporal span. Manually defining the scales limits the ability to handle complex variations. The GTAN [25] introduces nonlinear temporal modeling, cascades multiple feature maps with different temporal resolutions, and learns a Gaussian kernel for each temporal position to predict the relative offset. The PBRNet [26] cascades three detection modules; the first module generates coarse results and subsequent modules further the boundary position.

Inspired by the successful application of the anchor-free detector in object detection, many methods adopt the anchor-free method, which directly predicts the boundary position without manually specifying the proposal scale. The AFSD [27] proposes a purely anchor-free framework that directly predicts the distance of the boundary (start and end) from each temporal position. However, the predicted proposal relies heavily on local information and does not make full use of context relations. In order to model long-range context, some current works, such as the RTD-Net [5] and TadTR [28], regard video as a temporal sequence and introduce a self-attention transformer structure. Because using the attention mechanism in the whole sequence is inefficient and will introduce irrelevant noise interference, ActionFormer [4] proposed a local attention mechanism that limits the attention range within a fixed window. Considering the anchor base and anchor free have the advantages of stability and flexibility, respectively, the A2net [12] integrates these two methods into one framework to achieve complementary advantages.

Bottom–up methods mainly focus on evaluating "probabilities". The SSN [29] directly predicts the binary action probabilities for each frame in the video. Then, continuous frames with high action probabilities are grouped by the watershed algorithm to generate candidate proposals. The BSN [30], BMN [31], and BSN++ [32] predict the probabilities of being a start/end/action for each frame and then adopt a boundary-matching strategy to match pairs of start and end, generating candidate proposals with a flexible duration. These approaches fail to take full advantage of contextual information by focusing only on the confidence of isolated frames; this makes it sensitive to noise and prone to generating false positives and incomplete action instances. The BU-MR [33] exploits potential constraints between frame-level probabilities to provide more complementary information. The P-GCN [34] and G-TAD [35] take the proposals generated by the BSN as input and then use a graph convolution to explore the semantic relationships between proposals, providing more clues to facilitate boundary refinement.

3. Approach

3.1. Chunk Convolution

Given a sequence $X \in R^{C \times T}$, we evenly divide X into $N = T/(\omega + k)$ chunks, where k, C, and T denote kernel size, feature dimension, and temporal length, respectively, and ω is a manually set parameter used to adjust the chunk size. The j-th position of the i-th chunk can be represented as (i, j), where $i \in [1, N]$, $j \in [1, \omega + k]$. When extracting features at

position (i, j), the difference from traditional 1D convolution is that in addition to applying standard 1D convolution here, 1D convolution block is also applied to adjacent chunks $(i-1, j)$ and $(i+1, j)$, respectively. Three convolution blocks form a chunk convolution, and the outputs of all convolution blocks are fused by summation as the output of the chunk convolution. To facilitate implementation, we transform the temporal dimension from 1D to 2D, $X \in R^{C\times T} \rightarrow X' \in R^{C\times N\times(\omega+k)}$, $(\omega + k)$ represents the temporal length of each chunk, and N represents the number of chunks. In order to keep the feature dimension constant, the operation of padding 0 around X' is adopted, $X' \in R^{C\times N\times(\omega+k)}$ $\rightarrow X'' \in R^{C\times(N+2)\times(\omega+k+2)}$. Then, we apply 2D convolution to X'' as we did for extracting image features.

$$H = W * X'', \quad H \in R^{C\times N\times(\omega+k)} \tag{1}$$

where $*$ represents the convolution operation, $W \in R^{C\times C\times K\times K}$ is the convolution kernel. Subsequently, the temporal dimension of H is restored from 2D to 1D, $H \in R^{C\times N\times(\omega+k)} \rightarrow$ $Y_{CC} \in R^{C\times T}$, whose dimensions are consistent with the input features. Taking into account videos with various redundancy rates, we parallelize multiple branches with different chunk sizes to extract features simultaneously. In our experiments, the chunk size between d $\in \{4, 7, 9\}$, and the kernel size of all 1D convolutions is 3. It is easy to conclude that the dilate rates of the three branches are 1, 4, and 6, and the corresponding temporal receptive fields are 11, 17, and 21, respectively. The output of all branches is aggregated by max operation along the temporal dimension.

3.2. Multiple Temporal Coordination

A simple and effective strategy to extend the temporal receiving field is to stack multiple 1D convolutional layers. However, the duration of the action instances in the videos vary significantly. We adopt a split–transform–merge approach to deal with multiscale problems. As shown in Figure 1, given an input feature $Z \in R^{C\times T}$, we feed it into four branches with identical structure. Each branch is composed of a 1D convolution with kernel size 1, followed by a 1D convolution with kernel size 3. The relationship between adjacent branches is transformed from parallel to cascade through residual connections. Thus, the output can be expressed as:

$$F_i = W_3 * (W_1 * Z), \quad i = 1; \tag{2}$$

$$F_i = W_3 * (W_1 * (Z + F_{i-1})), \quad i = 2, 3, 4 \tag{3}$$

where $*$ represents the convolution operation, W_1 and W_3 denote the 1D convolution with kernel sizes of 1 and 3, respectively. Here, Z is the input features and F_{i-1} is the output features from the previous branch. The operation $Z + F_{i-1}, i \in [2, 4]$ is implemented by element-wise addition, where Z and F_{i-1} have equal dimensions. Each convolution operation is followed by a nonlinear activation function Relu, which is omitted for simplifying formula. The function W_1 is used to learn the residual mapping, and W_3 is used to expand the temporal receptive field.Obviously, after the feature passes through a branch, the temporal receptive field will be enlarged one time. Moreover, except for the first branch, each branch aggregates the feature information from former branch. Therefore, this module not only expands the temporal receptive field but also aggregates features of different receptive field. The output of this module is a multiscale temporal feature set $\{F_1, F_2, F_3, F_4\}$. Compared with the input feature Z, its temporal receptive field is enlarged by one, two, three, or four times, respectively. Finally, we adopted MAX operation along the temporal dimension on the set S to generate multiscale temporal features.

$$Y_{\mathrm{MTC}} = MAX([F_1, F_2, F_3, F_4]), \quad F_i \in R^{C\times T}, Y_{\mathrm{MTC}} \in R^{C\times T} \tag{4}$$

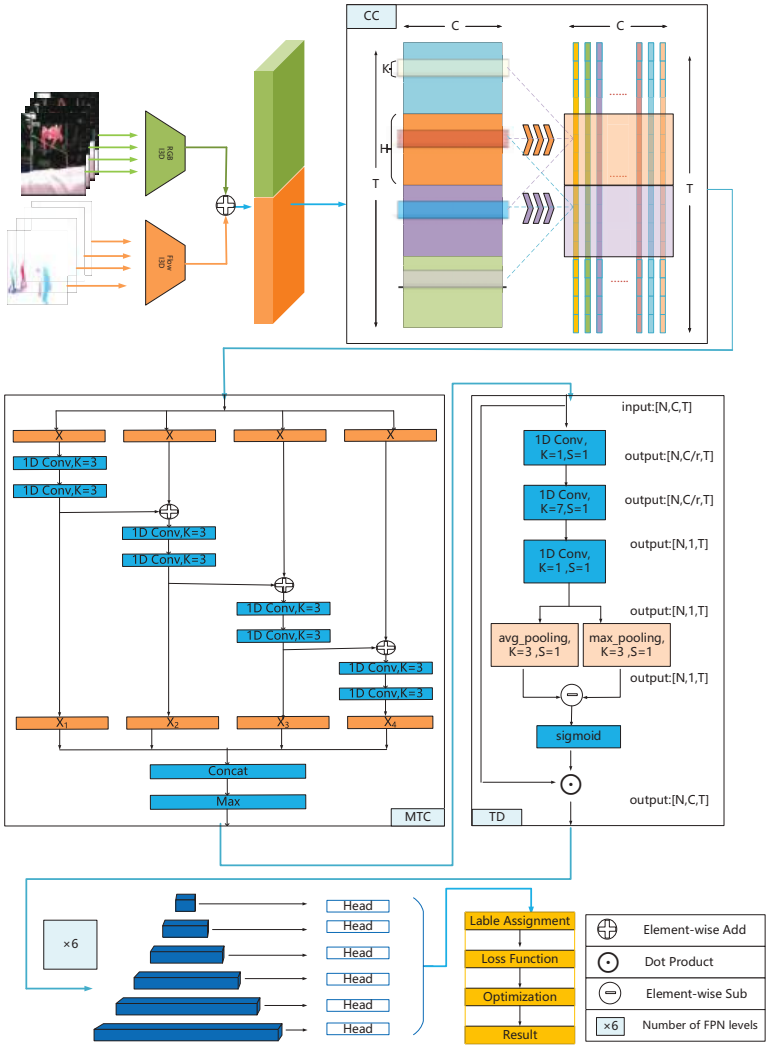


Figure 1. Overview of our proposed NTD. NTD first transforms the raw video into a sequence of clip features. Then, the clip features sequentially go through three modules. CC module, delivering information over long distances. MTC module, aggregating multiscale temporal features. TD module, adaptively enhancing motion information and boundary features. Finally, the encoded features pass through the temporal feature pyramid module, and the prediction head is responsible for generating detection results.

3.3. Temporal Difference

There is no doubt that among all features, boundary features and motion information play a particularly important role in the accurate localization of action instance. However, the boundary points of action instances are relatively sparse, and motion information delivered from distant frames are weakened. Therefore, in order to solve the problem of information weakening caused by stacking convolution layers, it is necessary to find motion-sensitive temporal locations, and then enhance its features. We will implement it in two steps, squeeze and excitation. As depicted in Figure 1, given an input sequence $V \in R^{C \times T}$, in order to generate the temporal attention weights, we first consider using squeeze-channel operations. Specifically, stacking three 1D convolution layers, transform

the feature dimension from $V \in R^{C\times T}$ to $G \in R^{1\times T}$. The first convolutional layer is used to reduce channel dimensions from C to C/r, r is set 4 in our work. Aims to capture long-range information, we followed a 1D convolution layer to extend receiving field, which set the kernel size as 7 and step size is 1 in our paper. The last 1D convolutional layer squeezes the channel dimensions into one. In addition, each convolutional layer is followed by an activation function Relu. In this way, we obtain temporal attention weights $G \in R^{1\times T}$ for each temporal position.

$$G' = \sigma(conv1 * V), \quad G' \in R^{C/r\times T} \tag{5}$$

$$G'' = \sigma(conv7 * G'), \quad G'' \in R^{C/r\times T} \tag{6}$$

$$G = \sigma(conv1 * G''), \quad G \in R^{1\times T} \tag{7}$$

where $*$ denotes convolution operation, σ refers to the Relu function, $conv1$ and $conv7$ indicate the 1D convolution whose kernel size is 1 and 7, respectively. We follow the squeeze operation with an excitation operation which aims to salient the boundary features. In practice, the attention weights between adjacent temporal location vary significantly, and this location can be approximately confirmed as a motion-sensitive position. It is achieved by feeding the attention weights into two independent branches, average pooling and max pooling, respectively. Among them, the difference calculation of the two branches can be formulated as:

$$S = \delta(MAX(G) - AVG(G)), \quad S \in R^{1\times T} \tag{8}$$

Here, δ represents the activation function Sigmoid, MAX and AVG denote max pooling and average pooling along the temporal dimension, respectively. Finally, the purpose of this module is to enhance boundary features and motion information; a straightforward way is to rescale features $V \in R^{C\times T}$ with the attention weights $S \in R^{1\times T}$.

$$Y = V@S, \quad Y \in R^{C\times T} \tag{9}$$

where @ refers to temporal-wise multiplication between the scalar S and feature V.

4. Training and Inference

4.1. Training

Before the encoded features are fed into the prediction layer, it goes through a six-level temporal feature pyramid module to be compatible with multiscale action instances. The hierarchical architecture is responsible for generating feature set M = $\{m_1, m_2, \cdots, m_6\}$ with varying temporal resolutions. Precisely, we adopt the 1D convolution with stride s = 2 (except for the first level) to decrease the temporal length of each level. Our prediction layer consists of two independent lightweight convolutional networks for classification and regression, and both branches are implemented by three consecutive 1D convolutional with kernel size 3. Our network outputs the predicted result $y_t = (p_t^i, d_t^s, d_t^e)$ for every moment t across all pyramid levels. $p_t^i \in \{0,1\}_1^c$ is the probability of action categories (c pre-defined categories). $d_t^e \geqslant 0$ and $d_t^s \geqslant 0$ are the distance from the current moment t to boundary. d_t^s and d_t^e are valid if time t falls within the range of any action instances; otherwise, they are not counted as loss. In practice, the proportion of the background in the video is much higher than that of the foreground. To alleviate the imbalance, we adopt focal loss [36] as our classification loss function. According to the predicted classification score $p_t = (s_0, s_1, \cdots, s_c)$, the total classification loss can be calculated using the following formula:

$$\mathcal{L}_{cls} = \frac{1}{T}\sum_{t=0}^{T}\sum_{i=0}^{c} -\alpha_t^i \left(1 - p_t^i\right)^{\gamma} \log\left(p_t^i\right) \tag{10}$$

$$p_t^i = \begin{cases} s_i & \text{if } y_t^i = 1, \ \ i = 0, 1, 2, \cdots, c \\ 1 - s_i & \text{otherwise, } \ i = 0, 1, 2, \cdots, c \end{cases} \tag{11}$$

$$\alpha_t^i = \begin{cases} \alpha & \text{if } y_t^i = 1, \ \ i = 0, 1, 2, \cdots, c \\ 1 - \alpha & \text{otherwise, } \ i = 0, 1, 2, \cdots, c \end{cases} \tag{12}$$

In the above, class label $y_t \in R^c$ (c pre-defined categories), $y_t^i \in \{0, 1\}$. γ and α are manually specified hyper-parameters, which are set to 2 and 0.25, respectively, in our paper. We use temporal Intersection over Union (tIoU) as the loss function for regression of the distance between the predicted instance $\hat{\phi}_i = (\hat{\psi}_i, \hat{\xi}_i)$ and the corresponding ground truth $\phi_i = (\psi_i, \xi_i)$, and only the foreground moment that falls into an action instance is selected. The formula can be expressed as:

$$\mathcal{L}_{reg} = \frac{1}{T_p} \sum_i \mathbb{I}(y_i \geq 1) \left(1 - \frac{|\hat{\phi}_i \cap \phi_i|}{|\hat{\phi}_i \cup \phi_i|}\right) \tag{13}$$

where T_P represents the number of foreground moments, the indicator function $\mathbb{I}(y_i \geqslant 1)$ is used to indicate whether the temporal location $t \in [1, N]$ falls within the range of any ground truth.The total loss function employed during training is defined as follows:

$$\mathcal{L} = \mathcal{L}_{cls} + \mathcal{L}_{reg} \tag{14}$$

4.2. Inference

At inference stage, we directly input the feature sequences into the network. Our prediction layer outputs the predicted result $y_t = (p_t^i, d_t^s, d_t^e)$ for every moment t across all pyramid levels. For each moment t in the j-th pyramid level, the predicted action instance is denoted as

$$a_j^t = \arg\max\left(p_t^i\right), \quad s_j^t = t - d_t^s, \quad e_j^t = t + d_t^e \tag{15}$$

where s_j^t and e_j^t are the left and right boundaries of an action instance, and a_j^t is the category score. Next, in order to remove the highly overlapping action instances, we aggregate the candidate action instances from all positions together, perform Soft-NMS [37], and obtain the final result.

5. Experiments

5.1. Datasets

We conduct the experiment on two popular benchmark datasets, ActivityNet-v1.3 [18] and THUMOS-14 [19], for the TAD task. THUMOS-14 collected videos of human daily activities, including 20 categories. The training and testing set contain 200 and 212 untrimmed videos, respectively. The average temporal length of the videos in the dataset is 4.4 min, each video contains more than 15 action instances, each instance has an average duration of 5 s, and more than 70% of the moments belong to the background. These action instances are densely distributed and disordered within the video, making it extremely challenging to perform TAD on this dataset. ActivityNet-v1.3 contains around 10 K, 5 K, and 5 K videos in the training, testing, and validation sets. As a larger dataset with 200 action categories, the average temporal length is 2 min, each video contains an average of 1.7 action instances, and each instance has an average duration of 48 s.

5.2. Evaluation Metrics

To compare with previous TAD methods, we adopt mean average precision (mAP) to evaluate our NTD network on both datasets. On THUMOS-14, the temporal Intersection-over-Union (tIoU) thresholds are selected from {0.3, 0.4, 0.5, 0.6, 0.7}. On Activitynet-v1.3, the tIoU thresholds are chosen from {0.5, 0.75, 0.95}. According to the official evaluation metrics, THUMOS-14 pays more attention to the performance on mAP@0.5, and Activitynet-v1.3 focuses on the results on mAP@avg [0.5:0.05:0.95].

5.3. Feature Extraction and Implementation Details

On THUMOS14, following [4,12], we adopt two-stream inflated 3D ConvNet (I3D [38]) module which pre-trained on Kinetics-400 [39] to extract spatial-temporal features from raw video. We sample 16 consecutive RGB and optical flow with the overlap rate of 75% as clips. Then, feed clip into I3D network and extract features of dimension 1024×2 at the first fully connected layer. Finally, the two-stream features are concatenated along the temporal dimension ($1024D \times 2 \Longrightarrow 2048D$). We use Adam [40] to optimize the network, setting the batch size, initial learning rate, total epoch number as 2, 1×10^{-4}, 35, respectively. We visualize four instances predicted by our model on this dataset and compare them with the corresponding ground truth in Figure 2.

On Activitynet v1.3, following [4,12], we use R(2+1)D [41] pre-trained on TSP [42] to extract features. We sample 16 consecutive RGB with non-overlapping as clips. We use linear interpolation to rescale the feature sequence to a fixed length of 128. We use Adam to optimize the network, setting the batch size, initial learning rate, total epoch number as 16, 1×10^{-3}, 15, respectively.

Our model is implemented based on PyTorch 1.1, Python 3.8, and CUDA 11.6. We conduct experiments with one NVIDIA GeForce RTX 3090 GPU, Intel i5-10400 CPU and 128 G memory.

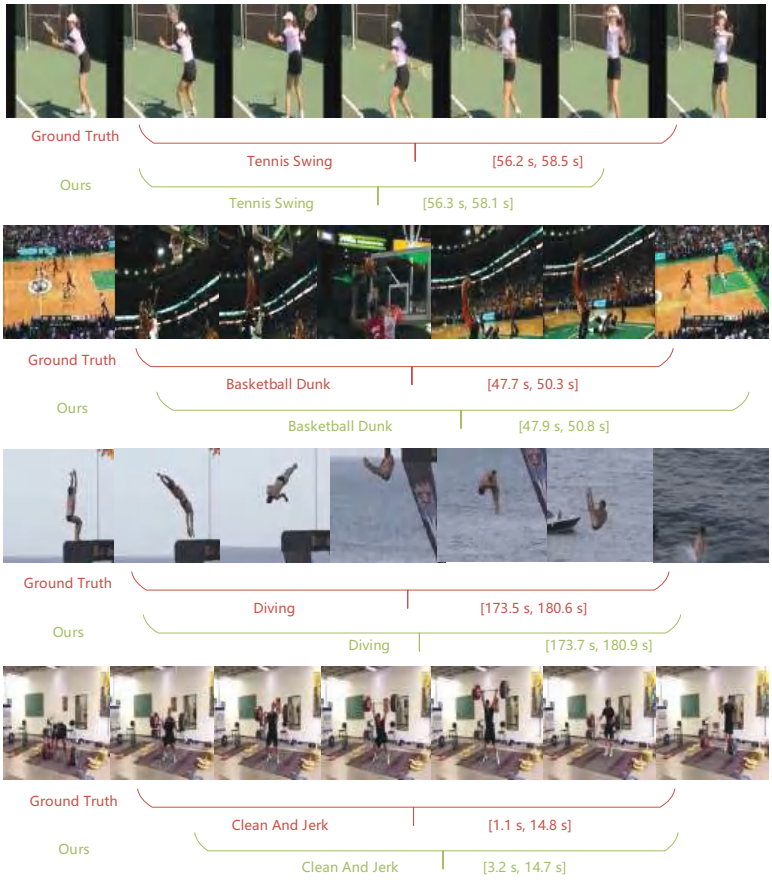


Figure 2. Qualitative Results. Visualize ground truth and corresponding predicted instances on THUMOS-14.

6. Results

6.1. Comparison with State-of-the-Art Methods

We compare the NTD with several state-of-the-art temporal action detection methods on the THUMOS-14 dataset. As shown in Table 1, the performances at different tIoU thresholds (mAP@tIoU) vary from 0.3 to 0.7 as well as an average mAP [0.3:0.1:0.7] (mAP@avg). In comparison, our proposed NTD outperforms the other methods at all thresholds. In particular, under the official evaluation index mAP@0.5, our method achieves 71.6%, exceeding the concurrent state-of-the-art work of ActionFormer [4] by a large margin 6.0% (71.6% vs. 65.6%). We also achieve a state-of-the-art performance with an average mAP of 66.8% ([0.3:0.1:0.7]).

On the ActivityNet-v1.3 database, our model also achieves a competitive result, significantly outperforming the recent representative works, the AES [43], ActionFormer [4], BCNet [44], and RCL [1]; the performers are shown in Table 2. Our model achieves a 54.4% mAP@0.5, outperforming all of the previous methods. With an average mAP ([0.5:0.05:0.95]), our method reaches 36.3% which is 0.7% higher than the recent state of the art 35.6% by ActionFormer. This improvement is significant because the results are averaged over many tIoU thresholds, including those that are tight, such as tIoU = 0.95.

Table 1. Comparison with state of the art (THUMOS-14). We report the precision at different tIoU thresholds (mAP@tIoU) as well as average mAP in [0.3:0.1:0.7] (mAP@avg). The best results are in bold.

Method	Year	Backbone	0.3	0.4	0.5	0.6	0.7	AVG
S-CNN [20]	CVPR-2016	DTF	36.3	28.7	19.0	10.3	5.3	19.9
TURN [21]	ICCV-2017	Flow	44.1	34.9	25.6	-	-	-
R-C3D [24]	ICCV-2017	C3D	44.8	35.6	28.9	-	-	-
BSN [30]	ECCV-2018	TSN	53.5	45.0	36.9	28.4	20.0	36.8
TAL-Net [11]	CVPR-2018	I3D	53.2	48.5	42.8	33.8	20.8	39.8
GTAN [25]	CVPR-2019	P3D	57.8	47.2	38.8	-	-	-
P-GCN [34]	ICCV-2019	TSN	60.1	54.3	45.5	33.5	19.8	42.6
BMN [31]	ICCV-2019	TSN	56.0	47.4	38.8	29.7	20.5	36.8
A2Net [12]	TIP-2020	I3D	58.6	54.1	45.5	32.5	17.2	41.6
G-TAD [35]	CVPR-2020	TSN	54.5	47.6	40.2	30.8	23.4	39.3
BU-MR [33]	ECCV-2020	TSN	53.9	50.7	45.4	38.0	28.5	43.3
VSGN [45]	ICCV-2021	TSN	66.7	60.4	52.4	41.0	30.4	50.2
CSA [46]	ICCV-2021	TSN	64.4	58.0	49.2	38.2	27.8	47.5
AFSD [27]	CVPR-2021	I3D	67.3	62.4	55.5	43.7	31.1	52.0
MUSES [47]	ICCV-2021	I3D	68.3	63.8	54.3	41.8	26.2	50.9
RefactorNet [16]	CVPR-2022	I3D	70.7	65.4	58.6	47.0	32.1	54.8
ActionFormer [4]	2022	I3D	75.5	72.5	65.6	56.6	42.7	62.6
RCL [1]	CVPR-2022	TSN	70.1	62.3	52.9	42.7	30.7	51.7
AES [43]	CVPR-2022	SF R50	69.4	64.3	56.0	46.4	34.9	54.2
BCNet [44]	AAAI-2022	I3D	71.5	67.0	60.0	48.9	33.0	56.1
NTD (Ours)		I3D	**82.7**	**78.7**	**71.6**	**58.3**	**42.8**	**66.8**

The excellent performance demonstrates the effectiveness and generalizability of our proposed method for the TAL. This indicates that modeling long-range temporal context dependence while taking into account multiscale and motion information enhancement can improve the ability of the network to model complex video structures.

Table 2. Comparison with state of the art (ActivityNet-1.3). We report the precision at tIoU = 0.5, 0.75, and 0.95 (mAP@tIoU) as well as average mAP in [0.5:0.05:0.95] (mAP@avg). The best results are in bold.

Method	Year	0.5	0.75	0.95	AVG
TAL-Net [11]	CVPR-2018	38.2	18.3	1.3	20.2
BSN [30]	ECCV-2018	46.5	30.0	8.0	30.0
GTAN [25]	CVPR-2019	52.6	34.1	8.9	34.3
BMN [31]	ICCV-2019	50.1	34.8	8.3	33.9
BC-GNN [48]	ECCV-2020	50.6	34.8	9.4	34.3
G-TAD [35]	CVPR-2020	50.4	34.6	9.0	34.1
TCANet [49]	CVPR-2021	52.3	36.7	6.9	35.5
BSN++ [32]	AAAI-2021	51.3	35.7	8.3	34.9
MUSES [47]	CVPR-2021	50.0	35.0	6.6	34.0
ActionFormer [4]	2022	53.5	36.2	8.2	35.6
BCNet [44]	AAAI-2022	53.2	36.2	**10.6**	35.5
AES [43]	CVPR-2022	50.1	35.8	10.5	35.1
RCL [1]	CVPR-2022	51.7	35.3	8.0	34.4
NTD (Ours)		**54.4**	**37.4**	8.2	**36.2**

6.2. Ablation Study of MTC Module

In response to the problem that the duration of different actions varies widely, we design the multiple temporal coordination (MTC) module. In our experiments, different numbers of branches were tried, and the results are listed in Table 3. By comparing the first to sixth rows of the table, we can observe that with the increase in branches, the performance continues to improve. The best results, 71.6% mAP@0.5 and 66.8% mAP@avg, were obtained when the number of branches is four. However, the fifth and sixth rows show that using more branches does not achieve a better performance. In the last two rows, we also show the effectiveness of different feature fusion strategies between branches. By comparison, the MAX operation works best. This benefits from the feature selectivity of the MAX function, which improves the saliency of the feature maps within the regions. With each additional branch, the equivalent temporal receptive field will be enlarged one time. A multi-branch cross-scale association is beneficial to capture the multiscale feature information, but the scale span is significantly different, which will affect the stability of the module to capture local features.

Table 3. Ablation Study (impact of MTC module). Comparing the effects of different branch numbers and fusion strategies between branches on THUMOS-14, measured by mAP@tIoU at different thresholds and the average mAP (mAP@avg) [0.3:0.1:0.7].

Number	Strategy	0.3	0.4	0.5	0.6	0.7	AVG
6	MAX	81.3	77.3	69.9	59.4	43.2	66.2
5	MAX	81.2	77.3	70.0	58.2	44.4	66.2
4	MAX	82.7	78.7	71.6	58.3	42.8	66.8
3	MAX	81.7	77.8	71.2	60.0	44.9	67.1
2	MAX	81.6	78.0	70.5	57.7	43.2	66.2
1	MAX	81.0	77.0	69.4	59.0	43.8	66.1
4	AVG	81.9	77.9	69.8	57.2	42.7	65.9
4	Conv1D	81.8	77.3	70.1	58.9	44.4	66.5

6.3. Ablation Study of TD Module

We study the effects of a temporal receptive field (convolution kernel size) for the TD module on THUMOS-14, as shown in Table 4. Comparing the first to third rows, it shows that the TD benefits more from larger kernel sizes (K = 7 vs. K = 3). However, as the convolution kernel continues to expand, the mAP@0.5 drops by more than 1%. We

also compared the effect of the max pooling size on the results when computing attention weights, and the results show that a smaller size (S = 3) performs better. Larger convolution kernels mean that the TD can capture the contextual information in a longer temporal range, thus mitigating random noise interference. However, an excessively large convolution kernel will smooth the difference between neighboring features.

Table 4. Ablation Study (impact of TD module). Comparing the effects of convolution kernel size (K) and max pooling size (S) on THUMOS-14, measured by mAP@tIoU at different thresholds and the average mAP (mAP@avg) [0.3:0.1:0.7].

K	*S*	0.3	0.4	0.5	0.6	0.7	AVG
3	3	82.1	77.5	71.2	58.0	43.3	66.4
5	3	82.1	78.4	71.3	59.6	43.4	66.9
7	3	82.7	78.7	71.6	58.3	42.8	66.8
9	3	81.1	77.3	70.5	57.7	44.4	66.2
11	3	81.6	77.2	70.6	58.3	43.8	66.3
7	5	81.9	77.9	70.2	58.1	43.4	66.3
7	7	81.2	77.3	70.5	58.4	43.7	66.2

6.4. Ablation Study of CC Module

We compare the performances of the aggregating chunk convolution features with different dilation rates, and the results are shown in Table 5. As can be observed, aggregating chunk convolutional features with a larger dilation rate generally yields a higher mAP. However, as the dilation rate continues to increase, it resulted in a performance degradation. In addition, we also compared the replacement of the ordinary 1D convolution with a dilated 1D convolution and did not obtain a better performance. This shows that taking into account different redundancy rates helps to improve the generalization of the model, but an excessive dilation rate hinders the capture of adjacent information, resulting in insufficient features information.

Table 5. Ablation Study (impact of CC module). Comparing the effect of chunk convolutions with different dilation rates (D). DC represents dilated convolution, SC represents standard convolution, measured by mAP@tIoU at different thresholds and the average mAP (mAP@avg) [0.3:0.1:0.7]. $\surd$ indicate the selected dilation rate.

SC	DC	D = 1	D = 3	D = 6	D = 9	D = 13	0.3	0.4	0.5	0.6	0.7	AVG
$\surd$		$\surd$					81.7	77.5	70.9	58.5	43.3	66.4
$\surd$		$\surd$	$\surd$				82.1	78.0	70.3	57.6	43.4	66.3
$\surd$		$\surd$	$\surd$	$\surd$			82.7	78.7	71.6	58.3	42.8	66.8
$\surd$		$\surd$	$\surd$	$\surd$	$\surd$		81.7	77.2	69.7	57.9	42.6	65.8
$\surd$		$\surd$	$\surd$	$\surd$	$\surd$	$\surd$	81.4	77.8	70.8	58.7	43.2	66.4
	$\surd$	$\surd$	$\surd$	$\surd$			82.1	77.9	70.5	58.5	43.9	66.6

6.5. Ablation Study of Combination Strategies

In order to verify the effect of the three independent modules working together, the CC, MTC, and TD, we tried a variety of combination strategies, and the results are listed in Table 6. Obviously, the best performance is achieved when the combined strategy is CC $\rightarrow$ MTC $\rightarrow$ TD. Compared with the three-paths parallel mode, its performance exceeds at least 1.2% (mAP@0.5). In addition, it also has at least a 0.9% (mAP@0.5) advantage compared with other series combination strategies. This suggests that the optimal strategy is to take three steps. The CC module not only establishes long-range context dependencies but also effectively alleviates the local information redundancy. The MTC module and the following lightweight TD module are responsible for providing multiscale information and enhanced motion information. The three modules work together to better model the complex video structure.

Table 6. Ablation Study (impact of combination strategies). Comparing the effects of different combined strategies, measured by mAP@tIoU at different thresholds and the average mAP (mAP@avg) [0.3:0.1:0.7].

Strategy	0.3	0.4	0.5	0.6	0.7	AVG
CC → MTC → TD	82.7	78.7	71.6	58.3	42.8	66.8
CC → TD → MTC	81.9	77.7	70.3	57.6	42.5	66.0
TD → CC → MTC	81.8	77.8	70.4	58.3	42.5	66.2
TD → MTC → CC	81.7	77.6	70.0	56.8	43.2	65.8
MTC → TD → CC	81.9	78.1	70.7	57.7	43.5	66.4
MTC → CC → TD	82.2	77.6	70.4	58.7	43.7	66.5
Stack (avg)	82.2	77.9	70.4	57.8	43.5	66.3
Stack (max)	82.0	77.7	70.1	58.2	43.7	66.3
Cancat	81.1	76.9	69.8	58.3	44.0	66.0

6.6. Qualitative Results

Figure 2 visualizes the localization results and predicted categories of four action instances on THUMOS14 and compares the predicted results (green) with the corresponding GT (yellow). These instances include short (first and second), medium (third), and long (fourth) durations. It can be observed that the middle (third) and long (fourth) instances were correctly localized. However, the boundary positions of the short-duration instances (first and second) were imprecise. The reason lies in two aspects: the motion of a short instance changes rapidly, and the lack of contextual relevance makes it difficult to provide sufficient clues for the prediction layer. In addition, relative to the long instance, its IoU is extremely sensitive to the offset, so it is easy to be judged as a negative sample.

7. Conclusions

In this paper, we introduce a novel network for the temporal activity detection (TAD) task in untrimmed videos. Our proposed model consists of three modules that process input features in a serial manner. Specially, the input features are first passed through the CC module to reduce the redundant content while capturing long-range contextual information. Then, the output features of the previous step are processed by the MTC to aggregate the multiscale local features. Finally, the aggregated features are input to the TD module to enhance the weakened motion information and boundary features. Benefiting from the complementarity of three independent modules, our model outperforms the state-of-the-art methods by a big margin on two large-scale benchmarks, ActivityNet-v1.3 and THUMOS-14. Extensive experiments demonstrate the generalization ability and effectiveness of our approach.

Discussions: Temporal action detection is still an extremely challenging task, where the complexity of the video structure is an important factor. So far, it is still unclear how to effectively model complex temporal structures. The video has contextual relevance, which requires modeling long-range temporal structures. The usual practice is to stack 1D convolutions. However, the original video has the following characteristics: (1) the duration of different actions varies widely; (2) information redundancy between neighboring frames; and (3) the information is weakened in the long-distance delivery. According to the above characteristics, we designed the MTC, CC, and TD modules, respectively. The experimental results show that each module can help to improve the performance of the model. In addition, the video also has the characteristics of overlap, nonlinearity, spatio-temporal correlation, an inconsistent motion rate, and sparse boundary points. How to design an efficient network to model long-range temporal relationships while taking into account the video characteristics is the key to further improve the performance of the TAD task.

Author Contributions: Conceptualization, Y.H. and Y.Z.; methodology, Y.H. and Y.Z.; software, Y.H., X.H. and L.W.; validation, Y.H. and X.H.; formal analysis, X.H.; investigation, L.W.; resources, Y.H.; data curation, Y.H. and X.H.; writing—original draft preparation, X.H.; writing—review and editing, X.H.; visualization, L.W.; supervision, Y.Z.; project administration, Y.Z.; funding acquisition, Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Sichuan Sciences and Technology Program (No. 2019ZDZX0005) and the Sichuan Sciences and Technology Program (No. 2022ZHCG0007).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wang, Q.; Zhang, Y.; Zheng, Y.; Pan, P. RCL: Recurrent Continuous Localization for Temporal Action Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–20 June 2022; pp. 13566–13575.
2. Dai, R.; Das, S.; Minciullo, L.; Garattoni, L.; Francesca, G.; Bremond, F. Pdan: Pyramid dilated attention network for action detection. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Nashville, TN, USA, 19–25 June 2021; pp. 2970–2979.
3. Dai, X.; Singh, B.; Ng, J.Y.H.; Davis, L. Tan: Temporal aggregation network for dense multi-label action recognition. In Proceedings of the 2019 IEEE Winter Conference on Applications of Computer Vision (WACV), Honolulu, HI, USA, 7–11 January 2019; pp. 151–160.
4. Zhang, C.; Wu, J.; Li, Y. Actionformer: Localizing moments of actions with transformers. *arXiv* **2022**, arXiv:2202.07925.
5. Tan, J.; Tang, J.; Wang, L.; Wu, G. Relaxed transformer decoders for direct action proposal generation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 13526–13535.
6. Dai, R.; Das, S.; Kahatapitiya, K.; Ryoo, M.S.; Bremond, F. MS-TCT: Multi-Scale Temporal ConvTransformer for Action Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 21–24 June 2022; pp. 20041–20051.
7. He, L.; Todorovic, S. DESTR: Object Detection with Split Transformer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 21–24 June 2022; pp. 9377–9386.
8. Li, Y.; Wu, C.Y.; Fan, H.; Mangalam, K.; Xiong, B.; Malik, J.; Feichtenhofer, C. MViTv2: Improved Multiscale Vision Transformers for Classification and Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 21–24 June 2022; pp. 4804–4814.
9. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008.
10. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv* **2018**, arXiv:1810.04805.
11. Chao, Y.W.; Vijayanarasimhan, S.; Seybold, B.; Ross, D.A.; Deng, J.; Sukthankar, R. Rethinking the faster r-cnn architecture for temporal action localization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 1130–1139.
12. Yang, L.; Peng, H.; Zhang, D.; Fu, J.; Han, J. Revisiting anchor mechanisms for temporal action localization. *IEEE Trans. Image Process.* **2020**, *29*, 8535–8548. [CrossRef]
13. Chen, G.; Zheng, Y.D.; Wang, L.; Lu, T. DCAN: Improving temporal action detection via dual context aggregation. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 22 February–1 March 2022; Volume 36, pp. 248–257.
14. Gao, S.H.; Cheng, M.M.; Zhao, K.; Zhang, X.Y.; Yang, M.H.; Torr, P. Res2net: A new multi-scale backbone architecture. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *43*, 652–662. [CrossRef] [PubMed]
15. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
16. Xia, K.; Wang, L.; Zhou, S.; Zheng, N.; Tang, W. Learning To Refactor Action and Co-Occurrence Features for Temporal Action Localization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–20 June 2022; pp. 13884–13893.
17. Simonyan, K.; Zisserman, A. Two-stream convolutional networks for action recognition in videos. *Adv. Neural Inf. Process. Syst.* **2014**, *27*. [CrossRef]
18. Caba Heilbron, F.; Escorcia, V.; Ghanem, B.; Carlos Niebles, J. Activitynet: A large-scale video benchmark for human activity understanding. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7 June 2015; pp. 961–970.

19. Idrees, H.; Zamir, A.R.; Jiang, Y.G.; Gorban, A.; Laptev, I.; Sukthankar, R.; Shah, M. The THUMOS challenge on action recognition for videos "in the wild". *Comput. Vis. Image Underst.* **2017**, *155*, 1–23. [CrossRef]
20. Shou, Z.; Wang, D.; Chang, S.F. Temporal action localization in untrimmed videos via multi-stage cnns. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 1049–1058.
21. Gao, J.; Yang, Z.; Chen, K.; Sun, C.; Nevatia, R. Turn tap: Temporal unit regression network for temporal action proposals. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 3628–3636.
22. Gao, J.; Chen, K.; Nevatia, R. Ctap: Complementary temporal action proposal generation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 68–83.
23. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*. [CrossRef] [PubMed]
24. Xu, H.; Das, A.; Saenko, K. R-c3d: Region convolutional 3d network for temporal activity detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 5783–5792.
25. Long, F.; Yao, T.; Qiu, Z.; Tian, X.; Luo, J.; Mei, T. Gaussian temporal awareness networks for action localization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 344–353.
26. Liu, Q.; Wang, Z. Progressive boundary refinement network for temporal action detection. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 11612–11619.
27. Lin, C.; Xu, C.; Luo, D.; Wang, Y.; Tai, Y.; Wang, C.; Li, J.; Huang, F.; Fu, Y. Learning salient boundary feature for anchor-free temporal action localization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 3320–3329.
28. Liu, X.; Wang, Q.; Hu, Y.; Tang, X.; Zhang, S.; Bai, S.; Bai, X. End-to-end temporal action detection with transformer. *IEEE Trans. Image Process.* **2022**, *31*, 5427–5441. [CrossRef]
29. Zhao, Y.; Xiong, Y.; Wang, L.; Wu, Z.; Tang, X.; Lin, D. Temporal action detection with structured segment networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2914–2923.
30. Lin, T.; Zhao, X.; Su, H.; Wang, C.; Yang, M. Bsn: Boundary sensitive network for temporal action proposal generation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
31. Lin, T.; Liu, X.; Li, X.; Ding, E.; Wen, S. Bmn: Boundary-matching network for temporal action proposal generation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 3889–3898.
32. Su, H.; Gan, W.; Wu, W.; Qiao, Y.; Yan, J. Bsn++: Complementary boundary regressor with scale-balanced relation modeling for temporal action proposal generation. In Proceedings of the AAAI Conference on Artificial Intelligence, Online, 2–9 February 2021; pp. 2602–2610.
33. Zhao, P.; Xie, L.; Ju, C.; Zhang, Y.; Wang, Y.; Tian, Q. Bottom-up temporal action localization with mutual regularization. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 539–555.
34. Zeng, R.; Huang, W.; Tan, M.; Rong, Y.; Zhao, P.; Huang, J.; Gan, C. Graph convolutional networks for temporal action localization. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27 October–2 November 2019; pp. 7094–7103.
35. Xu, M.; Zhao, C.; Rojas, D.S.; Thabet, A.; Ghanem, B. G-tad: Sub-graph localization for temporal action detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10156–10165.
36. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2980–2988.
37. Bodla, N.; Singh, B.; Chellappa, R.; Davis, L.S. Soft-NMS–improving object detection with one line of code. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 5561–5569.
38. Carreira, J.; Zisserman, A. Quo vadis, action recognition? A new model and the kinetics dataset. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6299–6308.
39. Kay, W.; Carreira, J.; Simonyan, K.; Zhang, B.; Hillier, C.; Vijayanarasimhan, S.; Viola, F.; Green, T.; Back, T.; Natsev, P.; et al. The kinetics human action video dataset. *arXiv* **2017**, arXiv:1705.06950.
40. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
41. Tran, D.; Wang, H.; Torresani, L.; Ray, J.; LeCun, Y.; Paluri, M. A closer look at spatiotemporal convolutions for action recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 6450–6459.
42. Alwassel, H.; Giancola, S.; Ghanem, B. Tsp: Temporally-sensitive pretraining of video encoders for localization tasks. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 3173–3183.
43. Liu, X.; Bai, S.; Bai, X. An Empirical Study of End-to-End Temporal Action Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–20 June 2022; pp. 20010–20019.
44. Yang, H.; Wu, W.; Wang, L.; Jin, S.; Xia, B.; Yao, H.; Huang, H. Temporal Action Proposal Generation with Background Constraint. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 22 February–1 March 2022; Volume 36, pp. 3054–3062.
45. Zhao, C.; Thabet, A.K.; Ghanem, B. Video self-stitching graph network for temporal action localization. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 13658–13667.

46. Sridhar, D.; Quader, N.; Muralidharan, S.; Li, Y.; Dai, P.; Lu, J. Class semantics-based attention for action detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, QC, Canada, 10–17 October 2021; pp. 13739–13748.
47. Liu, X.; Hu, Y.; Bai, S.; Ding, F.; Bai, X.; Torr, P.H. Multi-shot temporal event localization: A benchmark. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 12596–12606.
48. Bai, Y.; Wang, Y.; Tong, Y.; Yang, Y.; Liu, Q.; Liu, J. Boundary content graph neural network for temporal action proposal generation. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 121–137.
49. Qing, Z.; Su, H.; Gan, W.; Wang, D.; Wu, W.; Wang, X.; Qiao, Y.; Yan, J.; Gao, C.; Sang, N. Temporal context aggregation network for temporal action proposal refinement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 485–494.

sensors

MDPI

Article

A New Method for Image Protection Using Periodic Haar Piecewise-Linear Transform and Watermarking Technique †

Andrzej Dziech, Piotr Bogacki * and Jan Derkacz

Institute of Telecommunications, Faculty of Computer Science, Electronics and Telecommunications, AGH University of Science and Technology, Mickiewicza 30, 30-059 Kraków, Poland

* Correspondence: pbogacki@agh.edu.pl

† This paper is an extended version of "A Novel Watermark Method for Image Protection Based on Periodic Haar Piecewise-Linear Transform" published in the Proceedings of the 10th International Conference, MCSS 2020, Kraków, Poland, 8–9 October 2020.

Abstract: The paper presents a novel data-embedding method based on the Periodic Haar Piecewise-Linear (PHL) transform. The theoretical background behind the PHL transform concept is introduced. The proposed watermarking method assumes embedding hidden information in the PHL transform domain using the luminance channel of the original image. The watermark is embedded by modifying the coefficients with relatively low values. The proposed method was verified based on the measurement of the visual quality of an image with a watermark with respect to the length of the embedded information. In addition, the bit error rate (BER) is also considered for different sizes of a watermark. Furthermore, a method for the detection of image manipulation is presented. The elaborated technique seems to be suitable for applications in digital signal and image processing where high imperceptibility and low BER are required, and information security is of high importance. In particular, this method can be applied in systems where the sensitive data is transmitted or stored and needs to be protected appropriately (e.g., in medical image processing).

Keywords: watermarking; image protection; PHL transform; data embedding; multimedia systems

Citation: Dziech, A.; Bogacki, P.; Derkacz, J. A New Method for Image Protection Using Periodic Haar Piecewise-Linear Transform and Watermarking Technique. *Sensors* **2022**, 22, 8106. https://doi.org/10.3390/s22218106

Academic Editor: Manuel José Cabral dos Santos Reis

Received: 19 September 2022
Accepted: 18 October 2022
Published: 22 October 2022

1. Introduction

There is a large number of areas where the security of multimedia content is crucial for ensuring privacy and citizens' rights in general. Digital watermarking is an efficient and versatile technical means for embedding secret information into multimedia objects, such as still images, videos, and audio files. An example of such secret, sensitive information can be medical data related to patients. Watermarking technology can assure protection of the digital content against unauthorized access, tampering, sensitive information disclosure, or copyright infringement. Methods based on watermarking may be also used for such applications as steganography and pseudonymization of private data. A graphic or audio file marked in this way can help locate websites or FTP servers where these files are unlawfully shared. As a result, a digital watermark now has high hopes for an effective fight against fraud.

The efficient watermark should be characterized by the following features: Imperceptibility—the watermark should be imperceptible to the human eye, and the inserted information should not deteriorate the visual quality of an original image. Robustness—the watermark is detectable even after the original image transformation and is difficult to be removed. Consideration of local image properties—the watermark is inserted with varying intensity in different areas, depending on the characteristics of the area (e.g., brightness) Watermark decoding method—the watermark can be read based on the watermarked image only, without the need to verify against the original image.

Image watermarking can be performed in the spatial or transform domain. Spatial domain methods usually result in direct modifications of image data, such as color bands,

and brightness. The common method for embedding a watermark in the spatial domain is the Least Significant Bit (LSB) method where the secret information is inserted into the original image by modifying or replacing the least significant bits of pixels. On the other hand, transform-based techniques rely on changing spectral factors in the domain of a specific transform. To retrieve the image with an embedded watermark one needs to perform the corresponding inverse transform operation. Watermarks embedded in the transform domains are typically more reliable in comparison with the watermarks inserted in the spatial domain [1,2].

The most widely used transforms used in digital watermarking include discrete cosine transform (DCT) [3–6], discrete wavelet transforms (DWT) [7,8] and discrete Fourier transform (DFT) [9,10]. Combination of different transform methods can be implemented, (e.g., DCT and DWT transform) [11–13]. Additionally, transform-based techniques can be used jointly with other methods, such as, (e.g., singular value decomposition (SVD) [14] or discrete fractional random transform (DFRNT)) [2]. There are also new approaches that apply novel types of transforms that are orthogonal and can be parameterized [15].

In [16], Yan et al. presented a data hiding scheme based on LSB modification in the Piecewise-Linear Haar transform for audio signals. Yang et al. in [17] proposed a reversible data hiding method dedicated to images using symmetrical histogram expansion also in the domain of this transform.

However, Periodic Haar Piecewise-Linear (PHL) transform is only mentioned in the literature with regard to image compression tasks [18].

For obvious reasons medical images are private to the patient and authorized medical personnel and should be protected from unauthorized viewers. One method to protect such images is using cryptography including traditional symmetric cryptosystems and biometrics [19–21]. Digital content, in particular this related to medical images, is more and more often protected by a combination of tools, such as encryption and watermarking. As defined in [22] encryption algorithms can be considered as an "a priori" protection mechanism since once data is decrypted, it is no longer protected. A complement to "a priori" mechanism is "a posteriori" protection, which can be provided by watermarking.

Apart from unauthorized access to sensitive content, another potential threat to medical multimedia content is possible manipulations. Existing, widely available, image editing software and image altering tools allow us to easily manipulate a digital image nowadays. Studies of various image manipulation detection techniques are available in the literature. Numerous image forgeries that can be performed on the image and different image manipulation detection and localization methods were presented in [23]. Image manipulation can also concern biomedical sciences where the use of images to depict laboratory results is widely disseminated. Results published in [24] have shown an alarming level of image manipulation in the published record. A dedicated tool was used to detect some of the most common misbehaviors, running tests on a random set of papers and the full publishing record of a journal.

Currently, image tampering detection can be also realized with the use of Convolutional Neural Networks [25]. Image protection and manipulation detection are extremely relevant in all applications where the sensitive data is transmitted from the imaging sensor to a remote destination where it is further processed and analyzed [26]. Such protection can be realized in aerial photography, area monitoring, and satellite imagery [27]. The same applies to medical applications of remote sensing where electromagnetic radiation is most commonly the sensing medium and the sensors of diagnostic devices, which are exterior to the body of a patient, can detect various features of human tissues in a noninvasive way [28].

The paper is organized as follows. The next section is dedicated to Periodic Haar Piecewise-Linear Transform. Section 3 introduces a new method for data embedding. Section 4 presents the potential application of the proposed algorithm for the detection of image manipulations. In Section 5 the experimental results are presented and the

comparison between the proposed solution and the DCT approach is discussed. Finally, Section 6 contains the conclusions and future work.

2. Periodic Haar Piecewise-Linear PHL Transform

This section covers the most important theoretical aspects related to Periodic Haar Piecewise-Linear (PHL) transform. The thorough description and further information are presented in detail in [29]. The Haar functions are defined by the following formulas:

$$har(0,t) = 1 \qquad \text{for } t \in (-\infty,\infty), \quad \text{usually } T = 1 \tag{1a}$$

$$har(i,t) = \begin{cases} 2^{\frac{k-1}{2}} & \text{for } [\frac{i}{2^{k-1}} - 1] \le t < [\frac{i+\frac{1}{2}}{2^{k-1}}] \\ -2^{\frac{k-1}{2}} & \text{for } [\frac{i+\frac{1}{2}}{2^{k-1}} - 1] \le t < [\frac{i+1}{2^{k-1}} - 1] \\ 0 & \text{otherwise} \end{cases} \tag{1b}$$

where $0 < k < \log_2 N$, $1 \le i \le 2^k$.

In turn, the PHL functions can be calculated by performing the integration of these Haar functions. It can be realized by using the below formulas:

$$PHL(0,t) = 1 \quad t \in (-\infty,\infty) \tag{2a}$$

$$PHL(1,t) = [\frac{2}{T}\int_{mT}^{t+mT} har(1,\tau)d\tau] + \frac{1}{2} \tag{2b}$$

$$PHL(i+1,t) = \frac{2^{k+1}}{T}\int_{mT}^{t+mT} har(i+1,\tau)d\tau \tag{2c}$$

where $i = 1,2,...,N-2; k = 1,2,...,\log_2 N - 1; m = 0,1,2,...$;

k—index of group of PHL functions;

m—number of period.

Figure 1 depicts the derivatives (in distributive sense) of Haar functions. The PHL functions are linearly independent but they do not satisfy the orthogonality condition.

2.1. One-Dimensional PHL Transform

To perform forward and inverse PHL transform, the following matrix equations can be used:

a. Forward transform

$$[\mathrm{C(N)}] = [-\frac{1}{2^{k+1}}][\mathrm{PHL(N)}][\mathrm{X(N)}] \tag{3}$$

b. Inverse transform

$$[\mathrm{X(N)}] = [\mathrm{IPHL(N)}][\mathrm{C(N)}] \tag{4}$$

where [C(N)]—vector of PHL coefficients (PHL spectrum);

[X(N)]—vector of sampled signal;

[PHL(N)]—matrix of forward transform;

[IPHL(N)]—matrix of inverse transform;

$[-\frac{1}{2^{k+1}}]$—diagonal matrix of normalization.

$$[-\frac{1}{2^{k+1}}] = diag[1, -\frac{1}{2^1}, -\frac{1}{2^2}(2\ times), -\frac{1}{2^3}(4\ times), ...,, -\frac{1}{2^k}(2^{k-1}\ times)] \tag{5}$$

The first row of the forward transform matrix consists of number one at the first position and the remaining elements are equal to zero. Other rows are composed of derivatives (in a distributive sense) of periodic Haar functions. The matrix for the inverse transform [IPHL(N)] is constructed in such a way that particular rows consist of PHL

function values calculated for the same argument. For instance, the [PHL(N)] and [IPHL(N)] matrices, for N = 8, are presented below:

$$[\mathrm{PHL}(8)] = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 0 & 0 & 0 & -2 & 0 & 0 & 0 \\ \sqrt{2} & 0 & -2\sqrt{2} & 0 & \sqrt{2} & 0 & 0 & 0 \\ \sqrt{2} & 0 & 0 & 0 & \sqrt{2} & 0 & -2\sqrt{2} & 0 \\ 2 & -4 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & -4 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & -4 & 2 & 0 \\ 2 & 0 & 0 & 0 & 0 & 0 & 2 & -4 \end{bmatrix} \quad (6)$$

$$[\mathrm{IPHL}(8)] = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & \frac{1}{4} & \frac{\sqrt{2}}{2} & 0 & 2 & 0 & 0 & 0 \\ 1 & \frac{1}{2} & \sqrt{2} & 0 & 0 & 0 & 0 & 0 \\ 1 & \frac{3}{4} & \frac{\sqrt{2}}{2} & 0 & 0 & 2 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & \frac{3}{4} & 0 & \frac{\sqrt{2}}{2} & 0 & 0 & 2 & 0 \\ 1 & \frac{1}{2} & 0 & \sqrt{2} & 0 & 0 & 0 & 0 \\ 1 & \frac{1}{4} & 0 & \frac{\sqrt{2}}{2} & 0 & 0 & 0 & 2 \end{bmatrix} \quad (7)$$

In this case, according to Equation (5), the diagonal matrix of normalization takes the following form:

$$[-\frac{1}{2^{k+1}}] = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{4} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\frac{1}{8} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -\frac{1}{8} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -\frac{1}{8} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\frac{1}{8} \end{bmatrix} \quad (8)$$

It can be observed that:

$$[-\frac{1}{2^{k+1}}][\mathrm{PHL}(8)][\mathrm{IPHL}(8)] = [\mathrm{I(N)}] \quad (9)$$

where [I(N)] is the identity matrix.

2.2. Two-Dimensional PHL Functions and Transform

The 2D PHL transform can be formulated in the following way:

a. Forward transform

$$[\mathrm{C}(N_x, N_y)] = [-\frac{1}{2^{k_y+1}}][\mathrm{PHL}(N_y)][\mathrm{F}(N_x, N_y)][\mathrm{PHL}(N_x)]^T[-\frac{1}{2^{k_x+1}}]^T \quad (10)$$

b. Inverse transform

$$[\mathrm{F}(N_x, N_y)] = [\mathrm{IPHL}(N_y)][\mathrm{C}(N_x, N_y)][\mathrm{IPHL}(N_x)]^T \quad (11)$$

where $[\mathrm{F}(N_x, N_y)]$—matrix of 2D signal;
$[\mathrm{C}(N_x, N_y)]$—matrix of coefficients (2D PHL spectrum);
$[\mathrm{PHL}(N_y)]$, $[\mathrm{PHL}(N_x)]$—matrices of 1D PHL forward transform;
$[\mathrm{IPHL}(N_y)]$, $[\mathrm{IPHL}(N_x)]$—matrices of 1D PHL inverse transform;
$[-\frac{1}{2^{k_y+1}}]$, $[-\frac{1}{2^{k_x+1}}]$—diagonal matrices of normalization.

The non-periodic Haar Piecewise-Linear Transforms have an order $(N + 1)$ while the PHL Transforms have an order (N). Due to this fact, PHL transforms can be applied in digital signal and image processing since the data usually has a dimension that is a power of 2.

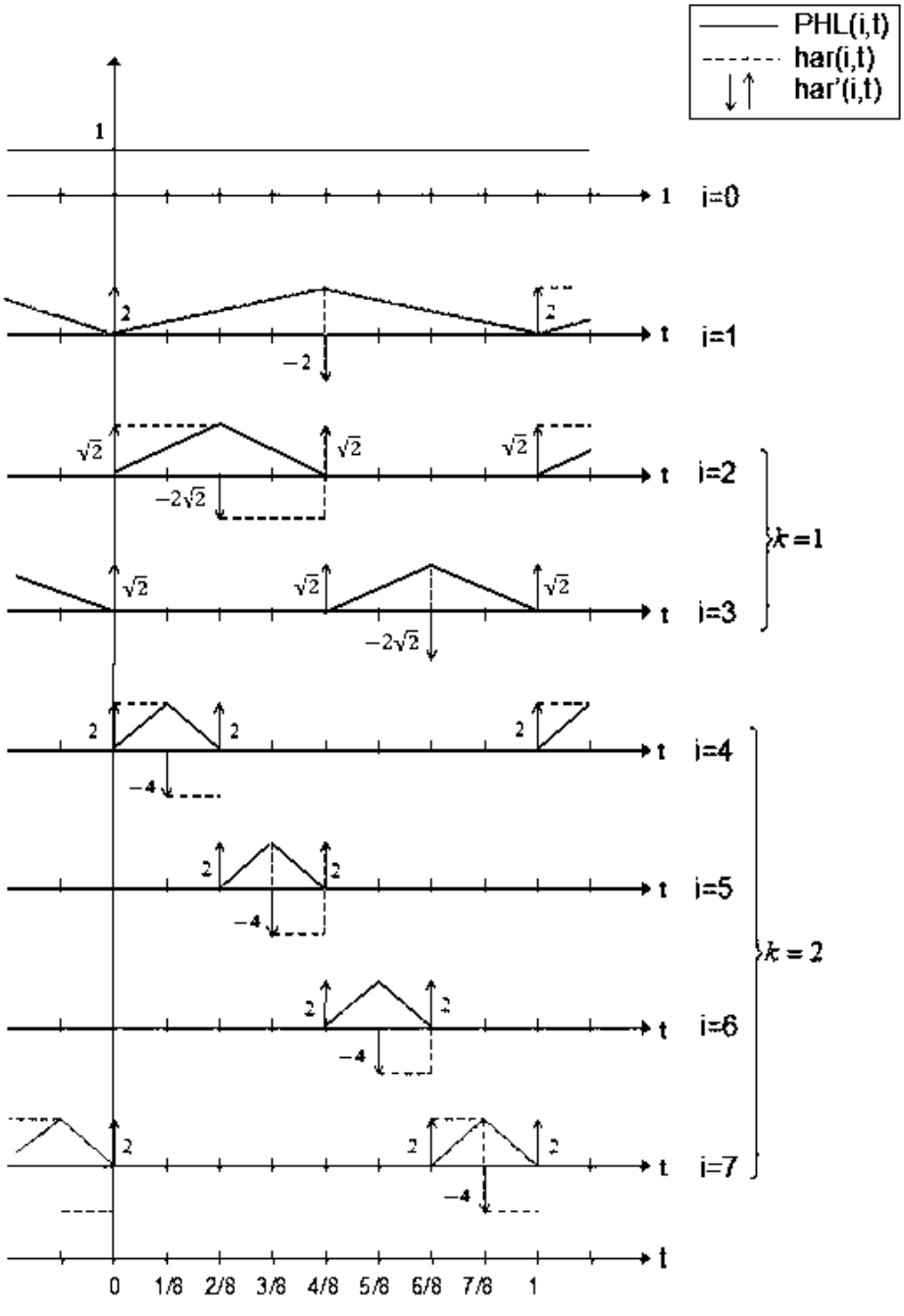


Figure 1. Set of PHL functions for N = 8.

3. Data Embedding in PHL Spectrum

The watermarking approach, presented in this paper, is based on inserting secret information in the PHL transform domain. The method assumes that the PHL spectrum is calculated only for the luminance channel of the given image, representing its grayscale version. To speed up the computations, the forward transform operation is performed on smaller subimages, i.e., blocks with the size: 8×8 pixels, using Equation (10) and the matrices (6) and (7).

As a result, after performing the above process to the input signal, we get its spectral coefficients in the PHL domain. Typically, a limited number of these coefficients carry most of the signal energy [30,31].

The PHL transform may be used for image compression purposes [32]. In this task, the spectral coefficients that are above a given threshold are kept while the remaining ones are set to zero. Following this approach, our method assumes embedding of the watermark by modification of the coefficients having relatively low values. To perform this operation, the PHL coefficients are split into channels. Each channel groups the spectral coefficients with the same indices from each block processed in the forward transform step. This way, we obtain 64 PHL transform channels. The study of a set of various images and their spectra indicates that the top-left channel cumulates most of the signal energy. It is well depicted in Figure 2 which shows the PHL spectrum coefficients after grouping into 64 channels.

For the testing purpose and the presentation of the image manipulation detection method in the following section, the Optical Coherence Tomography (OCT) images, having the resolution of 1536×496 pixels, were used [33]. The OCT is a non-invasive imaging examination that uses light waves to take cross-section pictures of the human retina. One sample image of this type is shown in Figure 3. The tests show that the blocks: 37–39, 45–47, and 53–55, marked in Figure 4, should be usually selected for the process of inserting secret information. This conclusion is based on the analysis of spectra of diverse images with varying content and characteristics. For the selection of the best channel for watermark embedding, the mean of all absolute values from each block is calculated. The channel with the lowest mean is chosen as the first candidate for the subsequent data embedding operation. To increase the capacity of the watermark, other blocks can be selected afterward, considering their mean values sorted in ascending order.

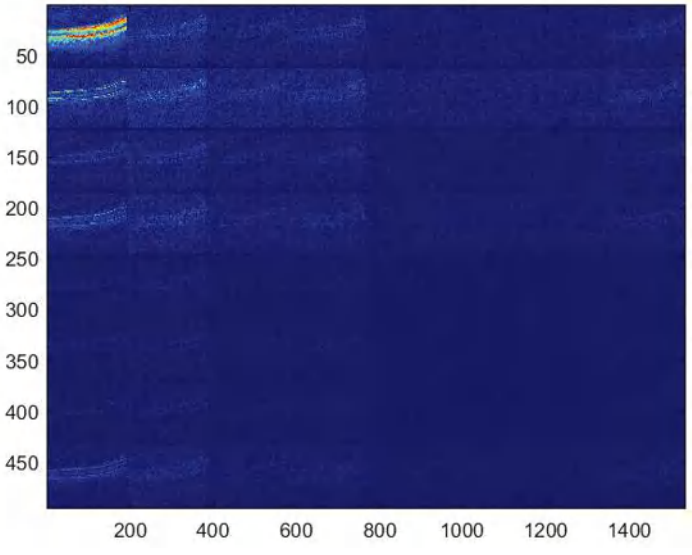


Figure 2. PHL spectrum coefficients grouped into channels.

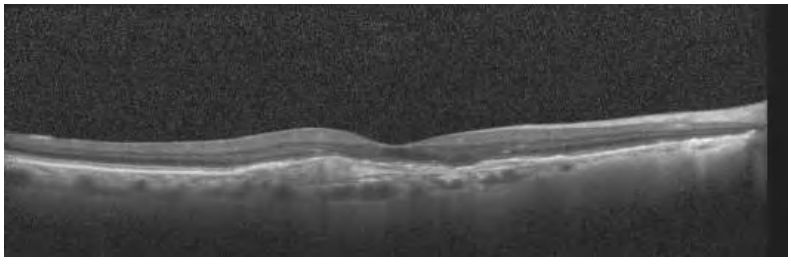


Figure 3. Sample OCT image.

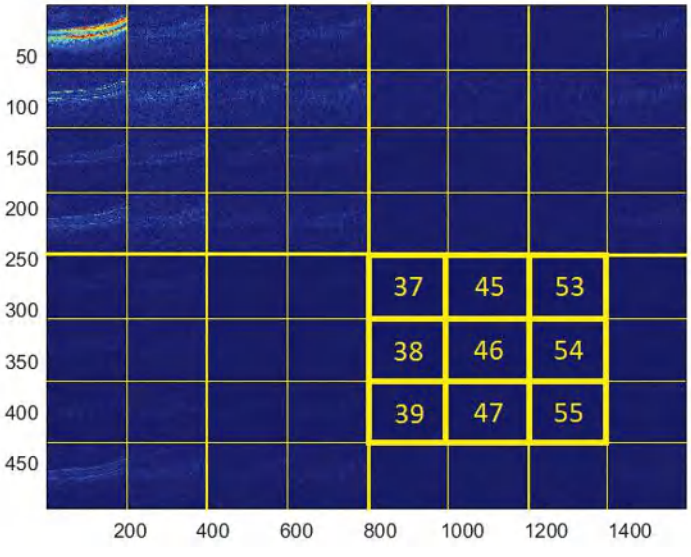


Figure 4. Blocks selected for data embedding.

The selected channel coefficients are replaced with the consecutive bits of the message that is to be hidden in the image. Subsequently, the channel coefficients need to be relocated back to their previous positions. The final step is the inverse PHL transform of the modified image spectrum that results in the image with an inserted watermark. The stages of the whole embedding process are presented in Figure 5.

For the recovery of the embedded information, the same steps as previously need to be performed—the forward transform, the grouping of PHL coefficients, and finally extracting information from the selected channel or channels.

The selection of nine blocks for watermark embedding can be performed adaptively, as described previously, or arbitrarily. In this way, the chosen order can be used as an additional key at the watermark extraction phase.

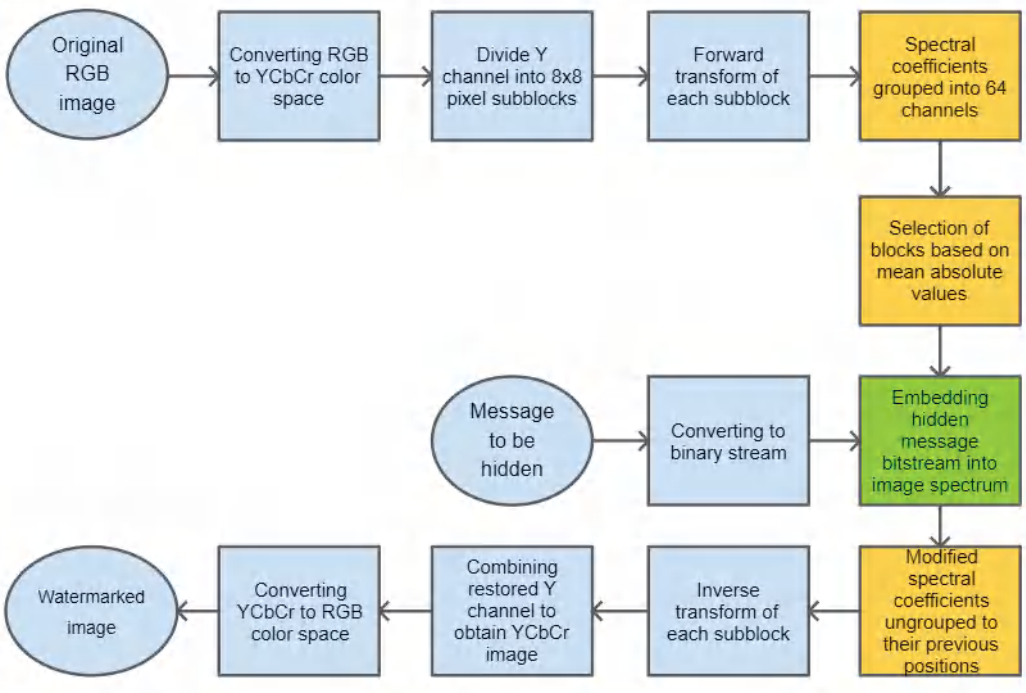


Figure 5. Block diagram for the base process of watermark embedding.

4. Image Manipulation Detection

The information embedded as a watermark can be used to detect potential manipulations of the image. It would be beneficial if the hidden message could somehow describe the content of the image so that later, during the recovery phase, it could be compared with a newly generated description for the watermarked image. In case these two descriptions differ significantly, it could be stated that the watermarked image has been tampered with.

In this paper, as a method for image description, MPEG-7 Edge Histogram descriptor (EHD) has been selected. It is a visual texture descriptor that captures the spatial distribution of five types of edges in an image: vertical, horizontal, two diagonals, and non-directional edge. It is created by dividing an input image into 16 (4×4) blocks, which is depicted in Figure 6. For each block, a histogram of all the above-mentioned types of edges is calculated. Therefore, it consists of $4 \times 4 \times 5 = 80$ values that compose this descriptor [34].

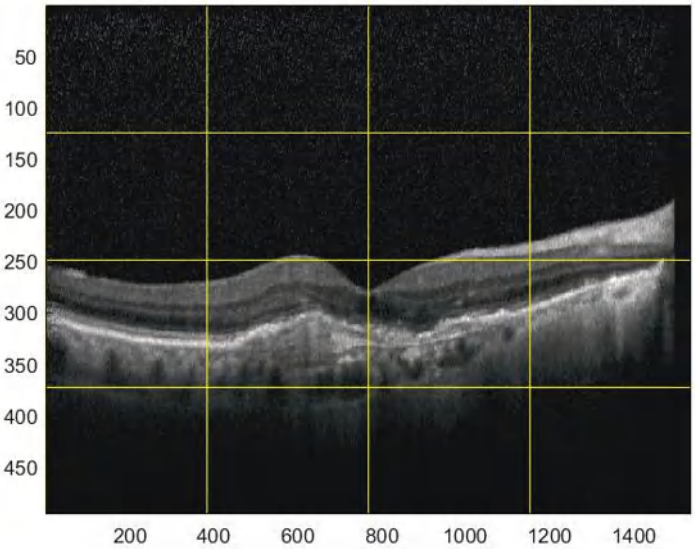


Figure 6. Blocks for which EHD descriptor is calculated.

In the first stage, the Edge Histogram descriptor is calculated for the given image. Its values are binarized to create a message bitstream which is then embedded into the image.

To detect potential manipulation of the watermarked image, it is necessary to calculate the EHD descriptor again and compare it with the one recovered from the watermark. The particular steps for image manipulation detection are shown in Figure 7.

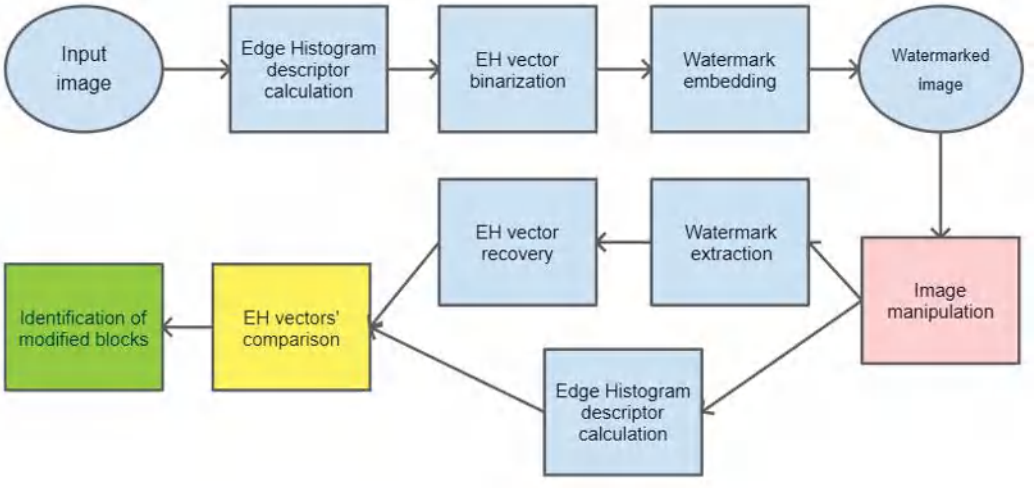
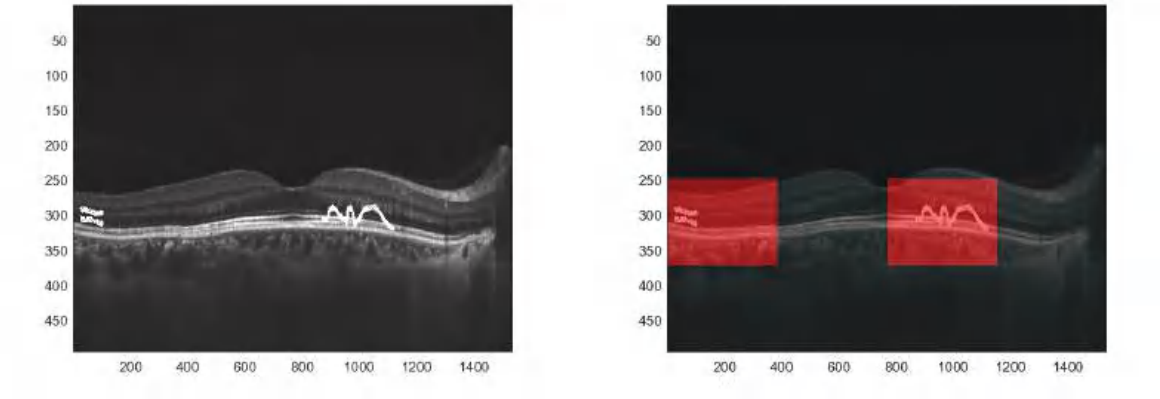


Figure 7. Block diagram for image manipulation detection process.

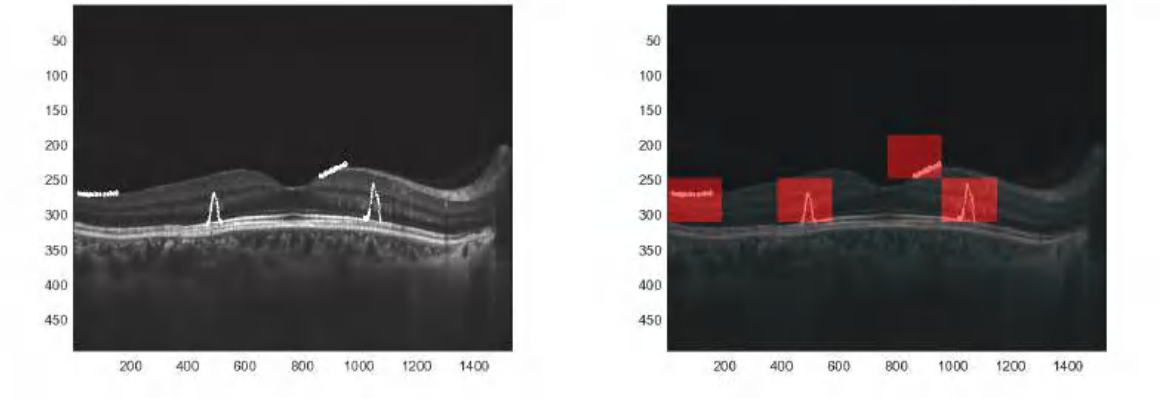
When the difference between particular values of both descriptors is significant, one can determine that the image has been modified. Furthermore, since the EHD descriptor returns 5 values for each of the 16 blocks, the proper analysis of differences at the given positions can precisely indicate which of these 16 blocks have been tampered with. This is presented in Figure 8. A sample tampered image is presented in Figure 8a and the image with selected blocks that have been modified is shown in Figure 8b.



(**a**) (**b**)

Figure 8. Result of image manipulation detection. (**a**) Tampered watermarked image. (**b**) Detected regions where the image has been manipulated.

To obtain better precision for image manipulation detection the image can be initially divided into smaller sub-images which are then further processed following the same steps as in the previous example. In such a way, the blocks that are identified to have been tampered with are of smaller dimensions. This is depicted in Figure 9.



(**a**) (**b**)

Figure 9. Result of image manipulation detection with greater precision. (**a**) Tampered watermarked image. (**b**) Detected regions with greater precision.

5. Experimental Results

The verification of the proposed algorithm is performed by measuring of Peak Signal to Noise Ratio (PSNR), which represents the visual quality of a watermarked image in relation to the total size of a watermark. Additionally, to consider the human visual system (HVS), Structural Similarity (SSIM) metric [35] and Universal Quality Image (UQI) index [36] are measured to assess the quality of the image with an embedded watermark. Furthermore, the bit error rate (BER) is also analyzed, for different lengths of the hidden message. The measurements of these ratios were performed for watermarks inserted in DCT and PHL transform domains so that the performance of both approaches may be compared. For test purposes, a random bit stream is used as a watermark message. The tests were carried out

in a MATLAB environment. The referenced DCT method originates from the one described in [3].

For test purposes, 23 images from 'Images 4k' dataset [37] have been selected. The dataset contains 2057 files. The test images were selected in such a way that they represent different visual characteristics, i.e., low and high contrast and brightness as well as various color distributions. The dimensions of the images were reduced by half to 1920 × 1080 size so that the calculations and the watermark embedding process are speeded up.

The relation between the PSNR ratio and the length of a hidden bit stream is presented in Figure 10. It can be observed that a perceptual quality of an image with a watermark inserted in the PHL spectrum is consistently better than in the case of a watermark embedded in the DCT domain. It is assumed that the PSNR above 35 dB indicates that the two images being compared are visually identical, with no perceptual loss of quality [38]. Therefore, both techniques provide satisfying results as far as the imperceptibility of a watermark is concerned, for the size of a watermark exceeding even 100,000 bits.

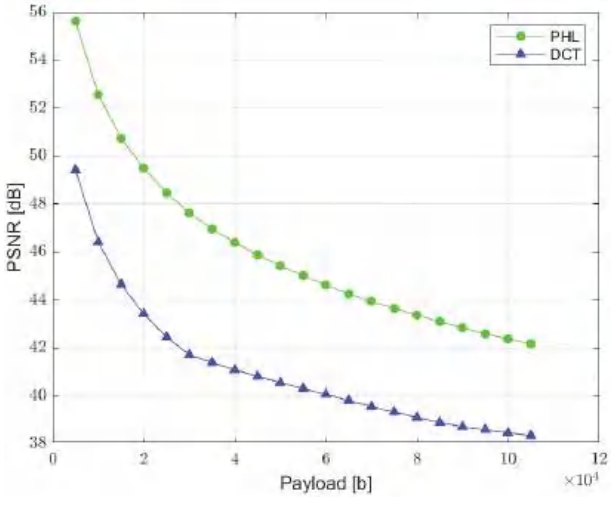


Figure 10. Relation between PSNR ratio and the watermark capacity (PHL vs. DCT).

SSIM is a quality assessment metric based on the visual changes in local structure and contrast between two images. It provides a good approximation of human visual perception. The metric values can range from 0 to 1, where 1 indicates perfect similarity [35]. The relation between SSIM and the total size of a watermark is presented in Figure 11. The results measured for the PHL method are slightly better than the ones achieved in the DCT approach. However, both methods according to this metric provide satisfying results.

UQI index is designed to model image distortion as a combination of three factors: loss of correlation, luminance distortion, and contrast distortion. Although it does not employ any human visual system model, it was proved to be consistent with subjective quality assessment [36]. UQI index can vary between −1 and 1, where value 1 indicates no distortion present in the image. The relation between UQI and the length of a hidden message is presented in Figure 12.

The relation between the BER ratio and the size of a watermark is shown in Figure 13. It can be noticed that both methods guarantee a low bit error rate (<0.1%) for the watermark size ranging from 5000 to 105,000 bits. Therefore, both solutions are useful when a limited, but still, in most applications, sufficient, amount of information needs to be hidden in an image.

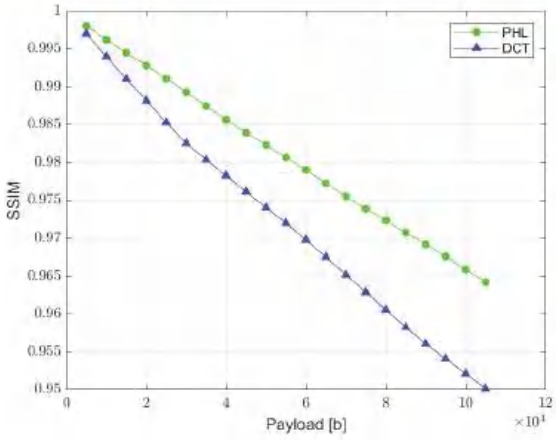


Figure 11. Relation between SSIM metric and the watermark capacity (PHL vs. DCT).

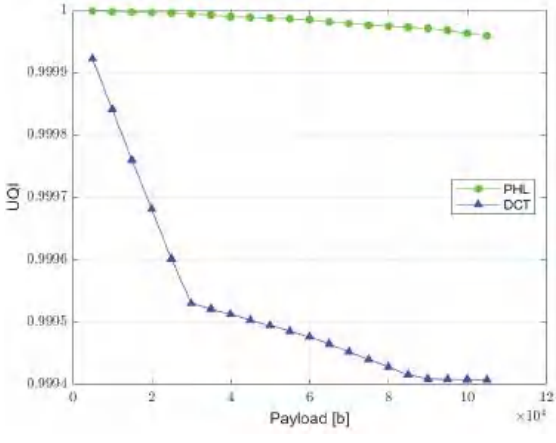


Figure 12. Relation between UQI index and the watermark capacity (PHL vs. DCT).

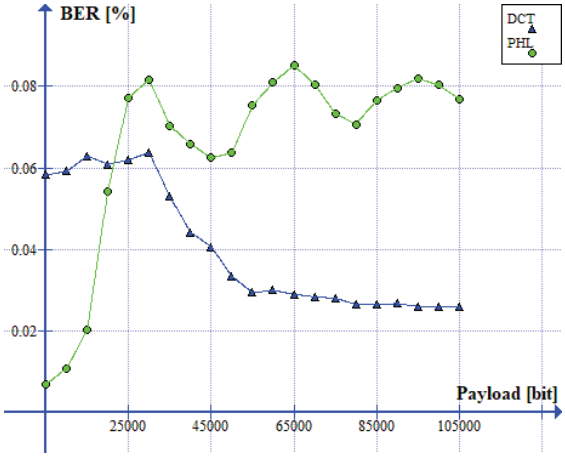


Figure 13. Relation between BER ratio and the watermark capacity (PHL vs. DCT).

6. Conclusions and Future Work

We have presented a new watermarking scheme that is based on inserting a message bitstream in the PHL transform domain. The method offers a high capacity for hidden information and simultaneously satisfies the initial requirements of low image distortion and high accuracy during the watermark recovery stage. Therefore, it is a promising technique that can be used in a wide range of multimedia systems and services with emphasis put on medical applications where the aforementioned conditions need to be met. In addition, a method for the detection of image manipulation has been presented.

Further investigations will cover potential enhancements so that the method could be robust to various types of attacks. Finally, we plan to apply our solution in many applications in the upcoming future.

Author Contributions: Conceptualization, P.B. and A.D.; methodology, P.B. and A.D.; software, P.B.; validation, P.B. and J.D.; formal analysis, P.B. and J.D.; investigation, P.B.; writing—original draft preparation, P.B.; writing—review and editing, J.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the European Union's Horizon 2020 Research and Innovation Programme, under Grant Agreement no. 830943, the ECHO project.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Sharma, P.K.; Sau, P.C.; Sharma, D. Digital image watermarking: An approach by different transforms using level indicator. In Proceedings of the 2015 Communication, Control and Intelligent Systems (CCIS), Mathura, India, 7–8 November 2015; pp. 259–263.
2. Zhou, N.R.; Hou, W.M.X.; Wen, R.H.; Zou, W.P. Imperceptible digital watermarking scheme in multiple transform domains. *Multimed Tools Appl.* **2018**, *77*, 30251–30267. [CrossRef]
3. Lan, T.-H.; Tewfik, A.H. A novel high-capacity data-embedding system. *IEEE Trans. Image Process.* **2006**, *15*, 2431–2440.
4. Kim, W.-H.; Hou, J.-U.; Jang, H.-U.; Lee, H.-K. Robust Template-Based Watermarking for DIBR 3D Images. *Appl. Sci.* **2018**, *8*, 911. [CrossRef]
5. Li, H.; Guo, X. Embedding and Extracting Digital Watermark Based on DCT Algorithm. *J. Comput. Commun.* **2018**, *6*, 287–298. [CrossRef]
6. Xu, Z.J.; Wang, Z.Z.; Lu, Q. Research on Image Watermarking Algorithm Based on DCT. *Procedia Environ. Sci.* **2011**, *10*, 1129–1135. [CrossRef]
7. Zhou, X.; Zhang, H.; Wang, C. A Robust Image Watermarking Technique Based on DWT, APDCBT, and SVD. *Symmetry* **2018**, *10*, 77. [CrossRef]
8. Narang, M.; Vashisth, S. Digital Watermarking using Discrete Wavelet Transform. *Int. J. Comput. Appl.* **2013**, *74*, 34–38. [CrossRef]
9. Li, L.; Bai, R.; Lu, J.; Zhang, S.; Chang, C.-C. A Watermarking Scheme for Color Image Using Quaternion Discrete Fourier Transform and Tensor Decomposition. *Appl. Sci.* **2021**, *11*, 5006. [CrossRef]
10. Liao, X.; Li, K.; Yin, J. Separable data hiding in encrypted image based on compressive sensing and discrete fourier transform. *Multimed Tools Appl.* **2017**, *76*, 20739–20753. [CrossRef]
11. Hasan, N.; Islam, M.S.; Chen, W.; Kabir, M.A.; Al-Ahmadi, S. Encryption Based Image Watermarking Algorithm in 2DWT-DCT Domains. *Sensors* **2021**, *21*, 5540. [CrossRef]
12. Hazim, N.; Saeb, Z.; Hameed, K. Digital Watermarking Based on DWT (Discrete Wavelet Transform) and DCT (Discrete Cosine Transform). *Int. J. Eng. Technol.* **2019**, *7*, 4825–4829.
13. Akter, A.; Nur-E-Tajnina; Ullah, M. Digital image watermarking based on DWT-DCT: Evaluate for a new embedding algorithm. In Proceedings of the 2014 International Conference on Informatics, Electronics & Vision (ICIEV), Dhaka, Bangladesh, 23–24 May 2014. [CrossRef]
14. He, Y.; Hu, Y. A Proposed Digital Image Watermarking Based on DWT-DCT-SVD. In Proceedings of the 2018 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Xi'an, China, 25–27 May 2018; pp. 1214–1218.
15. Bogacki, P.; Dziech, A. Analysis of New Orthogonal Transforms for Digital Watermarking. *Sensors* **2022**, *22*, 2628. [CrossRef]

16. Yan, D.; Wang, R. Data Hiding for Audio Based on Piecewise Linear Haar Transform. In Proceedings of the 2008 Congress on Image and Signal Processing, Sanya, China, 27–30 May 2008; pp. 688–691.
17. Yang, L.; Hao, P.; Zhang, C. Progressive Reversible Data Hiding by Symmetrical Histogram Expansion with Piecewise-Linear Haar Transform. In Proceedings of the 2007 IEEE International Conference on Acoustics, Speech and Signal Processing—ICASSP '07, Honolulu, HI, USA, 15–20 April 2007; pp. II-265–II-268.
18. Dziech, A.; Tibken, B.; Slusarczyk, P. Image compression using periodic Haar piecewise-linear PHL transform. In Proceedings of the 2002 14th International Conference on Digital Signal Processing Proceedings, Santorini, Greece, 1–3 July 2002; Volume 2, pp. 1333–1336.
19. Abdallah, H.A.; ElKamchouchi, D.H. Signing and Verifying Encrypted Medical Images Using Double Random Phase Encryption. *Entropy* **2022**, *24*, 538. [CrossRef] [PubMed]
20. Lim, E.Y.S. Data security and protection for medical images. *Biomed. Inf. Technol.* **2008**, 249–257. [CrossRef]
21. Fornazin, M.; Netto, D.B.; Cavenaghi, M.A.; Marana, A.N. Protecting Medical Images with Biometric Information. In *Advances in Computer and Information Sciences and Engineering*; Springer: Dordrecht, The Netherlands, 2008.
22. Bouslimi, D.; Coatrieux, G. Encryption and Watermarking for medical Image Protection. In *Medical Data Privacy Handbook*; Springer: Cham, Switzerland, 2015.
23. Thakur, R.; Rohilla, R. Recent advances in digital image manipulation detection techniques: A brief review. *Forensic Sci. Int.* **2020**, *312*, 110311. [CrossRef] [PubMed]
24. Bucci, E.M. Automatic detection of image manipulations in the biomedical literature. *Cell Death Dis.* **2018**, *9*, 400. [CrossRef]
25. Wei, X.; Wu, Y.; Dong, F.; Zhang, J.; Sun, S. Developing an Image Manipulation Detection Algorithm Based on Edge Detection and Faster R-CNN. *Symmetry* **2019**, *11*, 1223. [CrossRef]
26. Yuan, G.; Hao, Q. Digital watermarking secure scheme for remote sensing image protection. *China Commun.* **2020**, *17*, 88–98. [CrossRef]
27. Zhu, P.; Jiang, Z.; Zhang, J.; Zhang, Y.J.; Wu, P. Remote Sensing Image Watermarking Based on Motion Blur Degeneration and Restoration Model. *Optik* **2021**, *248*, 168018. [CrossRef]
28. Short, N.M. Remote Sensing Tutorial: Medical Applications of Remote Sensing. Available online: https://drr.ikcest.org/remote-sensing-tutorial/introduction/Part2_26b.html (accessed on 15 September 2022).
29. Dziech, A.; Bogacki, P.; Derkacz, J. A Novel Watermark Method for Image Protection Based on Periodic Haar Piecewise-Linear Transform. In Proceedings of the International Conference on Multimedia Communications, Services and Security, Communications in Computer and Information Science, Kraków, Poland, 8–9 October 2020; Volume 1284.
30. Dziech, A.; Belgassem, F.; Nern, H.J. Image Data Compression using Zonal Sampling and Piecewise-Linear Transforms. *J. Intell. Robot. Syst.* **2000**, *28*, 61–68. [CrossRef]
31. Baran, R.; Wiraszka, D. Application of Piecewise-Linear Transforms in Threshold Compression of Contours. *Logistyka* **2015**, *4*, 2341–2348.
32. Dziech, A.; Ślusarczyk, P.; Tibken, B.R. Methods of Image Compression by PHL Transform. *J. Intell. Robot. Syst.* **2004**, *39*, 447–458. [CrossRef]
33. Kermany, D.; Goldbaum, M.; Cai, W. Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning. *Cell* **2018**, *172*, 1122–1131. [CrossRef]
34. Won, C.; Park, D.; Park, S. Efficient Use of MPEG7 Edge Histogram Descriptor. *Etri J.* **2002**, *24*, 23–30. [CrossRef]
35. Wang, Z.; Bovik, A.; Sheikh, H.; Simoncelli, E. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef] [PubMed]
36. Wang, Z.; Bovik, A.C. A universal image quality index. *IEEE Signal Process. Lett.* **2002**, *9*, 81–84. [CrossRef]
37. 'Images 4k' Dataset from Kaggle. Available online: https://www.kaggle.com/evgeniumakov/images4k (accessed on 19 September 2022).
38. Aherrahrou, N.; Tairi, H. PDE based scheme for multi-modal medical image watermarking. *Biomed Eng. Online* **2015**, *14*, 108. [CrossRef]

sensors

MDPI

Article

Material Translation Based on Neural Style Transfer with Ideal Style Image Retrieval

Gibran Benitez-Garcia [1,*], Hiroki Takahashi [1,2] and Keiji Yanai [1]

[1] Graduate School of Informatics and Engineering, The University of Electro-Communications, Chofugaoka 1-5-1, Chofu-shi 182-8585, Japan
[2] Artificial Intelligence eXploration Research Center, The University of Electro-Communications, Chofugaoka 1-5-1, Chofu-shi 182-8585, Japan
* Correspondence: gibran@ieee.org

Abstract: The field of Neural Style Transfer (NST) has led to interesting applications that enable us to transform reality as human beings perceive it. Particularly, NST for material translation aims to transform the material of an object into that of a target material from a reference image. Since the target material (style) usually comes from a different object, the quality of the synthesized result totally depends on the reference image. In this paper, we propose a material translation method based on NST with automatic style image retrieval. The proposed CNN-feature-based image retrieval aims to find the ideal reference image that best translates the material of an object. An ideal reference image must share semantic information with the original object while containing distinctive characteristics of the desired material (style). Thus, we refine the search by selecting the most-discriminative images from the target material, while focusing on object semantics by removing its style information. To translate materials to object regions, we combine a real-time material segmentation method with NST. In this way, the material of the retrieved style image is transferred to the segmented areas only. We evaluate our proposal with different state-of-the-art NST methods, including conventional and recently proposed approaches. Furthermore, with a human perceptual study applied to 100 participants, we demonstrate that synthesized images of stone, wood, and metal can be perceived as real and even chosen over legitimate photographs of such materials.

Keywords: material translation; neural style transfer; instance normalization; human perception of materials

Citation: Benitez-Garcia, G.; Takahashi, H.; Yanai, K. Material Translation Based on Neural Style Transfer with Ideal Style Image Retrieval. *Sensors* **2022**, *22*, 7317. https://doi.org/10.3390/s22197317

Academic Editor: Manuel José Cabral dos Santos Reis

Received: 19 August 2022
Accepted: 23 September 2022
Published: 27 September 2022

1. Introduction

Since the introduction of AlexNet [1] in the early 2010s, Convolutional Neural Networks (CNNs) have become the central pillar of computer vision. Over the next decade, the field gradually shifted from engineering features to designing CNN architectures. This success is attributed to more efficient graphics processing units (GPUs), new regularization techniques, and data augmentation methods to generate more training samples by deforming the available datasets. CNN architectures are now leading the performance of almost all computer vision tasks, such as detection, segmentation, and recognition of different types of objects and regions in images and videos [2,3]. On the other hand, in 2016, Gatys et al. [4] first studied how to use CNNs for applying painting styles to natural images. They demonstrated that is possible to exploit CNN feature activation to recombine the content of a given photo and the style of artwork. Specifically, a pre-trained CNN architecture is used to extract content and style features from each image. Subsequently, the resultant image is optimized by minimizing the features' distance iteratively. This work opened up the field of Neural Style Transfer (NST), which is the process of rendering image content in different styles using CNNs [5].

NST has led to interesting applications that enable transforming the reality that human beings perceive, such as photo editing, image colorization, makeup transfer, material

translation, and more [6–9]. In particular, material translation aims to transform the material of an object (from a real photograph) into a target material synthesized from the reference image (from now on, just called the style image). Consequently, the generated images can change the human perception of the objects, as shown in Figure 1. In these examples, objects made from light materials such as fabric or wood can be perceived as heavy materials, such as metal. Further, this technique can be combined with Augmented Reality (AR) and Virtual Reality (VR) devices to develop applications that generate alternate reality experiences.

Figure 1. Examples of material translation results. Metal objects translated from different original materials: wood, fabric, glass, plastic, fabric, and stone (content image in red). From left to right and top to bottom, respectively.

An important issue of material translation is that the quality of the synthesized results totally depends on the chosen style image. For example, Figure 2 shows the results of material translation from plastic to paper using three different style images. From this figure, we can see that although the style images clearly show characteristics of paper, not all translation results can be recognized as paper toys. It is worth noting that the translation has to be localized only in the object region, keeping the background unaltered. Therefore, both problems need to be tackled to achieve realistic results that can challenge the perception of original objects.

Figure 2. Examples of material translation (plastic → paper) using different style images (right corner of each generated result). The first picture shows the content image (blue), and the last is the generated image using our proposed framework (red).

In general, we can summarize three crucial aspects to achieve realistic results for material translation. The chosen style image must be able to (i) represent the target material clearly and (ii) share semantic information with the original object (similarity with the content image). Moreover, (iii) the original object must be segmented to maintain the background. Taking into account these aspects, in this paper we propose a material translation method based on NST and automatic style image retrieval. In this way, we cover the problems related to style image selection, i.e., (i) and (ii). Furthermore, we apply real-time material segmentation as a postprocessing step to fulfill (iii).

Our material translation method is defined as follows. In order to select an *ideal style image*, we firstly refine the search process by automatically choosing the most discriminative

candidate images from each material class available. Secondly, we propose to remove the style information using instance normalization whitening (IN [10]) from the query (content) and the refined images (style) of the desired material. Thus, the final search is performed using normalized CNN features extracted from the VGG19 network [11]. Finally, to translate materials to object regions, we combine semantic segmentation with NST. Specifically, we obtain pseudo labels with a weakly supervised segmentation (WSS) framework [12] to train a real-time material segmentation model [13]. Thus, we can efficiently segment target regions (objects) to translate the material of the retrieved style image.

In this paper, we employ ten different material classes shared in two publicly available datasets: Flickr Material Database (FMD [14]) and the Extended-FMD (EFMD [15]). Some examples from these datasets are shown in Figure 3. We quantitatively evaluate our work on different metrics, including: Inception Score (IS), Frechet Inception Distance (FID), classification accuracy, and segmentation performance. Qualitatively, we show examples of synthesized images that can be evaluated by visual inspection. Furthermore, we conduct a human perceptual study to evaluate the realism of the generated results. The study is designed to analyze the capacity to fool human perception by translating the original materials of target objects. One hundred participants strictly evaluated image triplets from the same material, where two were real photographs, and one was translated from a different original material. Participants are asked to choose the image that they think does not belong to the mentioned material (strict question). Thus, if they do not pick the synthesized image, it means that the translated results are real enough to fool human perception. The results of our study indicate that using our NST-based approach, it is possible to generate images that can be recognized as real even over legitimate photographs, especially for objects made of stone, wood, or metal.

Figure 3. Example images from the ten material classes used in this paper. Each row depicts images from the same class, from top to bottom: fabric, foliage, glass, leather, metal, paper, plastic, stone, water, and wood.

In our previous work [16], we tested an image retrieval method for improving material translation and found that NST is better than GAN-based generation models for material translation. Therefore, in this paper, we focus the analysis on the cutting-edge NST methods, including conventional [4,17–19] and recent approaches [20–22].

In summary, in this work, we extend our workshop paper [16] findings. Hence, the novel contributions of this paper are threefold:

- We propose a single-material translation framework based on real-time material segmentation and neural style transfer with automatic style image retrieval.
- We evaluate our proposed method with state-of-the-art (SOTA) NST methods, including Gatys [4], Johnson's [17], AdaIN [18], WCT [19], LST [20], MetaStyle [21], and STROTSS [22].
- We present a human perceptual study applied to 100 participants to evaluate the capacity of our generated results to fool the human perception of objects with translated materials.

2. Related Work

Neural Style Transfer methods can be divided in two groups: image-optimization-based and model-optimization-based [5]. The seminal work of Gatys et al. [4] is part of the first group, since the style transfer is built upon an iterative image optimization in the pixel space. Specifically, the content is defined by features extracted from multiple layers of pre-trained CNN, and the style is by terms of the Gram matrix of features extracted from another set of layers. Recently, Style Transfer by Relaxed Optimal Transport and Self-Similarity (STROTSS [22]) was proposed as an alternative to Gatys team's work. In this approach, the style is defined as a distribution over features extracted by CNN, and the distance is measured between these using an approximation of the earth mover's distance. Further, the content is defined by using local self-similarity descriptors. With these original representations of content and style, STROTSS overcame the results of Gatys, which was for a long time considered the gold standard due to its visual quality [5].

To enable faster stylization, the second group of works trains Conv–Deconv Networks using content and style loss functions to approximate the results in a single forward pass. This method was first introduced by Johnson et al. with the well-known perceptual loss function [17]. An important drawback of Johnson's approach is that an independent model must be trained for each single style image. Therefore, some approaches aim to train one single model to transfer arbitrary styles [18–20]. Huang and Belongie [18] propose adaptive instance normalization (AdaIN) to achieve real-time performance. AdaIN transfers channel-wise statics between content and style, which are modulated with affine parameters (trainable). Concurrently, Li et al. [19] propose a pair of whitening and coloring transformations (WCT) to achieve the first style learning-free method. In the same line, Linear Style Transfer (LST) [20] is proposed as an arbitrary style transfer that learns the transformation matrix with a feed-forward network and presents general solutions to the linear transformation approaches (such as AdaIN and WCT). It is known that usually arbitrary style transfer models come at the cost of compromised style transfer quality compared to single-style model methods [5]. To overcome this issue, the recent MetaStyle approach [21] formulates NST as a bi-level optimization problem, which is solvable by meta-learning methods. MetaSttyle combines arbitrary style representation learning with only a few post-processing update steps to adapt to a fast approximation model with quality comparable to that of image-optimization-based methods. A more recent approach called IFFMStyle [23] introduced invalid feature filtering modules (IFFM) to an encoder–decoder architecture for filtering the content-independent features in the original and generated images. In this way, IFFMStyle is able to transfer the style of a collection of images rather than selecting a single style image. On the other hand, Total Style Transfer [24] resolves the limitation of transferring the scale across style patterns of a style image by utilizing intra/inter-scale statistics of multi-scaled feature maps. The process is achieved by a single decoder network using skip-connections to efficiently generate stylized images. It is worth

noting that all mentioned methods from both groups can be applied to material translation. Hence, we test our framework with different SOTA NST methods to find the most suitable approach for our task.

3. Proposed Method

Matsuo et al. [9] proposed combining conventional NST [4] with a weakly semantic segmentation (WSS) approach [25] to achieve realistic material translation results. Therefore, we build upon Matsuo's framework and extend it as follows: (1) we propose automatic image retrieval rather than manually finding the *ideal style image*; (2) we employ a real-time semantic segmentation model trained with pseudo labels generated with a SOTA WSS method; and (3) we analyze different SOTA NST approaches since we previously found that NST-based strategies usually generate more realistic results than the GAN-based methods [16]. Figure 4 illustrates our proposed inference process for material translation focused on a single object (wood $\rightarrow$ foliage). As an input, we take the content image and the label of the target material. Our main contribution resides in the style image retrieval process, where we propose to apply IN whitening to remove the style information and retrieve the *ideal style image* based on its semantic similarity with the content image. Subsequently, in the material translation stage, we apply the NST approach to synthesize the material of the content image using the retrieved style. At the same time, we apply semantic segmentation on the content image to get the foreground mask depicting the material region that will be translated. Finally, the output is generated by combining synthesized and content images using the foreground mask. In the following subsections, we describe both of the main stages: Style Image Retrieval and Material Translation.

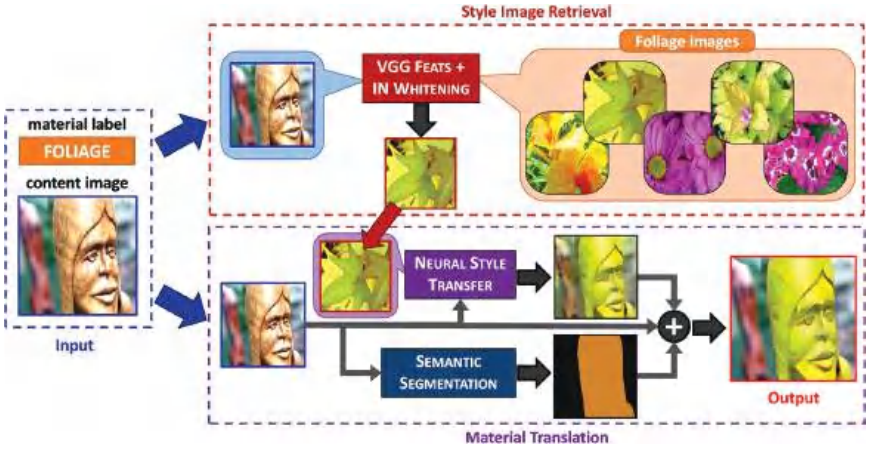


Figure 4. General overview of our proposal for material translation using style image retrieval.

3.1. Style Image Retrieval

We build our image retrieval process upon two key ideas: search refinement and style removal from CNN features. For search refinement, we assume that the *ideal style image* must reflect essential characteristics from its material while showing apparent differences from others. Therefore, to refine the style image search to the most discriminative samples, we train a CNN model to classify all possible style images from the target material to automatically choose the samples with the highest score (defined by a classification threshold Th_{clf}). Since this classification model is crucial to define high-quality candidates, we choose the robust CNN architecture of InceptionV3 [26]. Subsequently, we automatically choose the possible style images that present the widest area of the target material. To do so, we define the relative material area as the division of the material region by the image size, and we choose the samples with the most extensive regions using an area threshold Th_{area}. Note that the material region is depicted by the provided pixel annotation of the dataset or

is automatically detected by a semantic segmentation model. Then, the style image search is refined to the best-scored images with more extensive material regions from the target material. In practice, we set Th_{clf} and Th_{area} both to 0.99, so that the number of refined images drops to about 16% of samples per material. Figure 5 shows some examples of possible ideal style images that satisfy our designed requirements.

Figure 5. Fixed style images per material selected from the best-scored samples and the widest material regions. From left to right and top to bottom: fabric, foliage, glass, leather, metal, paper, plastic, stone, water, and wood.

Equally important, we employ instance normalization (IN) whitening for style removal, which was originally proposed to remove instance-specific contrast information from input images [10]. Huang et al. [27] experimentally proved that the distance between VGG [11] features of two samples is more domain-invariant when using IN whitening (experiment details on the supplementary material of [27]). In other words, the features of two images with the same content and different styles (domain) are closer in the euclidean space than those from the same style but with different contents. That is what we seek in our style image retrieval process: *to find the most similar style image based on its content (semantic) by excluding its style information.* Therefore, we build the style-free image retrieval on a VGG19, replacing all batch normalization (BN) layers with non-parametric IN. The formal definition of the IN is as follows:

$$y_{tijk} = \frac{x_{tijk} - \mu_{ti}}{\sqrt{\sigma_{ti}^2 + \epsilon}}, \tag{1}$$

where $x \in \mathbb{R}^{T \times C \times W \times H}$ is an input tensor; x_{tijk} denotes the $tijk$-th element, where k and j span spatial dimensions, i corresponds to the feature map (output from the current convolutional layer), and t is the index of the image in the batch; ϵ is an arbitrarily small constant used for numerical stability, and μ_{ti} and σ_{ti}^2, respectively are the per-instance mean and standard deviation, given by:

$$\begin{aligned} \mu_{ti} &= \frac{1}{HW} \sum_{l=1}^{W} \sum_{m=1}^{H} x_{tilm}, \\ \sigma_{ti}^2 &= \frac{1}{HW} \sum_{l=1}^{W} \sum_{m=1}^{H} (x_{tilm} - \mu_{ti})^2, \end{aligned} \tag{2}$$

where H and W represent the height and width of the feature map, respectively. It is worth noting that, different from the conventional IN layer, we exclude the affine parameters. That's why we call this process "whitening".

We L2-normalize the VGG-features from the fc7 layer before using the euclidean distance to evaluate the similarity between the content (query) and the possible style image. Finally, the image with the lowest distance is retrieved (*ideal style image*). Note that we search only within the refined images from the target material, making the retrieval process very efficient. Figure 6 shows examples of the retrieved images from different materials by using IN or BN. As can be seen, the IN version retrieves style-free images that can be useful

for material translation. Meanwhile, BN retrieves images that show apparent similarities to the content image (including color and style).

Figure 6. Retrieved results from our proposal using IN (**top**) and BN (**bottom**). From left to right: content image (stone); results of fabric, foliage, and wood materials.

3.2. Material Translation with NST

In order to design a robust and efficient material segmentation model, we first obtain pseudo labels of object regions with a WSS approach; then, we train a real-time fully supervised semantic segmentation. Subsequently, material translation is achieved in three steps: (1) material translation with NST using the *ideal style image*; (2) real-time semantic segmentation of the content image; and (3) style synthesis to the segmented regions. Each sub-process is described below.

3.2.1. Real-Time Material Segmentation

Since pixel annotation labels (semantic labels) are costly to acquire, WSS directly attacks the problem by generating segmentation labels of images given their image-level class labels. Particularly, Ahn and Kwak [12] propose to learn Pixel-level Semantic Affinity (PSA) from class activation maps (CAMs) [28] of a multi-label CNN network. The so-called AffinityNet predicts semantic affinity between a pair of adjacent image coordinates, and semantic propagation is done by a random walk. The training of AffinityNet is only supervised by the initial discriminative part segmentation (using CAMs), which is incomplete as a segmentation annotation but sufficient for learning semantic affinities within small image areas. Hence, we train AffinityNet with the Extended-FMD dataset, which contains image-level class labels only. As a result, we obtain coarse material region labels from the complete dataset (10,000 images), enough to train and fine-tune a case-specific semantic segmentation model.

On the other hand, Harmonic Densely Connected Network (HarDNet) [13] deals with real-time performance, an important issue of semantic segmentation methods. HarDNet achieves SOTA results by using harmonic densely connected blocks (HarDBlocks) instead of traditional bottleneck blocks [13]. A HarDBlock reduces most of the layer connections from a dense block, which heavily decreases concatenation cost. Moreover, the input/output channel ratio is balanced by increasing the channel width of a layer according to its connections. In particular, HarDNet for semantic segmentation is a U-shaped architecture with five encoder and four decoder blocks built of HarDBlocks. Compared to SOTA CNN architectures, HardDNet achieves comparable accuracy with significantly lower GPU runtime. Therefore, our material segmentation model is based on a HarDNet architecture. Particularly, we train a HarDNet model with the coarse labels obtained by AffinityNet. Subsequently, we fine-tune the model with the FMD dataset. Note that fine-tuning helps to enrich the quality of the HarDNet segmentation results by employing (in the supervision) the pixel-level annotations provided by the FMD dataset.

3.2.2. Material Translation

As a baseline, we use the conventional NST method from Gatys [4] for material translation, which uses a pre-trained VGG19 network to extract content and style features. The translated image is optimized by minimizing the features distance and their Gram matrices (correlation operations). Gatys et al. [4] experimentally proved that the Gram matrix of CNN activations from different layers efficiently represents the style of an image. As shown in Figure 4, we first translate the whole content image to the retrieved style. Finally, we integrate the material region of the synthesized image and the background region of the content image into the final output (I_{out}), which is defined by:

$$I_{out} = I_{gen} I_{mask} + I_{org}(1 - I_{mask}) \quad (3)$$

where I_{gen} is the synthesized image, $I_{mask} \in \{0, 1\}$ is the region mask obtained by HarDNet, and I_{org} is the content image with the original object.

4. Experimental Results

4.1. Implementation Details

We use PyTorch 1.2 with CUDA 10.2 for all experiments. For all trained methods, we used their respective pre-trained models on ImageNet [29]. AffinityNet (PSA) employed Adam as the optimization method. On the other hand, for HarDNet and InceptionV3, we used Stochastic Gradient Descent (SGD) with weight-decay 5×10^{-4} and momentum 0.9 as the optimizer. The rest of the parameters and data augmentation techniques were chosen as described in their original papers: PSA [12], HarDNet [13], and InceptionV3 [26]. The input image size for each network was 448×448, 512×512, and 299×299 for PSA, HarDNet, and InceptionV3, respectively. Note that all methods were tested with images in their original resolution (512×384 for FMD and EFMD datasets). Finally, we measured the inference time of each method (excluding I/O time) on an Intel Core i7-9700K desktop with a single NVIDIA GTX 1080Ti GPU.

4.2. Datasets

In this paper, we use two publicly available datasets: Flickr Material Database (FMD) and the Extended-FMD (EFMD). FMD [14] consists of 10 materials (fabric, foliage, glass, leather, metal, paper, plastic, stone, water, and wood). Each class contains 100 real-world images. The samples were selected manually from Flickr and were manually annotated at pixel-level. Some examples of this dataset are shown in Figure 3. EFMD [15] contains the same materials but includes 1000 images per class (10,000 in total). The samples were picked as close as possible to the FMD images, and only image-level annotations are provided. Images from both datasets are real-world RGB photographs with a size of 512×384 pixels. As shown in Table 1, for each method, we used a different number of training and testing images from both datasets. As mentioned before, HarDNet is firstly trained with the complete EFMD (HarDNet-base) and then fine-tuned with the FMD dataset (HarDNet). Note that for the material translation experiment (NST-based methods), each of the 100 testing images (10 per class) is transformed into each material class. Hence, we evaluate the NST-based methods with 1000 synthesized images.

Table 1. Number of training and testing images used for each method.

Method	Training Set	Test Set
PSA	10,000 (EFMD)	1000 (FMD)
HarDNet-base	10,000 (EFMD)	1000 (FMD)
InceptionV3	10,000 (EFMD)	1000 (FMD)
HarDNet	900 (FMD)	100 (FMD)
NST-based	-	100 (FMD)

4.3. Ablation Study

We first evaluate our proposal with classification and segmentation metrics: average accuracy (acc) and mean Intersection over the Union (mIoU). The classification accuracy shows the percentage of synthesized images that can be correctly classified with the trained InceptionV3 model. The intuition behind this evaluation is that the higher the accuracy, the better the quality of synthesized images. On the other hand, the segmentation metric presents a similar evaluation focused on pixel-level accuracy. In other words, the mIoU metric is stricter since it measures the accuracy of the translated materials by region rather than taking a single decision from the whole image.

As a baseline, we select one fixed style image per material based on the best-scored images and the widest material regions. Figure 5 shows the selected style images from each class. Note that these ten images are used to translate the entire testing set. As a result, ten translated images are generated from each content sample; hence, we evaluate 1000 synthesized images in total (100 per class). On the other hand, we apply our style image retrieval only to the refined images (about 15 per class) of the target material. We evaluate the results of our proposal by replacing the IN whitening (VGG19-IN) with BN layers (VGG19-BN) and without the normalization process (VGG19). We also evaluate the results with (w/refine) and without search refinement (w/o refine), which means searching for the ideal style image within 90 images per class.

Table 2 presents quantitative results of all variations. We observe that IN whitening significantly improves the results compared to the vanilla VGG19 and the BN (11% of accuracy and 4% of mIoU). These results concur with our hypothesis that style information must be removed from VGG features to retrieve *ideal style images*. Further, search refinement plays an essential role in the retrieving process. It boosts the material translation performance of our VGG19-IN by more than 15%. Surprisingly, the fixed-style image performance is comparable to that of the retrieving-based approaches and even outperforms the BN and vanilla VGG19 variations. This issue suggests that there is still a place for improvement in the retrieving process (to find better style images).

Table 2. Classification and segmentation evaluation of the ablation study: "w/o" and "w/ refine" refers to without and with search refinement, respectively.

	w/o Refine		**w/ Refine**	
Method	**acc**	**mIoU**	**acc**	**mIoU**
Baseline	-	-	0.556	0.4860
VGG19-IN	**0.409**	**0.3967**	**0.572**	**0.5062**
VGG19-BN	0.291	0.3612	0.543	0.4887
VGG19	0.270	0.3520	0.506	0.4845

We also evaluate per-material performance from our proposal. Figure 7 shows the average accuracy of content (translated from original material to the ten classes) and style (individually translated material from all content styles) materials. As expected, not all materials show the same level of realism after the translation process. Interesting results are those from glass and water. The former seems to be easy to synthesize but challenging to translate, while the latter presents the opposite situation. Likewise, water and leather materials are challenging to synthesize, while glass and wood are certainly easier. Furthermore, in Figure 8, we evaluate the translation performance from each pair of materials (A $\rightarrow$ B), where rows and columns represent original (content) and translated (style) materials, respectively. Stone to leather and leather to water are challenging to translate, while stone to wood and wood to plastic are more accessible.

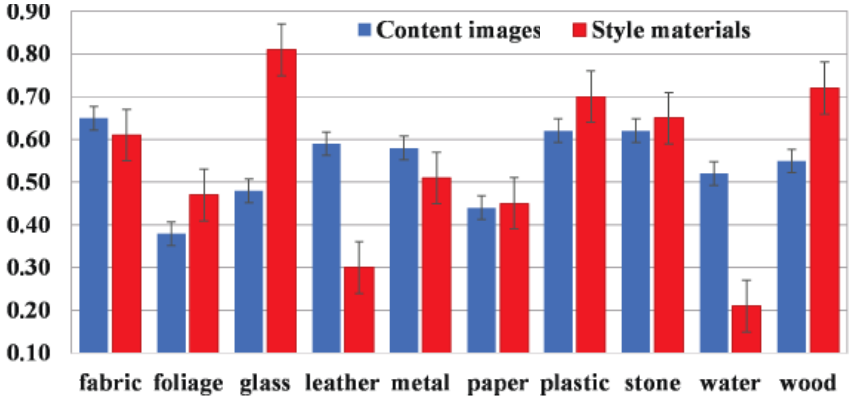


Figure 7. Classification accuracy per material class using our proposed VGG19-IN.

original materials (material A) \ translated materials (material B)	Fa	Fo	Gl	Le	Me	Pa	Pl	St	Wa	Wo
Fabric	-	64	88	27	68	62	80	72	23	62
Foliage	23	-	70	11	27	24	38	40	12	50
Glass	47	38	-	20	55	41	71	41	22	63
Leather	86	32	81	-	35	21	63	54	6	85
Metal	69	27	94	37	-	28	56	62	10	80
Paper	47	24	32	16	27	-	65	49	11	52
Plastic	68	33	86	45	73	26	-	72	30	48
Stone	71	66	87	7	49	72	74	-	23	94
Water	36	27	68	4	46	37	41	58	-	74
Wood	48	52	90	39	33	33	97	84	9	-

Figure 8. Classification accuracy (%) of translations from material A (rows) to material B (columns).

Figure 9 illustrates quantitative results of our VGG19-IN feature-based approach. We can see that all retrieved style images do not share style similarities with the content images explicitly (due to the IN whitening). Further, some of them show similar features: such as in the first example (from wood), the angular shape of the tooth-like part of the object has similar patterns on the foliage image. On the other hand, the difficulty in translating the water material might be related to the object shapes rather than the style itself. It is not natural to recognize water as certain shapes that do not exist in the real world.

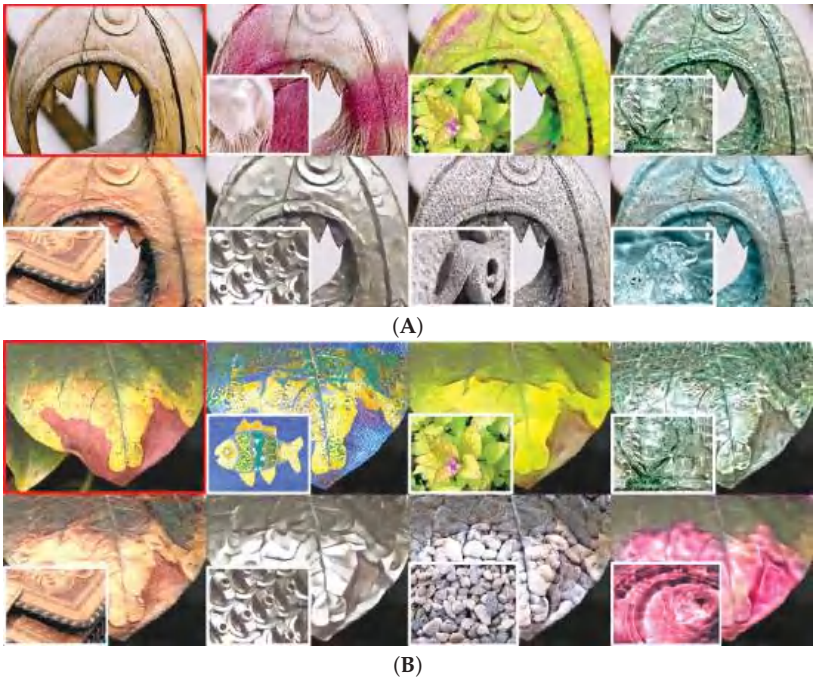


Figure 9. Translated results using our VGG19-IN proposal. (**A**) from wood, and (**B**) from foliage material (content image in red). From left to right, and top to bottom: content image, results of fabric, foliage, glass, leather, metal, stone, and water.

4.4. Comparison among SOTA NST Methods

We evaluate our approach with the conventional methods of Gatys [4] and Johnson [17]; the real-time learning-free methods of AdaIN [18], WCT [19], and LST [20]; as well as the recently proposed methods of MetaStyle [21] and STROTSS [22]. In the case of the Johnson and MetaStyle models, using the default parameters provided in their respectively open-source codes, we train ten models based on the fixed style images shown in Figure 5. For Gatys and STROTSS, we optimize each content image with its respective *ideal styles* from all materials, generating the same number of images in total (100 per material). Finally, we use the respectively pre-trained models provided by the authors of AdaIN, WCT, and LST.

In addition to the acc and mIoU, we evaluated all methods using GAN metrics, i.e., Inception Score (IS), and the Frechet Inception Distance (FID). The IS estimates the quality of the synthesized images based on how well the InceptionV3 model classifies them. This metric combines the confidence of the conditional class predictions for each synthetic image (quality) and the integral of the marginal probability of the predicted classes (diversity). However, IS does not capture how synthetic images compare to real ones. That's the main reason for introducing FID, which employs the coding layer of the InceptionV3 model to generate features from real and synthesized images. Thus, the distance between the distributions from both groups of images is then calculated using the Frechet distance. Finally, we use our pre-trained InceptionV3 model to calculate IS and FID metrics, and the final results are averaged over the 1000 synthesized images generated from the 100 content images. Note that as an accuracy score, the higher the IS is, the better. Contrarily, the smaller the FID is, the better, as it reflects the distance from real images.

Table 3 shows the results from all evaluated NST methods. As expected, the best results are obtained by the image-optimization-based approaches (Gatys and STROTSS). However, due to their iterative optimization process, these are the methods with the slowest inference time. STROTSS obtained the best FID score, which means that the translated images share stronger semantic similarities with real photographs than those translated by the Gatys method. Still, the latter gets better classification and segmentation accuracy. On the other hand, Johnson's, AdaIN, and MetaStyle are the most computationally efficient methods. Nevertheless, AdaIN can be preferred since it uses only a single model to transfer arbitrary styles. Finally, WCT and LST show similar performance, although LST is about two times faster than the former. Figure 10 shows qualitative results from all methods. We can see that in this example, almost all synthesized images show distinctive properties of the target material (stone), such as rough and porous texture rather than the polished surface of the original wood material. Even so, the results of Gatys and STROTSS look significantly more real than those of AdaIN and LST.

Table 3. Quantitative results of all evaluated NST methods. Inference time is measured on a single GTX 1080 Ti GPU.

Method	acc ↑	mIoU ↑	IS ↑	FID ↓	Inference Time ↓
Gatys [4]	**0.572**	**0.5062**	**4.181**	61.30	45.6545 s
STROTSS [22]	0.515	0.4887	4.046	**60.29**	89.1562 s
Johnson's [17]	0.506	0.4464	3.887	68.44	**0.0881 s**
MetaStyle [21]	0.442	0.4674	3.635	61.93	0.1868 s
WCT [19]	0.353	0.4079	3.604	64.53	1.0151 s
LST [20]	0.343	0.3606	3.569	62.95	0.4816 s
AdaIN [18]	0.304	0.2780	3.129	74.52	0.1083 s

Figure 10. Qualitative results from all evaluated NST methods, translating from wood to stone. From left to right and top to bottom: content image (red) and style (blue); results from Gatys, STROTSS, Johnson, MetaStyle, WCT, LST, and AdaIN.

4.5. Human Perceptual Study

Well-designed perceptual experiments with human observers are the most reliable known methodology for evaluating the quality of synthesized images. Human perceptual studies for NST approaches usually analyze the results from different NST approaches [22,27,30,31]. However, they do not assess if the generated images can be perceived as real over legit photographs of the same category. Therefore, we design a human perceptual study to analyze the capacity of the synthesized images to fool human perception by translating the original materials with our NST-based proposal.

Using the InceptionV3 model, for this study, we choose the top-6 synthesized images from each material, 60 images in total. An example of the top-6 translated results of metal are shown in Figure 1. These images are generated using the Gatys NST method and are usually translated from content images affine to the target material, as shown in

the confusion matrix of Figure 8. We present each synthesized image along with two real photographs of objects from the same material. These photographs were manually selected from the FMD dataset and considered the objects included in the 60 synthesized images. Then, users were asked to select the image that does not belong to the depicted material from the three options. An example of the user interface is shown in Figure 11. Furthermore, to avoid biased results generated from the background of the original photographs, we remove this from all images used in the study. In this way, objects only found outdoors may have the chance to be recognized if these are translated to materials that are found indoors and vice-versa.

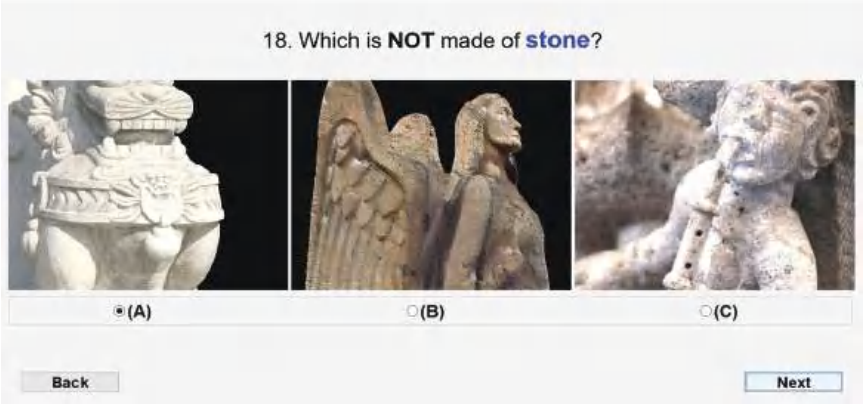


Figure 11. Human perceptual study interface: 70% of the participants chose (**A**), 18% chose (**C**), while the image generated using our approach (**B**) got only 12% of the votes.

One hundred different participants took part in this study. We randomly showed 30 questions to each of them, keeping a ratio of 3 images per material. In total, 60 different questions were defined, and the image order of the triplets was also randomized to ensure fair comparisons. Each question was answered by 50 different participants, so we collected 3000 votes in total. Unlimited time was given to select the fake image out of three options. Note that there was not an option to indicate that all photos are real. Thus, the participants were forced to carefully find the outlier image. Consequently, if they did not pick the synthesized image, it means that the translated results are real enough to fool human perception.

We counted the results when participants do not choose the synthesized images. Given that, the average results of the 3000 votes show that 44.86% of the time, participants took the translated results as representative pictures of their target material. These findings are more significant for some materials, as shown in Figure 12. Translated images from materials such as stone, wood, metal, and leather were taken as real over legitimate photographs by more than 50% of participants. In this way, we can prove that our NST-based approach can generate images that fool the human perception. Figure 13 shows some examples of the translated images that got the best acceptance in the human study. We can see that the synthesized images clearly exhibit elements from the target material, such as reflection and texture in the cases of metal and leather, respectively.

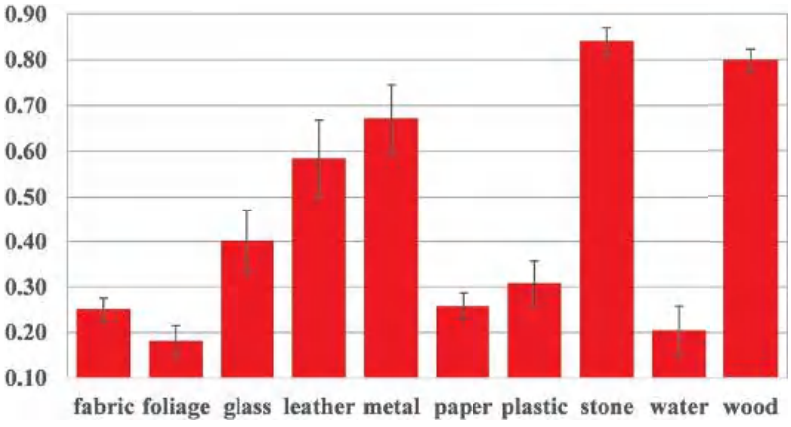


Figure 12. Realism results from the human perceptual study. *Y*-axis shows the average results when participants did not select the translated image as the outlier. Higher results represent more people being fooled by the synthesized images.

Figure 13. Examples of synthesized images with fewer votes (i.e., perceived as real). Each row shows the image triplets shown in one question (1st row: metal; 2nd row: leather). The most-voted pictures are shown from (**left**) to (**right**). The synthesized results of metal and leather got 4% and 14% of the votes, respectively (content image in red).

On the other hand, although we selected the best-scored synthesized images, the results of foliage, water, and fabric were not able to fool the human perception. Figure 14 shows some examples of these materials. As we can see, the shape and texture of the original materials (wood in both cases) limits the results for being selected over legitimate photographs, even though the texture and color of the target materials are still present (foliage and water). Finally, we believe the results for plastic, paper, and glass can become more real if the original object shares similarities with authentic objects of the target material. However, with the current study, we cannot prove this hypothesis.

Figure 14. Examples of synthesized images with more votes (i.e., perceived as fake). Each row shows the image triplets shown in one question (1st row: foliage; 2nd row: water). The most-voted pictures are shown from (**left**) to (**right**). The synthesized results of foliage and water got 88% and 85% votes of the votes, respectively (content image in red).

5. Conclusions and Future Work

In this paper, we introduced a material translation method based on real-time material segmentation and neural style transfer with automatic *ideal style image* retrieval. We build the image retrieval on VGG19 features whitened with instance normalization to remove the style information. Our results show that by excluding the style in the search process, the translated results are significantly better. We were able to translate the material of segmented objects using different NST methods, which we further analyzed quantitatively and qualitatively. Furthermore, we presented a human perceptual study to evaluate the quality of the synthesized images. The results of our study indicate that our NST-based approach can generate images of stone, wood, and metal that can be perceived as real even over legitimate photographs. Since we can alternate the material of some objects with the results being perceived as more real than fictional, we expect that our approach can be used to create alternate reality scenarios in which the user can feel a different environment based on the imperceptively modified objects.

As future work, we will further analyze different options to synthesize materials such as plastic, paper, and glass, which we believe can get more real if the original object shares similarities with authentic objects of the target material. Further, we would like to develop a real-time application that can translate the material of objects in-the-wild.

Author Contributions: Conceptualization, G.B.-G., H.T. and K.Y.; investigation, G.B.-G.; data curation, H.T.; software, G.B.-G.; supervision, H.T. and K.Y.; writing—original draft preparation, G.B.-G.; writing—review and editing, G.B.-G., H.T. and K.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by JSPS KAKENHI Grant Numbers, 21H05812, 22H00540, 22H00548, and 22K19808.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We would like to thank all the participants who selflessly took part in the human perceptual study.

Conflicts of Interest: The authors declare no conflict of interest for realization of this research. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Commun. ACM* **2017**, *60*, 84–90. [CrossRef]
2. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
3. Liu, Z.; Mao, H.; Wu, C.Y.; Feichtenhofer, C.; Darrell, T.; Xie, S. A ConvNet for the 2020s. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 19–24 June 2022; pp. 11976–11986.
4. Gatys, L.A.; Ecker, A.S.; Bethge, M. Image style transfer using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2414–2423.
5. Jing, Y.; Yang, Y.; Feng, Z.; Ye, J.; Yu, Y.; Song, M. Neural style transfer: A review. *IEEE Trans. Vis. Comput. Graph.* **2019**, *26*, 3365–3385. [CrossRef] [PubMed]
6. Siarohin, A.; Zen, G.; Majtanovic, C.; Alameda-Pineda, X.; Ricci, E.; Sebe, N. How to make an image more memorable? A deep style transfer approach. In Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval, Mountain View, CA, USA, 6–9 June 2017; pp. 322–329.
7. Yanai, K.; Tanno, R. Conditional fast style transfer network. In Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval, Mountain View, CA, USA, 6–9 June 2017; pp. 434–437.
8. Li, T.; Qian, R.; Dong, C.; Liu, S.; Yan, Q.; Zhu, W.; Lin, L. Beautygan: Instance-level facial makeup transfer with deep generative adversarial network. In Proceedings of the 26th ACM International Conference on Multimedia, Seoul, Korea, 22–26 October 2018; pp. 645–653.
9. Matsuo, S.; Shimoda, W.; Yanai, K. Partial style transfer using weakly supervised semantic segmentation. In Proceedings of the IEEE International Conference on Multimedia & Expo Workshops, Hong Kong, China, 10–14 July 2017; pp. 267–272.
10. Ulyanov, D.; Vedaldi, A.; Lempitsky, V. Instance normalization: The missing ingredient for fast stylization. *arXiv* **2016**, arXiv:1607.08022.
11. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
12. Ahn, J.; Kwak, S. Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 4981–4990.
13. Chao, P.; Kao, C.Y.; Ruan, Y.S.; Huang, C.H.; Lin, Y.L. Hardnet: A low memory traffic network. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 27 October– 2 November 2019; pp. 3552–3561.
14. Sharan, L.; Rosenholtz, R.; Adelson, E. Material perception: What can you see in a brief glance? *J. Vis.* **2009**, *9*, 784. [CrossRef]
15. Zhang, Y.; Ozay, M.; Liu, X.; Okatani, T. Integrating deep features for material recognition. In Proceedings of the 23rd International Conference on Pattern Recognition, Cancun, Mexico, 4–8 December 2016; pp. 3697–3702.
16. Benitez-Garcia, G.; Shimoda, W.; Yanai, K. Style Image Retrieval for Improving Material Translation Using Neural Style Transfer. In Proceedings of the 2020 Joint Workshop on Multimedia Artworks Analysis and Attractiveness Computing in Multimedia (MMArt-ACM '20), Dublin, Ireland, 26–29 October 2020; pp. 8–13.
17. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual losses for real-time style transfer and super-resolution. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 694–711.
18. Huang, X.; Belongie, S. Arbitrary style transfer in real-time with adaptive instance normalization. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 1501–1510.
19. Li, Y.; Fang, C.; Yang, J.; Wang, Z.; Lu, X.; Yang, M.H. Universal style transfer via feature transforms. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 386–396.
20. Li, X.; Liu, S.; Kautz, J.; Yang, M.H. Learning linear transformations for fast arbitrary style transfer. *arXiv* **2018**, arXiv:1808.04537.
21. Zhang, C.; Zhu, Y.; Zhu, S.C. Metastyle: Three-way trade-off among speed, flexibility, and quality in neural style transfer. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; Volume 33, pp. 1254–1261.
22. Kolkin, N.; Salavon, J.; Shakhnarovich, G. Style Transfer by Relaxed Optimal Transport and Self-Similarity. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 10051–10060.
23. Xu, Z.; Hou, L.; Zhang, J. IFFMStyle: High-Quality Image Style Transfer Using Invalid Feature Filter Modules. *Sensors* **2022**, *22*, 6134. [CrossRef] [PubMed]
24. Kim, M.; Choi, H.C. Total Style Transfer with a Single Feed-Forward Network. *Sensors* **2022**, *22*, 4612. [CrossRef] [PubMed]
25. Shimoda, W.; Yanai, K. Distinct class-specific saliency maps for weakly supervised semantic segmentation. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 218–234.
26. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
27. Huang, X.; Liu, M.Y.; Belongie, S.; Kautz, J. Multimodal unsupervised image-to-image translation. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018; pp. 172–189.
28. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning deep features for discriminative localization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2921–2929.

29. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. ImageNet Large Scale Visual Recognition Challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]
30. Wang, T.C.; Liu, M.Y.; Zhu, J.Y.; Tao, A.; Kautz, J.; Catanzaro, B. High-resolution image synthesis and semantic manipulation with conditional gans. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 8798–8807.
31. Chen, Q.; Koltun, V. Photographic image synthesis with cascaded refinement networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1511–1520.

sensors

MDPI

Article

Estimation of Respiratory Frequency in Women and Men by Kubios HRV Software Using the Polar H10 or Movesense Medical ECG Sensor during an Exercise Ramp

Bruce Rogers [1,*], Marcelle Schaffarczyk [2] and Thomas Gronwald [2]

[1] College of Medicine, University of Central Florida, 6850 Lake Nona Boulevard, Orlando, FL 32827-7408, USA
[2] Interdisciplinary Institute of Exercise Science and Sports Medicine, MSH Medical School Hamburg, University of Applied Sciences and Medical University, Am Kaiserkai 1, 20457 Hamburg, Germany
* Correspondence: bjrmd@knights.ucf.edu

Abstract: Monitoring of the physiologic metric, respiratory frequency (RF), has been shown to be of value in health, disease, and exercise science. Both heart rate (HR) and variability (HRV), as represented by variation in RR interval timing, as well as analysis of ECG waveform variability, have shown potential in its measurement. Validation of RF accuracy using newer consumer hardware and software applications have been sparse. The intent of this report is to assess the precision of the RF derived using Kubios HRV Premium software version 3.5 with the Movesense Medical sensor single-channel ECG (MS ECG) and the Polar H10 (H10) HR monitor. Gas exchange data (GE), RR intervals (H10), and continuous ECG (MS ECG) were recorded from 21 participants performing an incremental cycling ramp to failure. Results showed high correlations between the reference GE and both the H10 (r = 0.85, SEE = 4.2) and MS ECG (r = 0.95, SEE = 2.6). Although median values were statistically different via Wilcoxon testing, adjusted median differences were clinically small for the H10 (RF about 1 breaths/min) and trivial for the MS ECG (RF about 0.1 breaths/min). ECG based measurement with the MS ECG showed reduced bias, limits of agreement (maximal bias, −2.0 breaths/min, maximal LoA, 6.1 to −10.0 breaths/min) compared to the H10 (maximal bias, −3.9 breaths/min, maximal LoA, 8.2 to −16.0 breaths/min). In conclusion, RF derived from the combination of the MS ECG sensor with Kubios HRV Premium software, tracked closely to the reference device through an exercise ramp, illustrates the potential for this system to be of practical usage during endurance exercise.

Keywords: respiratory rate; breathing frequency; heart rate variability; endurance exercise

Citation: Rogers, B.; Schaffarczyk, M.; Gronwald, T. Estimation of Respiratory Frequency in Women and Men by Kubios HRV Software Using the Polar H10 or Movesense Medical ECG Sensor during an Exercise Ramp. *Sensors* **2022**, *22*, 7156. https://doi.org/10.3390/s22197156

Academic Editor: Manuel José Cabral dos Santos Reis

Received: 13 August 2022
Accepted: 19 September 2022
Published: 21 September 2022

1. Introduction

While respiratory frequency or breathing rate (RF) has been shown to be of value in monitoring both health and disease states, its application as an exercise measurement tool has lagged behind other internal and external load parameters such as heart rate (HR), heart rate variability (HRV), cycling, and running power [1]. However, there are scenarios in which RF estimation can be helpful in exercise science. These include ventilatory threshold measurements, assessments of work rate intensity, and decreases in exercise performance [2–4]. Though minute ventilation (the tidal volume x RF) has received attention for intensity estimation purposes [5], the respiratory rate itself has been shown to be quite sensitive in this regard as well [1]. Interestingly, the RF curve closely mimics that of the rise in lactate seen with progressive increases in exercise intensity [1]. It also appears to be a good predictor for constant intensity time to exhaustion and a means of differentiating effort/perceived exertion over long time periods [6]. Additionally, RF appears to respond to exercise load change faster than HR or VO_2, making this metric ideal for high intensity interval physiologic tracking [7].

There are a wide variety of methods to determine RF, including formal gas exchange metabolic carts, devices that track mechanical deformation of the chest wall, sounds of

breathing, and analysis of the photoplethysmogram (PPG) [8]. Unfortunately, both gas exchange carts and mechanical sensing vests can be cumbersome and costly. Photoplethysmography is problematic when used during dynamic exercise due to associated limb and body motion. However, it is possible to derive RF through the low-cost modality of HR monitoring technology [9]. One method is based on the alterations in HRV (as the variation of RR intervals) that accompany the process known as respiratory sinus arrythmia (RSA) [10]. On a simplistic basis, chest cavity expansion during inspiration induces an intrathoracic pressure drop with a secondary blood pressure reduction, leading to a reduced parasympathetic drive to the cardiac pacemaker apparatus causing heart rate to rise. Conversely, chest cavity volume contraction during expiration results in a return of parasympathetic drive and a slowing of HR [11]. Ideally, these cyclic changes in RR/HRV pattern are directly reflected to the RF and therefore measurable by HR monitoring or electrocardiogram (ECG) devices. In practice however, RF derived purely from RSA has relatively poor agreement with reference methods [9]. In an attempt to improve the accuracy of RSA-based RF, additional clues taken from the actual ECG waveform have been utilized to enhance those based on RSA alone. These methods involve several potential observations including the variation of R wave amplitude, QRS waveform analysis and/or QRS slope alteration during the breathing process [12–14].

Several reports have been published comparing the accuracy of different ECG and HRV algorithms to derive RF [13,15]. Many of these procedures are not easily reproduced by consumers, coaches, or sports professionals. Recently, one of the more popular commercial HRV software applications, Kubios HRV Premium, has been modified to calculate RF either using RR/HRV interval or ECG data recordings [16]. The HRV method is based on the cyclic cardiac beat to beat time domain changes in RR intervals associated with RSA, whereas the Kubios ECG procedure (ECG-derived RF; EDR) combines both the HRV estimation method with that of ECG-associated R wave amplitude changes seen during the respiratory cycle. To date, there has not been a published independent evaluation regarding the validity of these methods. Additionally, the question of whether adding the R wave amplitude information seen with ECG recording improves the RF estimation over simple RR interval analysis arises. In the context of comparing RF derived from two recording sources, it is also important to consider the effects of lead placement on HRV [17,18]. In other words, HRV measured from a conventional ECG lead placement may differ from that of HRV from a chest belt device. Fortunately, there are consumer HR monitoring devices with similar chest belt form factors able to accurately measure HRV and ECG waveforms, thus eliminating that particular variable. Therefore, to best compare HRV alone to that of HRV plus ECG-derived RF, we will compare data from two chest belt devices worn concurrently, the Polar H10 and Movesense Medical single-channel ECG, to a gas exchange-derived RF (GE).

2. Methods

2.1. Participants

Twenty-one participants (men: n = 12, age: 43 ± 13 years, height: 178 ± 8 cm, body weight: 83 ± 14 kg; women: n = 9, age: 35 ± 11 years, height: 169 ± 4 cm, body weight: 66 ± 10 kg) with no previous past medical history, current medications, or recent illness were recruited. They were all above 18 years of age and of any fitness level. All participants were asked to abstain from alcohol, caffeine, recreational drugs, tobacco, and vigorous exercise 24 h before testing, and provided written informed consent. Ethical approval for the study was acquired through the University of Hamburg, Department of Psychology and Movement Science, Germany (reference no.: 2021_400) and was in accordance with the principles of the Declaration of Helsinki.

2.2. Exercise Protocol and Data Recording

An incremental ramp protocol until exhaustion was performed on a mechanically braked cycle (Ergoselect 4 SN, Ergoline GmbG, Bitz, Germany) by all participants. Testing procedure included a warmup of three minutes with an initial workload of 50 watts then increasing by 1 watt every 3.6 s (equivalent to 50 watts/3 min). The exercise ramp was terminated when the participant could not maintain a cadence of 60 rpm or when they reached subjective exhaustion or a heart rate > 90% of the maximum predicted heart rate, or respiratory quotient > 1.1. Maximum oxygen uptake (VO_{2MAX}) and maximum HR (HR_{MAX}) were defined as the average VO_2 and HR over the last 30 s of the test. Recordings of RR intervals and ECG were taken continuously with two devices at the same time, the Movesense Medical sensor (firmware version 2.0.99) single-channel ECG with chest belt (Movesense, Vantaa, Finland; sampling rate: 512 Hz; app software: Movesense Showcase version 1.0.9), and the Polar H10 sensor chest belt device (Polar Electro Oy, Kempele, Finland; sampling rate: 1000 Hz; app software: Elite HRV App, Version 5.5.1). Placement of both chest belt devices was just below the pectoral muscles with a similar horizontal alignment (see 18). Gas exchange kinetics including RF were recorded with a metabolic analyzer (Quark CPET, module A-67-100-02, Cosmed, Italy; desktop software: Omnia version 1.6.5).

2.3. Data Processing

RR data .txt files were exported from the Elite HRV app then processed by Kubios HRV Premium Software version 3.5 (Biosignal Analysis and Medical Imaging Group, Department of Physics, University of Kuopio, Kuopio, Finland). Movesense Medical sensor ECG tracings were recorded by the Movesense showcase app via an iPhone, converted into .csv files and also processed by Kubios HRV Premium. Preprocessing settings were set to the default values including the RR detrending method which was kept at "smoothness priors" (Lambda = 500). The RR series was then corrected by the Kubios HRV Premium "automatic method" [19]. For RF calculation, the window width was set to 30 s with a recalculation done every 1 s (grid interval = 1 s). Data sets with artefacts >3% were excluded from analysis. A 30 s window was based on recommendations from Kubios HRV [16]. A particular RF value was therefore based on the time 15 s before and 15 s after each given time stamp. The reference RF measured by the Quark CPET (breath by breath) was exported to Microsoft Excel 365 and time aligned with both the Polar H10 and Movesense Medical ECG sensor-derived RF. Since both the Polar H10 and Movesense Medical sensor ECG RF were recalculated every 1 s for both devices, only those values that time matched the gas exchange RF values were included for analysis.

2.4. Statistics

Normal distribution of data was checked by Shapiro–Wilk testing and visual inspection of data histograms. Descriptive statistical analysis was performed for the tested variables using Microsoft Excel 365 for the calculation of means, medians, and standard deviations (SD). The agreement of the derived RF during incremental exercise was assessed via linear regression, Pearson's r correlation coefficient, coefficient of determination (R^2), standard error of estimate (SEE), and Bland–Altman plots with limits of agreement (LoA) [20]. The size of Pearson's r correlations was evaluated as follows: $0.3 \leq r < 0.5$ low, $0.6 \leq r < 0.8$ moderate, and $r \geq 0.8$ high [21]. For non-normalized data, estimates of the adjusted median difference (AMD) were calculated using the Hodges–Lehmann shift method along with Wilcoxon testing of paired groups [22]. Agreement between groups was assessed by Bland–Altman analysis, but if proportional bias was detected, regression-based calculation of mean differences and limits of agreement were performed [23]. Bland–Altman mean differences for data comparisons were expressed as the absolute difference in RF as breaths/min (b/min). Inspection of the distribution of the mean differences in the Bland–Altman analysis was performed to confirm normality. For all tests, the statistical significance was accepted as $p \leq 0.05$. Analytical statistics were performed using Microsoft Excel 365 with Real Statistics Resource Pack software (Release

7.6, copyright 2013–2021, Charles Zaiontz, www.real-statistics.com, accessed on 13 August 2022) and Analyse-it software (Leeds, UK, Version 6.01).

3. Results

During the incremental exercise ramp, participants achieved a mean VO_{2MAX} of 40.3 ± 7.9 mL/kg/min and HR_{MAX} of 176 ± 13 bpm, which was associated with a maximal power (P_{MAX}) of 260 ± 53 watts. Five participants were excluded from exercise analysis due to artefacts > 3%. These were caused by both atrial and ventricular ectopic beats that were noted in the ECG. Artifacts attributed to noise were virtually nonexistent and ECG waveforms were well shaped in the analysis group.

The total number of paired RF observations between devices was 7543 from 16 participants and the distribution of values was not normal. The level of correlation was high (Figure 1, Table 1) between the reference gas exchange device and both the Polar H10 (r = 0.85, SEE = 4.2) and Movesense Medical ECG sensor (r = 0.95, SEE = 2.6). Although median values were statistically different via Wilcoxon testing, adjusted median differences were clinically small for the Polar H10 (RF about 1 b/min) and trivial for the Movesense Medical ECG sensor (RF about 0.1 b/min). Bland–Altman plotting is shown in Figure 2. An analysis looking for both proportional bias (change in the bias over the RF range) and heteroscedasticity (change in scatter of the differences) did show significant findings for each comparison. A line of regression for the mean differences and limits of agreement was performed and displayed in Figure 2 according to the recommendations of Ludbrook [23]. Representative plots of RF over time for the three measurement modalities are shown in Figure 3. In 2 of the cases there were zero artifacts and in the other 2 the total artifacts (atrial premature beats) were below 1%.

Table 1. Mean, standard deviation (SD), median, minimum, maximum, adjusted median difference (AMD) as breaths/min (b/min) for the respiratory frequency (RF) comparison of the gas exchange (GE), Polar H10 (H10) and the Movesense Medical sensor ECG (MS ECG) data according to Hodges–Lehmann method (p-value estimated by Wilcoxon paired testing), Pearson's r and standard error of estimate (SEE) calculated from paired RF data during the incremental exercise test until voluntary exhaustion.

	GE	**H10**	**MS ECG**
Mean (b/min)	27.75	26.19	27.70
Median (b/min)	25.80	25.09	26.51
SD (b/min)	8.56	7.92	8.08
Max (b/min)	63.93	48.01	56.57
Min (b/min)	10.91	9.06	11.89
AMD (b/min)		−1.159	0.105
Wilcoxon *p* value		0.0001	0.004
Pearson's r		0.85	0.95
SEE (b/min)		4.2	2.6

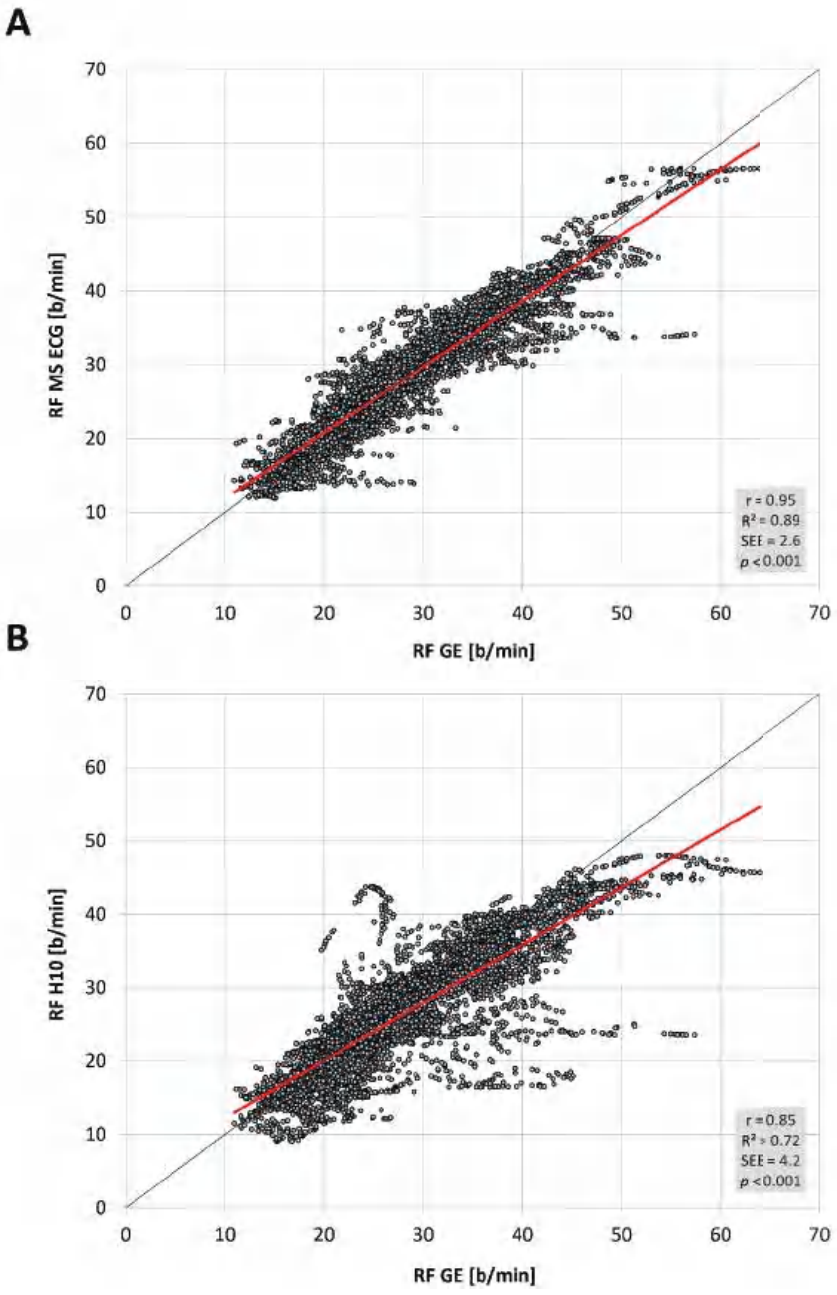


Figure 1. Regression plots for the comparison of respiratory frequency (RF) in breaths/min (b/min) for the (**A**) Movesense Medical sensor ECG (MS ECG) and the (**B**) Polar H10 sensor chest belt device (H10) vs gas exchange data (GE) during the incremental exercise test. Coefficient of determination (R^2), Pearson's r, standard error of estimate (SEE), and p value shown in the bottom right plot. Regression line in red, line of unity shown in dark grey.

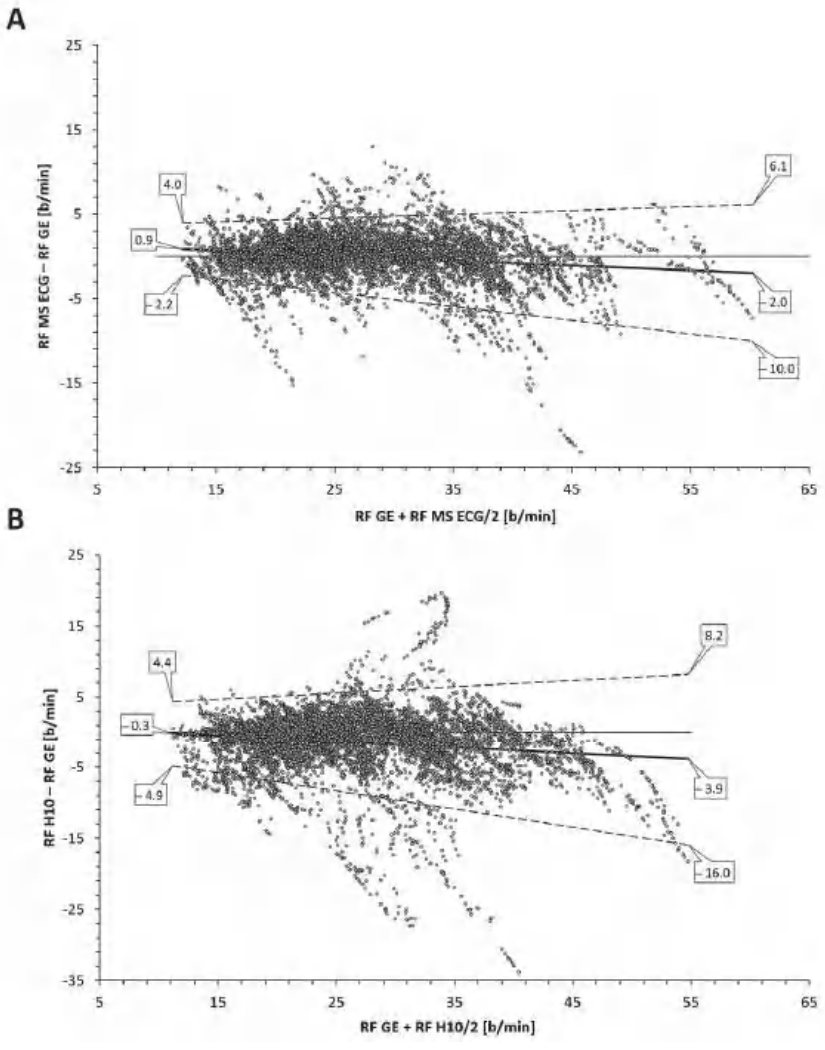


Figure 2. Bland–Altman analysis of respiratory frequency (RF) as breaths/min (b/min) for the (**A**) Movesense Medical sensor ECG (MS ECG) and the (**B**) Polar H10 sensor chest belt device (H10) vs the gas exchange data (GE) during the incremental exercise test until voluntary exhaustion. Center solid line in each plot represents the mean bias (difference) between each paired value as absolute values. The top and bottom dashed lines are LoA (1.96 standard deviations from the mean difference).

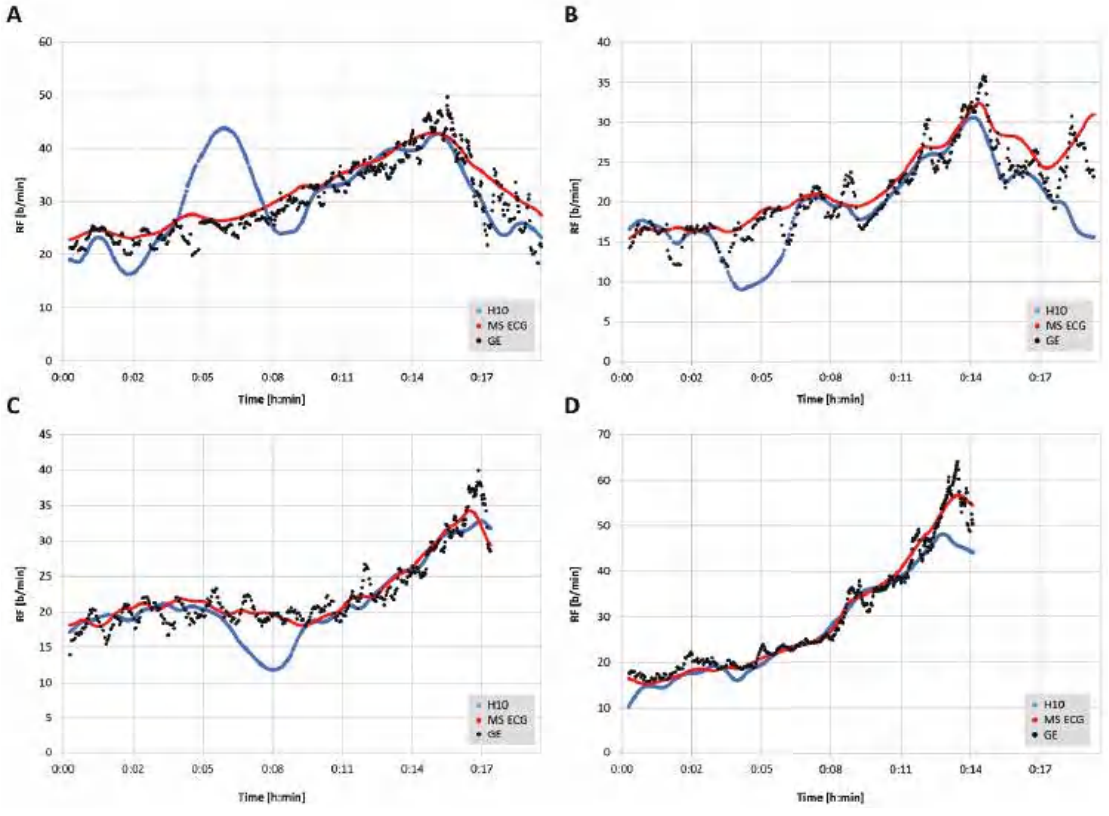


Figure 3. Respiratory frequency (RF) plotted over time for Movesense Medical sensor ECG (MS ECG), Polar H10 (H10) (Kubios window width: 30 s, grid interval: 1 s) and the gas exchange data (GE) in four representative participants. (**A**): 26-year-old female with a BMI of 30 kg/m^2, VO_{2MAX} of 38 mL/kg/min, Kubios artifact rate of 0.5%; (**B**): 27-year-old female with a BMI of 22 kg/m^2, VO_{2MAX} of 42 mL/kg/min, Kubios artifact rate of 0.0%; (**C**): 47-year-old male with a BMI of 38 kg/m^2, VO_{2MAX} of 31 mL/kg/min, Kubios artifact rate of 0.0%; (**D**): 25-year-old female with a BMI of 21 kg/m^2, VO_{2MAX} of 40 mL/kg/min, Kubios HRV artifact rate of 0.7%. MS ECG (red circle), H10 (blue circle), GE (black). Ramp termination corresponds with peak GE respiratory rate. Length of post ramp recovery determined by data recording cessation and artifacts below 3%.

4. Discussion

The aim of this study was to assess the level of agreement for RF detection between a reference gas exchange analyzer and either RR interval/HRV analysis alone (Polar H10) or a combination of RR interval analysis and ECG waveform fluctuation (Movesense Medical ECG sensor) using a commercial software application, Kubios HRV Premium. The findings show that RF derived from a combination of both RR intervals and ECG were closer to the reference values than from RR intervals only. Both correlation coefficients, adjusted median differences, Bland–Altman bias and LoAs were superior for the combination approach. In both methodologies, diminished accuracy appears to occur at higher respiratory rates and/or at the ramp termination with voluntary exhaustion. In almost all cases, this under reporting of RF was more prominent with the RR interval only method. However, for the most part, both absolute values as well as the shape of the RF over time curve were well preserved with the combination RR intervals and ECG data.

Inspection of the Bland–Altman plots did reveal significant proportional bias and heteroscedasticity (change in scatter of the differences) with both methods. With the

RR interval-based approach the mean bias varied from −0.3 to −3.9 b/min along with relatively wide LoA (maximum 8.2 to −16.0 b/min). Although the AMD was only about 1 b/min, there was a general failure to properly measure the higher RF with precision as well as more frequent outliers than with the combination approach. Improvement in agreement and correlation was seen using the combination of RR intervals and ECG data with a reduction of mean bias variation (0.9 to −2.0 b/min), LoA range (maximum 6.1 to −10.0 b/min) and a trivial AMD. Both r, R2, and SEE were superior with this method but even this algorithm still failed to fully capture the highest RF portions at ramp termination at voluntary exhaustion. This disparity in values at high RF was also seen with Kubios' own internal white paper report [16]. Their data did mirror the present findings in showing that the combination approach was superior to the RR interval only method especially in the high RF zone.

It should be noted that the Movesense Medical ECG sensor and the Polar H10 should display virtually identical base ECG waveforms, since they use the same subpectoral sensor pad placement. As background information, the Polar H10 is capable of transmitting ECG waveform data at a fixed sample rate of 130 Hz with several android and iOS applications available to read this data. This distinction is important as other studies regarding HRV indexes have shown differences based purely on the ECG lead chosen for comparison [17,18]. In the present case, each device had similar sensor electrode placements, suggesting that this issue should not be a concern. Both devices also had high sample rates at 512 and 1000 Hz, respectively, both above recommended levels [24]. In another report [14], a Polar H10 ECG waveform was upsampled (from 130 to 1000 Hz) and analyzed for RF using a combination of RR interval and ECG morphology change during various participant activities including running and cycling. Although correlations and participant specific plotting were not shown, the Bland–Altman differences were minimal with a bias of −0.5 and LoA of 2 b/min. The authors did report that the error rate was higher during running than with cycling.

From a practical standpoint, the EDR seems to yield equivalent RF patterns as the GE noted in Figure 3. Since previous studies showed potential for RF breakpoints to correspond with ventilatory thresholds [2], it would be of interest to see if EDR could achieve a similar result. A recent publication showed that both first and second ventilatory threshold identification is possible with EDR methodology [25]. This study used a Holter monitor ECG with a sample rate of 1000 Hz, and lead V6. Although we did not attempt to correlate gas exchange thresholds with the shape of the RF curve, prototypical RF over time plots shown in the cited report display many similarities to that of the Movesense Medical ECG sensor seen in Figure 3. Unfortunately, in many cases from the current report that used RR interval information only, the RF over time plot had skewed regions that would make breakpoint estimation difficult. Another endurance exercise characteristic that could be examined with EDR is the ability of RF to act as an index of "acute performance decrement (APD)" [4] as a consequence of training load. The APD appears to be a similar concept to that of athletic "durability" described as the time of onset and magnitude in deterioration in physiological-profiling characteristics over an exercise session [26]. Given the high concordance between EDR and GDR seen in this report, it seems plausible that EDR could be substituted for more equipment intensive measurements of RF. In the study overview by Passfield et al. [4], the APD was well correlated with the RF, supporting the potential of this metric to follow exercise load effects.

5. Limitations and Future Directions

Several potential limitations are apparent in using both RR interval related RSA patterns and/or ECG morphology for the purpose of RF estimation. The precision of RR measurement, amount of noise or artifact will certainly play a role in accurate delineation of any measure of HRV [27–29] and presumably derived RF as well. However, in addition to these concerns, the ECG based algorithm can be hampered by poor waveform signal strength, highlighting the need for optimal sensor pad/chest belt placement. In a similar fashion, even RR measurement can be affected by lead placement and waveform morphol-

ogy, as noted. To compound matters, the induction of body motion and muscular contractile effects will make a suboptimal waveform even more difficult to parse. It is interesting to note that in the subjects with deviation in the RR interval-derived RF seen in Figure 3, the addition of the ECG algorithm caused almost complete correction of the abnormal tracking. It is also important to realize that the data presented here represent a best-case scenario, with excellent ECG waveforms, little-to-no noise/missed beat artifacts and rare premature beats. The effects of both cardiac arrythmia, missed beats, and noisy ECG tracings are unclear. We strongly suggest that individuals inspect their ECG waveforms before testing to optimize QRS morphology (to achieve best R peak voltage) and signal-to-noise ratio. It is also of note that few validations have been performed investigating EDR during high intensity activity, let alone incremental exercise ramps until voluntary exhaustion [14]. In the future, a promising area of technology involving wearable, washable fabric sensors may provide a solution to ECG-related noise and arrhythmia issues in respiratory monitoring as well [30]. Additionally, progress in compensating for motion artifacts in PPG-derived indexes [31] may lead to better accuracy in forthcoming applications related to RF calculation [32].

The present study was performed with participants cycling indoors. Extrapolation to either outdoor cycling or other sport modalities (running, row, ski) needs to be made with caution until further validation is done. Some evidence points to subtle changes in the patterns of RF between exercise modalities such as cycling, running, and rowing that are dependent on entrainment effects [32–35]. In addition to issues related to exercise modality, higher EDR error rates were observed in participants running rather than cycling [14]. Since that specific study used a similar chest belt sensor (Polar H10), similar preprocessing and EDR methodology to Kubios HRV Premium software (Pan Tompkins R peak identification with quantification of R peak voltage combined with RR interval timing), this may indicate some limitation of using the current implementation with certain sporting activities. Regarding RF data point matching in the current study, the RR interval/EDR values were calculated over a measurement window of 30 s. It is possible that the failure to fully reach the peak RF seen with the gas exchange reference device may be related to the limited time duration of that RF. We also did not time-average the gas exchange values, which, if executed, may have led to less scattering of differences on the Bland–Altman plots. Some recommendations have been made to time-average the RF to remove effects of swallows, coughs, and sighs [1]. It was felt, that for the most part, breathing patterns during a cycling ramp would contain little of the above. Our intent was to directly compare devices with as little data manipulation as possible. Finally, the issue of device specific HRV precision should not be a factor since both the Polar H10 [36] and the Movesense Medical ECG sensor [18] have had formal RR validation studies performed.

Despite the above considerations, the degree of similarity between the EDR and GDR were impressive. Except for the slight underrepresentation of maximal values at ramp termination at voluntary exhaustion, the overall shape and agreement of the incremental rise of RF values were clinically meaningful. In the context of insights into exercise intensity assessment and threshold demarcation, the RF values seen with the Movesense Medical sensor ECG were virtually indistinguishable to that of the reference device values. With a similar form factor to a conventional chest belt monitor and only a minimal additional cost, the Movesense Medical sensor ECG appears very promising for athletic RF estimation in conjunction with Kubios HRV Premium software. One additional consideration is the cost of Kubios HRV Premium which is required for both ECG interpretation and any RF analysis. Beyond price, what are the prospects for future wearable devices (watches, cycling head units) to include incorporation of ECG-derived RF into dedicated apps which record from the Movesense Medical ECG sensor directly? Although this may appear to be unrealistic given the hardware and software constraints of mobile units, the accomplishment of real time computation of the nonlinear HRV index DFA a1 for the purpose of athletic monitoring by several apps [37] illustrates what is potentially possible with skillful software design.

6. Conclusions

The ability of a commercial HRV software package, Kubios HRV Premium, to estimate respiratory frequency throughout an exercise ramp was assessed in two consumer heart rate monitoring devices, the Polar H10 and Movesense Medical ECG sensor. Bland–Altman analysis, linear regression, and adjusted median differences indicate that the ECG centric system (single-channel chest belt ECG plus Kubios HRV Premium ECG algorithm) is superior to that of RR interval-derived respiratory frequency. The ECG based methodology also captured the pattern and shape of the respiratory frequency rise over time during the incremental ramp, whereas the RR interval-based system displayed variable accuracy especially at high exercise intensities. Future confirmation of these findings needs to be carried out with other exercise modalities as well as evaluation of the effects of artifact and noise. However, the use of commercially available software and hardware for the purpose of respiratory frequency monitoring appears promising.

Author Contributions: Conceptualization, B.R. and T.G.; physiologic testing, M.S. writing—original draft preparation, B.R.; data analysis, B.R.; writing—review and editing, B.R., M.S. and T.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Ethical approval for the study was acquired through the University of Hamburg, Department of Psychology and Movement Science, Germany (reference no.: 2021_400) and was in accordance with the principles of the Declaration of Helsinki.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Conflicts of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

ECG:	Electrocardiogram
EDR:	ECG derived respiration
HRV:	Heart rate variability
GE:	Gas exchange
PPG:	Photoplethysmogram
RF:	Respiratory frequency
RR interval:	Time between 2 successive R peaks in the ECG
RSA:	Respiratory sinus arrythmia

References

1. Nicolò, A.; Massaroni, C.; Passfield, L. Respiratory Frequency during Exercise: The Neglected Physiological Measure. *Front. Physiol.* **2017**, *8*, 922. [CrossRef] [PubMed]
2. Cross, T.J.; Morris, N.R.; Schneider, D.A.; Sabapathy, S. Evidence of break-points in breathing pattern at the gas-exchange thresholds during incremental cycling in young, healthy subjects. *Eur. J. Appl. Physiol.* **2012**, *112*, 1067–1076. [CrossRef] [PubMed]
3. Nicolò, A.; Marcora, S.M.; Sacchetti, M. Respiratory frequency is strongly associated with perceived exertion during time trials of different duration. *J. Sports Sci.* **2016**, *34*, 1199–1206. [CrossRef] [PubMed]
4. Passfield, L.; Murias, J.M.; Sacchetti, M.; Nicolò, A. Validity of the Training-Load Concept. *Int. J. Sports Physiol. Perform.* **2022**, *17*, 507–514. [CrossRef]
5. Gastinger, S.; Sorel, A.; Nicolas, G.; Gratas-Delamarche, A.; Prioux, J. A comparison between ventilation and heart rate as indicator of oxygen uptake during different intensities of exercise. *J. Sports Sci. Med.* **2010**, *9*, 110–118.
6. Pires, F.O.; Lima-Silva, A.E.; Bertuzzi, R.; Casarini, D.H.; Kiss, M.A.; Lambert, M.I.; Noakes, T.D. The influence of peripheral afferent signals on the rating of perceived exertion and time to exhaustion during exercise at different intensities. *Psychophysiology* **2011**, *48*, 1284–1290. [CrossRef]
7. Nicolò, A.; Marcora, S.M.; Bazzucchi, I.; Sacchetti, M. Differential control of respiratory frequency and tidal volume during high-intensity interval training. *Exp. Physiol.* **2017**, *102*, 934–949. [CrossRef]

8. Nicolò, A.; Massaroni, C.; Schena, E.; Sacchetti, M. The Importance of Respiratory Rate Monitoring: From Healthcare to Sport and Exercise. *Sensors* **2020**, *20*, 6396. [CrossRef]
9. Charlton, P.H.; Birrenkott, D.A.; Bonnici, T.; Pimentel, M.A.F.; Johnson, A.E.W.; Alastruey, J.; Tarassenko, L.; Watkinson, P.J.; Beale, R.; Clifton, D.A. Breathing Rate Estimation from the Electrocardiogram and Photoplethysmogram: A Review. *IEEE Rev. Biomed Eng.* **2018**, *11*, 2–20. [CrossRef]
10. Blain, G.; Meste, O.; Bermon, S. Influences of breathing patterns on respiratory sinus arrhythmia in humans during exercise. *Am. J. Physiol. Heart Circ. Physiol.* **2005**, *288*, H887–H895. [CrossRef]
11. Grossman, P.; Karemaker, J.; Wieling, W. Prediction of tonic parasympathetic cardiac control using respiratory sinus arrhythmia: The need for respiratory control. *Psychophysiology* **1991**, *28*, 201–216. [CrossRef]
12. Arunachalam, S.P.; Brown, L.F. Real-time estimation of the ECG-derived respiration (EDR) signal using a new algorithm for baseline wander noise removal. In Proceedings of the 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Minneapolis, MN, USA, 3–6 September 2009; pp. 5681–5684. [CrossRef]
13. Varon, C.; Morales, J.; Lázaro, J.; Orini, M.; Deviaene, M.; Kontaxis, S.; Testelmans, D.; Buyse, B.; Borzée, P.; Sörnmo, L.; et al. A Comparative Study of ECG-derived Respiration in Ambulatory Monitoring using the Single-lead ECG. *Sci. Rep.* **2020**, *10*, 5704. [CrossRef]
14. Alikhani, I.; Noponen, K.; Hautala, A.; Ammann, R.; Seppänen, T. Spectral fusion-based breathing frequency estimation; experiment on activities of daily living. *BioMed Eng. OnLine* **2018**, *17*, 99. [CrossRef]
15. Sobron, A.; Romero, I.; Lopetegi, T. Evaluation of methods for estimation of respiratory frequency from the ECG. In Proceedings of the 2010 Computing in Cardiology, Belfast, UK, 26–29 September 2010; pp. 513–516.
16. Accuracy of Kubios HRV Software Respiratory Rate Estimation Algorithms. Available online: https://www.kubios.com/downloads/RESP_white_paper.pdf (accessed on 2 July 2022).
17. Jeyhani, V.; Mantysalo, M.; Noponen, K.; Seppanen, T.; Vehkaoja, A. Effect of Different ECG Leads on Estimated R-R Intervals and Heart Rate Variability Parameters. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; IEEE: Piscataway Township, NJ, USA, 2019; Volume 2019, pp. 3786–3790.
18. Rogers, B.; Schaffarczyk, M.; Clauß, M.; Mourot, L.; Gronwald, T. The Movesense Medical Sensor Chest Belt Device as Single Channel ECG for RR Interval Detection and HRV Analysis during Resting State and Incremental Exercise: A Cross-Sectional Validation Study. *Sensors* **2022**, *22*, 2032. [CrossRef]
19. Lipponen, J.A.; Tarvainen, M.P. A robust algorithm for heart rate variability time series artefact correction using novel beat classification. *J. Med. Eng. Technol.* **2019**, *43*, 173–181. [CrossRef]
20. Bland, J.M.; Altman, D.G. Measuring agreement in method comparison studies. *Stat. Methods Med. Res.* **1999**, *8*, 135–160. [CrossRef]
21. Chan, Y.H. Biostatistics 104: Correlational analysis. *Singap. Med. J.* **2003**, *44*, 614–619.
22. Jiang, X.; Guo, X.; Zhang, N.; Wang, B.; Zhang, B. Robust multivariate nonparametric tests for detection of two-sample location shift in clinical trials. *PLoS ONE* **2018**, *13*, e0195894. [CrossRef]
23. Ludbrook, J. Confidence in Altman-Bland plots: A critical review of the method of differences. *Clin. Exp. Pharmacol. Physiol.* **2010**, *37*, 143–149. [CrossRef]
24. Kwon, O.; Jeong, J.; Bin Kim, H.; Kwon, I.H.; Park, S.Y.; Kim, J.E.; Choi, Y. Electrocardiogram Sampling Frequency Range Acceptable for Heart Rate Variability Analysis. *Healthc. Inform. Res.* **2018**, *24*, 198–206. [CrossRef]
25. García, D.; Kontaxis, S.; Hernández-Vicente, A.; Hernando, D.; Milagro, J.; Pueyo, E.; Garatachea, N.; Bailon, R.; Lázaro, J. Ventilatory Thresholds Estimation Based on ECG-derived Respiratory Rate. In Proceedings of the 2021 Computing in Cardiology (CinC), Brno, Czech Republic, 13–15 September 2021; pp. 1–4. [CrossRef]
26. Maunder, E.; Seiler, S.; Mildenhall, M.J.; Kilding, A.E.; Plews, D.J. The Importance of 'Durability' in the Physiological Profiling of Endurance Athletes. *Sports Med.* **2021**, *51*, 1619–1628. [CrossRef]
27. Stapelberg, N.J.C.; Neumann, D.L.; Shum, D.H.K.; McConnell, H.; Hamilton-Craig, I. The sensitivity of 38 heart rate variability measures to the addition of artefact in human and artificial 24-hr cardiac recordings. *Ann. Noninvasive Electrocardiol.* **2018**, *23*, e12483. [CrossRef]
28. Rincon Soler, A.I.; Silva, L.E.V.; Fazan, R., Jr.; Murta, L.O., Jr. The impact of artefact correction methods of RR series on heart rate variability parameters. *J. Appl. Physiol.* **2018**, *124*, 646–652. [CrossRef]
29. Rogers, B.; Giles, D.; Draper, N.; Mourot, L.; Gronwald, T. Influence of Artefact Correction and Recording Device Type on the Practical Application of a Non-Linear Heart Rate Variability Biomarker for Aerobic Threshold Determination. *Sensors* **2021**, *21*, 821. [CrossRef]
30. Dong, H.; Sun, J.; Liu, X.; Jiang, X.; Lu, S. Highly Sensitive and Stretchable MXene/CNTs/TPU Composite Strain Sensor with Bilayer Conductive Structure for Human Motion Detection. *ACS Appl. Mater. Interfaces* **2022**, *14*, 15504–15516. [CrossRef]
31. Lam, E.; Aratia, S.; Wang, J.; Tung, J. Measuring Heart Rate Variability in Free-Living Conditions Using Consumer-Grade Photoplethysmography: Validation Study. *JMIR Biomed. Eng.* **2020**, *5*, e17355. [CrossRef]
32. Jarchi, D.; Charlton, P.; Pimentel, M.; Casson, A.; Tarassenko, L.; Clifton, D.A. Estimation of respiratory rate from motion contaminated photoplethysmography signals incorporating accelerometry. *Healthc. Technol. Lett.* **2019**, *21*, 19–26. [CrossRef]
33. Power, G.A.; Handrigan, G.A.; Basset, F.A. Ventilatory response during an incremental exercise test: A mode of testing effect. *Eur. J. Sport Sci.* **2012**, *12*, 491–498. [CrossRef]

34. Elliott, A.D.; Grace, F. An examination of exercise mode on ventilatory patterns during incremental exercise. *Eur. J. Appl. Physiol.* **2010**, *110*, 557–562. [CrossRef] [PubMed]
35. Siegmund, G.P.; Edwards, M.R.; Moore, K.; Tiessen, D.A.; Sanderson, D.J.; McKenzie, D.C. Ventilation and locomotion coupling in varsity male rowers. *J. Appl. Physiol.* **1999**, *87*, 233–242. [CrossRef] [PubMed]
36. Gilgen-Ammann, R.; Schweizer, T.; Wyss, T. RR interval signal quality of a heart rate monitor and an ECG Holter at rest and during exercise. *Eur. J. Appl. Physiol.* **2019**, *119*, 1525–1532. [CrossRef] [PubMed]
37. Rogers, B.; Gronwald, T. Fractal Correlation Properties of Heart Rate Variability as a Biomarker for Intensity Distribution and Training Prescription in Endurance Exercise: An Update. *Front. Physiol.* **2022**, *13*, 879071. [CrossRef] [PubMed]

sensors

MDPI

Article

Automatic Stones Classification through a CNN-Based Approach

Mauro Tropea [1,*], Giuseppe Fedele [1], Raffaella De Luca [2], Domenico Miriello [2] and Floriano De Rango [1]

[1] Department of Informatics, Modeling, Electronics and Systems Engineering (DIMES), University of Calabria, Via P. Bucci, 87036 Rende, Italy
[2] Department of Biology, Ecology and Earth Sciences (DiBEST), University of Calabria, Via P. Bucci, 87036 Rende, Italy
* Correspondence: mtropea@dimes.unical.it; Tel.: +39-0984-494786

Abstract: This paper presents an automatic recognition system for classifying stones belonging to different Calabrian quarries (Southern Italy). The tool for stone recognition has been developed in the SILPI project (acronym of "*Sistema per l'Identificazione di Lapidei Per Immagini*"), financed by POR Calabria FESR-FSE 2014-2020. Our study is based on the *Convolutional Neural Network*s (CNNs) that is used in literature for many different tasks such as speech recognition, neural language processing, bioinformatics, image classification and much more. In particular, we propose a two-stage hybrid approach based on the use of a model of *Deep Learning* (DL), in our case the CNN, in the first stage and a model of *Machine Learning* (ML) in the second one. In this work, we discuss a possible solution to stones classification which uses a CNN for the feature extraction phase and the *Softmax* or *Multinomial Logistic Regression* (MLR), *Support Vector Machine* (SVM), *k-Nearest Neighbors* (kNN), *Random Forest* (RF) and *Gaussian Naive Bayes* (GNB) ML techniques in order to perform the classification phase basing our study on the approach called *Transfer Learning* (TL). We show the image acquisition process in order to collect adequate information for creating an opportune database of the stone typologies present in the Calabrian quarries, also performing the identification of quarries in the considered region. Finally, we show a comparison of different DL and ML combinations in our Two-Stage Hybrid Model solution.

Keywords: *Deep Learning* (DL); *Convolutional Neural Network* (CNN); *Machine Learning* (ML); *Softmax*; *Support Vector Machine* (SVM); *k-Nearest Neighbors* (kNN); *Random Forest* (RF); *Gaussian Naive Bayes* (GNB); Two-Stage Hybrid Model

Citation: Tropea, M.; Fedele, G.; De Luca, R.; Miriello, D.; De Rango, F. Automatic Stones Classification through a CNN-Based Approach. *Sensors* **2022**, *22*, 6292. https://doi.org/10.3390/s22166292

Academic Editor: Hsiao-Chun Wu

Received: 23 June 2022
Accepted: 18 August 2022
Published: 21 August 2022

1. Introduction

In the course of evolution, humans have developed complex skills to adapt to the surrounding environment and act on the basis of what has been observed. Depending on the situation, we are able to decide the most appropriate behavior to use according to a certain pattern, which can be, for example, recognizing a face, understanding another person's words, reading handwriting or distinguishing fresh food from its smell. The development of technology and the exponential improvement of computational sciences have made it possible to create computer learning software. This software acts by recognizing a certain scheme, depending on the application. *Pattern Recognition* (PR) is a branch of *Artificial Intelligence* (AI) that focuses on the recognition of patterns, forms and classifications in data by a computer. It is closely related to *Machine Learning* (ML), data mining and the discovery of knowledge. It aims to classify objects into a number of categories or classes. The main phase of a PR process concerns the *Feature Extraction and Classification*. Its goal is to characterize the data to be recognized by metrics that will provide the same results for the data in the same category and different results for the data in different categories. This leads to finding distinctive features that are invariant to any data transformation (ideally). The degree of classification of the input into different categories varies according to the characteristics of the data. In this work, we used the *Convolutional Neural Network* (CNN) to

perform the feature extraction phase which is one of the most important steps in the PR, and the TL [1] approach to avoid creating our network from scratch. In particular, we used a Two-Stage Hybrid Model solution that joins the use of a *Deep Learning* (DL) technique, a CNN model for the feature extraction phase, with a classical ML algorithm in order to perform image classification. We used four different CNNs, each one implementing five types of ML algorithms for classification: the *Softmax* or *Multinomial Logistic Regression* (MLR) [2], the *Support Vector Machine* (SVM) [3], the *k-Nearest Neighbors* (kNN) [4], the *Random Forest* (RF) [5] and the *Gaussian Naive Bayes* (GNB) [6]. We have obtained the confusion matrix of the performed object recognition for each type of used algorithm. Finally, we have presented a comparison between these algorithms in order to show the performances of each approach. In this scenario, the contribution of this paper is to give some indications into the development of a system for automatic stones classification from the Calabrian quarries. The tool for stone recognition has been developed in the SILPI project (acronym of "*Sistema per l'Identificazione di Lapidei Per Immagini*"), financed by POR Calabria FESR-FSE 2014-2020 [7]. The characterization and the determination of the provenance of stone materials, generally, represent a very long and complex process that requires not only the use of destructive and expensive diagnostic techniques, but also a specialized staff with scientific and technical know-how who are able to interpret and process the compositional data obtained from the analyses. Instead, the system developed in this project is intended to be a tool that can be easily used by non-geologists (such as restorers, archaeologists, architects, engineers, diagnostics and art historians) by helping them to solve problems about the provenance and the classification of stone materials. The system, based on image processing, is developed using rocks sampled from different Calabrian quarries, some of which were used in historical times for the construction of artifacts of historical and archaeological interest [8–10].

The main contributions of this work are listed in the following:

- The paper proposes a methodology to be used in the stone recognition context of the main Calabrian quarries that, to the best of our knowledge, represents the first attempt in the stone literature;
- The paper proposes to use in the context of stone classification a Two-Stage Hybrid Model that joins the DL approaches with ML algorithms;
- The paper shows a set of experiments by which it is possible to take out some considerations on the best combination of DL plus ML techniques to be used in the stone recognition task.

The remainder of the paper is organized as follows. After a review of related literature (Section 2), we give a description of the materials used in our research (Section 3). A brief description of pre-trained CNN models and classification methods are provided in Section 4. Section 5 provides an introduction of the Two-Stage Hybrid Model composed of a CNN network followed by a traditional ML algorithm. Section 6 describes the experiments to evaluate the performance of the provided Two-Stage Hybrid Model showing the achieved results. Section 7 concludes the paper with some final considerations.

2. Related Work

In the last few years, many researchers focused their studies on the DL approach for many different tasks. In particular, the attention has been concentrated on the CNN that represents an important technique able to resolve many different issues regarding different aspects such as speech recognition, natural language processing, bioinformatics, and image classification [11]. Our attention is focused on image recognition issues and, in particular, our application domain regards stone recognition. Many different works exist in literature about stone classification through image processing and many works exist on neural network and DL approaches applied to this domain. In the remainder of this section, we show the main works in order to contextualize our research.

2.1. Convolutional Neural Network (CNN) for Classification

Many papers have faced the topic of image processing and classification using DL and *Convolutional Neural Network* (CNN) solutions. In [12], an evaluation of an image classifier using traditional computer vision and DL approaches is provided. They use an Inception-V3 architecture and their own CNN called TinyNet built from scratch. The accuracy and loss attributes are provided as a result of the evaluation. In [13] the use of DL approach for image classification is provided. The authors analyzed and implemented a VGG-16 model for performing image classification into different categories. Moreover, they provide a methodology for more accurate classification of images. In [14] and in [15] the authors show the use of the CNN approach for visual object recognition using only SVM, in the first case, and Softmax and SVM classifiers, in the second one. Moreover, the authors of the second paper demonstrate a small but consistent advantage of replacing the Softmax layer with a linear SVM. A work based on pedestrians using CNN and SVM techniques is proposed in [16]. Their tests show that the proposal is able to quickly and reliably detect the pedestrian targets on the Caltech data set. In [17] the authors propose an image classification model applied for identifying the display of the online advertisement using a *Convolutional Neural Network* (CNN). The proposed CNN considers two parameters (n, m) where n is a number of layers and m is the number of filters in convolutional layers that are chosen on the basis of a series of experiments that they present in the paper. In [18] the authors investigate the use of a deep convolutional neural network CNN for scene classification. They experiment with two simple and effective strategies to extract CNN features, first using pre-trained CNN models as universal feature extractors, and then, domain-specifically fine-tuning pre-trained CNN models on their scene classification dataset. In [19] the authors propose a CNN architecture using the MNIST handwritten dataset in order to validate it. They utilize an optimized hardware architecture with reduced arithmetic operations and faster computations implemented on an FPGA accelerator. Another paper focusing on computational architecture is [20]. The authors implement an image classification CNN using a multi-thread GPU on the CIFAR10 dataset. In [21] the authors deal with the problem of synthetic aperture radar (SAR) image classification. They design a deep CNN architecture proposing a microarchitecture called Compress Unit (CU). Their architecture, compared with other networks for SAR classification in literature, results in being more performed and efficient. Other works exist that compare classification approaches in order to show the best choice for their applicative domain. An investigation on supervised classification is in [22] where the authors evaluate the performances of two classifiers as well as two feature extraction techniques: Linear SVM and Quadratic SVM. An exploration of the hybrid CNN solution for image classification is provided in [23] where the authors provide a comparative study of seven CNN-based hybrid image classification techniques showing the results in terms of their accuracy. A specific topic of butterfly recognition is studied in [24]. The power of DL approaches has shown the capability of the CNN of discovering with accurate results the different varieties of these insects. They propose two CNN approaches building from scratch their neural model able to classify butterfly images. A problem of plant classification is analyzed in [25] through the use of two different hybrid CNN models implemented by the authors from scratch. They used three different datasets, namely LeafSnap, Flavia, and MalayaKew Dataset utilizing the data augmentation approach for better performing the training phase. Their study shows good results for the proposed models in terms of accuracy.

2.2. Stone Classification

A lot of works exist on this topic in literature. Many researchers face the stone classification issue taking into account many different approaches that involve earth science and the mining industry. In [26] the authors have presented some possible approaches to the development of an expert system for the automatic classification of granite tiles. Based on recent results on color texture analysis, they have proposed a set of visual descriptors which provide good classification accuracy with a limited number of features.

In [27] the authors investigate the problem of choosing adequate color representation for automated surface grading. Moreover, they discuss the pros and cons of different color spaces basing their study on a dataset of 25 classes of natural stone. In [28] the authors describe a methodology for a correct and automated granite identification and classification by processing spectral information captured by a spectrophotometer at various stages of processing using functional ML techniques. In [29] the authors describe an approach for texture classification on a dateset of different stones. They have worked on extracting statistical features from histogram of grain components. So, they have provided a computable feature vector which has most meaningful information of texture. In [30] a novel approach to rotation and scale invariant texture classification is introduced. The proposed approach is based on Gabor filters that have the capability to collapse the filter responses according to the scale and orientation of the textures. Their experiments have shown the goodness of the proposed approach compared with other methods existing in the literature. In [31] the authors deal with the texture classification issues. In this paper, the authors propose an approach that uses both the Gabor wavelet and the curvelet transforms on the transferred regular shapes of the image regions. They show some experiments on texture classification demonstrating the effectiveness of the proposed approach.

A computer-vision-based methodology for the purpose of gemstone classification on 68 different classes of gemstones is provided in [32]. The authors utilize a series of feature extraction techniques used in combination with different ML algorithms. Moreover, they also use a DL classification with two ResNet models: ResNet18 and ResNet50. They provide results of classification methods against three expert gemmologists with at least 5 years of experience in gemstone identification showing the difference in time response between human and automatic approaches.

Other literature works that use the DL approach for automatic stone classification is [33], where automatic recognition and classification of granite tiles is the object of study using CNN networks such as AlexNet and VGGNet for a fine-tuning pre-trained approach, or [34] where the authors implement a classification model of ornamental rocks through the analysis and classification of images, using machine learning algorithms.

The use of the *Transfer Learning* (TL) approach for mineral microscopic image classification is reported in [35]. The authors show the system behavior using four mineral image features extracted by an Inception-V3 CNN network. Moreover, the features extracted are used for classification purposes throughout different ML methods such as: *Logistic Regression* (LR), *Support Vector Machine* (SVM), *Random Forest* (RF), *k-Nearest Neighbors* (kNN), *Multilayer Perceptron* (MLP), and GNB. As a result, they found that LR, SVM, and MLP have a significant performance among all the models, with accuracy of about 90.0%. This last contribution, which is one of the literature works used for conceiving our idea, is of proposing a hybrid model composed of two stages based on DL and ML approaches: the first one used for feature extraction and the second one used for performing stone classification. So, on the basis of these literature works we have proposed a methodology and a model to be used in the context of stone recognition proposing the joining use of four different CNNs and five different ML classification algorithms, also showing the best combination to be used.

2.3. Main Paper Contributions

This literature review presents the scientific community effort in this research field, also showing how the new AI approaches are largely used in the context of stone classification. From this study, it emerges that many researchers propose DL- or ML-based approaches but our Two-Stage Hybrid Model is distinguished for the provided methodology/modeling and represents a good solution for image recognition. Many studies deal with stone recognition using classical approaches based on texture and color space that represent very complex and resource consuming techniques. Other works introduce approaches based on AI techniques, but no one considers more CNNs (four CNN models) combined with different classifier algorithms. So, on the basis of these studies, in this work we propose a

system model to be used in stone classification based on a hybrid approach that consists of a two-stage model in which, in the first stage, we apply the use of the DL approach based on four different CNN networks and, in the second stage we propose the use of ML techniques in order to perform image classification. Our study proposes a methodology and a modeling that can be used in different contexts of stone classification. Moreover, it uses the well-known TL approach in the first stage, in order to take advantage of feature extraction based on a large image database as ImageNet, passing this information to the second stage that, based on ML algorithms, performs the classification. The TL approach permits the avoidance of creating a CNN from scratch, making the project less complex and onerous in terms of time and resources.

So, in the following, the main contributions of this work are listed:

- Stone recognition of the main Calabrian quarries that, to the best of our knowledge, represents the first attempt in the stone literature;
- Two-Stage Hybrid Model proposal able to join the DL approaches with ML algorithms;
- Methodology for stone classification purpose giving indications to face with this specific task;
- Experimental tests for providing the best combination of DL and ML techniques to be used in the stone recognition task.

3. Materials

If we compare the quarries of stone materials currently exploited in Calabria with those known until the early 1900s and reported by [9], we find that today at least 70% of the historical quarries in Calabria are no longer exploited. Moreover, most of them have totally lost their historical knowledge and exact location. Other studies [10,36], recently conducted by the Calabrian Superintendence, show evidence of ancient quarries, located mostly on the coastal areas of Calabria, dating back to the Hellenistic and Roman period. This shows that Calabria, since ancient times, has been for many civilizations the place of preferential supply of stone materials used to realize artistic artifacts and ancient architectural buildings.

An easy-to-use tool, capable of identifying the quarries with which an ancient stone artifact was made, would make an important contribution to historical knowledge of trade relations between peoples of the same period. For this reason, it was decided to work on the most representative stone materials of the Calabria Region (Southern Italy), from the five provinces of Calabria (provinces of Reggio Calabria, Vibo Valentia, Catanzaro, Cosenza and Crotone). The location of the quarries is shown in Figure 1.

3.1. Stone Materials in Calabrian Provinces

The studied stone materials come from 25 quarries; 10 samples, representative of the geological outcrops, were sampled for each quarry. Figure 2 shows the 25 types of stone materials used for the classification in our Two-Stage Hybrid Model. Table 1 shows the historical name of the stone, the city in which the quarry is located and the geological classification of the stone. The studied rocks include magmatic rocks such as granodiorites, diorites and porphyrites, sedimentary rocks, such as sandstones, calcarenites and limestones, but also metamorphic rocks, such as marbles, schists, metabasites and serpentinites (Figure 2 and Table 1).

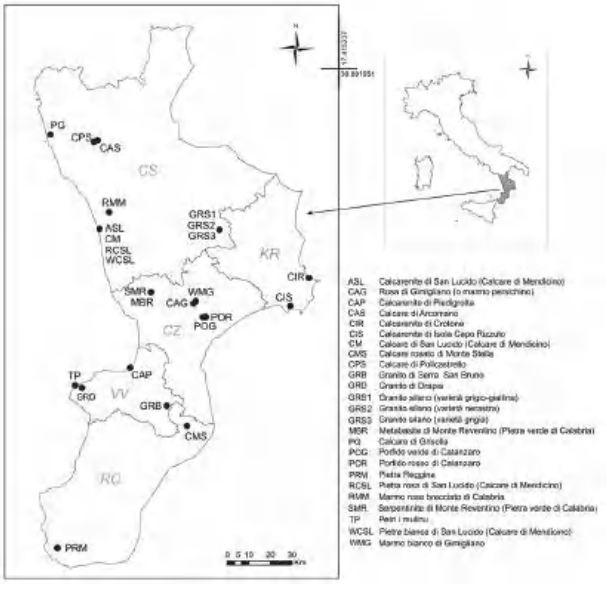


Figure 1. Location of the studied quarries in Calabria Region (Southern Italy). The legend shows the historical name of the stone materials studied.

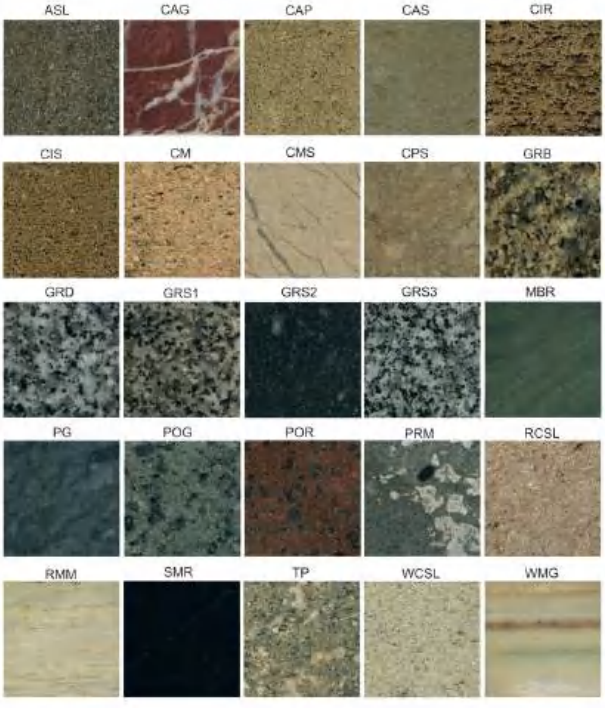


Figure 2. Macroscopic photos of the studied stone materials representative of each quarry. The photos were collected in reflected light using a flatbed scanner. The sizes of each photo are 5 cm × 5 cm.

Table 1. List of the stone materials studied from the five Calabrian provinces (Southern Italy).

Short Code of the Quarry	Historic Name of the Stone	Name of the City Where the Quarry Is Located	Geological Classification of the Stone
ASL	Calcarenite di San Lucido (Calcare di Mendicino)	San Lucido (Cosenza)	Calcarenite
CAG	Rosa di Gimigliano (o marmo persichino)	Gimigliano (Catanzaro)	Dolomitic Limestone
CAP	Calcarenite di Piedigrotta	Pizzo Calabro (Vibo Valenzia)	Calcarenite
CAS	Calcare di Arcomano	San Donato di Ninea (Cosenza)	Limestone
CIR	Calcarenite di Crotone	Crotone (Crotone)	Biocalcarenite
CIS	Calcarenite di Isola Capo Rizzuto	Isola Capo Rizzuto (Crotone)	Calcarenite
CM	Calcare di San Lucido (Calcare di Mendicino)	San Lucido (Cosenza)	Variable from limestone to dolomitic limestone
CMS	Calcare rosato di Monte Stella	Pazzano (Reggio Calabria)	Oolitic limestone (oosparite)
CPS	Calcare di Policastrello	San Donato di Ninea (Cosenza)	Evaporitic limestone
GRB	Granito di Serra San Bruno	Serra San Bruno (Vibo Valentia)	Granodiorite
GRD	Granito di Drapia	Drapia (Vibo Valentia)	Granodiorite
GRS1	Granito silano (varietà grigio-giallina)	San Giovanni in Fiore (Cosenza)	Granodiorite
GRS2	Granito silano (varietà nerastra)	San Giovanni in Fiore (Cosenza)	Diorite
GRS3	Granito silano (varietà grigia)	San Giovanni in Fiore (Cosenza)	Granodiorite
MBR	Metabasite di Monte Reventino (Pietra verde di Calabria)	Platania (Catanzaro)	Metabasite o greenschist
PG	Calcare di Grisolia	Grisolia (Cosenza)	Limestone
POG	Porfido verde di Catanzaro	Catanzaro (Catanzaro)	Dioritic green porphyry
POR	Porfido rosso di Catanzaro	Catanzaro (Catanzaro)	Monzonitic red porphyry
PRM	Pietra Reggina	Motta San Giovanni (Reggio Calabria)	Calcarenite
RCSL	Pietra rosa di San Lucido (Calcare di Mendicino)	San Lucido (Cosenza)	Variable from limestone or dolomitic limestone to calcarenite
RMM	Marmo rosa brecciato di Calabria	Montalto Uffugo (Cosenza)	Fine marble
SMR	Serpentinite di Monte Reventino (Pietra verde di Calabria)	Platania (Catanzaro)	Serpentinite
TP	Petri i mulinu	Tropea (Vibo Valentia)	Calcarenite
WCSL	Pietra bianca di San Lucido (Calcare di Mendicino)	San Lucido (Cosenza)	Biocalcarenite
WMG	Marmo bianco di Gimigliano	Gimigliano (Catanzaro)	Calce-schist

3.2. Image Acquisition System

To acquire the images, the stone samples coming from each Calabrian quarry have been cut through a petrographic cutter machine in order to obtain perfectly flat and smooth surfaces. The flat surface obtained, for all 250 samples, was acquired in three different modes, using two simple tools: a smartphone and a flatbed scanner.

1. The first typology of images was acquired using a smartphone Samsung Galaxy Note 4, with 16 Mpixel camera and a resolution of 4608 × 3456 pixels. The acquisition was performed under standard conditions, illuminating the sample with an LED illuminator, inserting the flash of the smartphone and always keeping constant the distance between the smartphone and the sample surface (10 cm).
2. The second typology of images was acquired by flatbed scanner, with reflected light, using an Epson Perfection 2400 Photo scanner, with a resolution of 600 dpi (image type: 24-bit colors). During the acquisition, all the filters were removed and the samples were carefully covered with a synthetic black and thermal cloth to normalize the acquisition and to perform it in standard condition.
3. The third typology of images was acquired using the same flatbed scanner and the same conditions of the previous typology, the only difference is that the acquisition was made on the wet surface of the samples, in order to simulate a polished effect of the stone.

4. Pre-Trained CNN Models and Classification Algorithms

Object recognition is a key technology based on AI techniques. Recently, ML and DL techniques have become commonly used approaches to solve problems related to object recognition. In DL, a computer model learns to perform classification tasks directly from images. Recent developments have allowed DL to progress to such an extent that it surpasses humans in some activities, such as the classification of objects in images. There are two approaches to performing object recognition using DL:

1. Train a model from scratch;
2. Using a pre-trained DL model (*Transfer Learning* (TL) technique [1]).

One of the biggest advantages of the current DL approach is the ability to have access to pre-trained networks. In this way, it is possible to avoid having to spend many hours, if not days, training the network, and directly use the architecture of the network and the weights obtained from the training, downloading them from the Internet. It is one of the advantages offered by the "Open Source" approach adopted to a large extent. Another advantage is the capability of solving the problem of lacking a large training dataset. The features from multiple fully-connected layers are combined with different weights and used to train different algorithms for image classification.

Our image classification system is based on the TL approach [1], a process that consists of refining a previously trained model through a re-training of the specific images used for the recognition. In our study, we utilized four CNN models and the performance of these models was evaluated. All the pre-trained models were trained on the ImageNet dataset, and each model is briefly explained in the following sections.

In particular, we used the pre-trained model as a feature extractor. We know that a DL model is basically a grouping of interconnected layers of neurons, where the last one acts as a classifier [37]. By removing the final layer of the considered pre-trained CNN network, the output of the penultimate layer, representing the feature vector, can be used as input to the ML classifier in order to perform the stone recognition on the basis of our dataset and, then classify stone as belonging to one of the 25 different classes. In fact, in the pre-trained network, the last layer classifies on the basis of the large database called ImageNet on 1000 different classes of objects. So, the main purpose of the pre-trained CNN is to provide the feature vector extracted by the large image database and use it to perform classification on our stone database.

We have used, in our hybrid solution, four different types of classifiers in order to compare the performance of each one and indicate the most promising solution in solving an image classification task. The network code is open source and is provided for the Tensorflow framework [38].

4.1. ImageNet Dataset

ImageNet is a large image database, created for use in the field of computer vision and in the field of object recognition [39]. The dataset consists of more than 14 million images with the indication of the objects they represent. Identified objects have been classified into more than 20,000 categories: some categories of frequent objects, such as "balloon" or "strawberry", consist of several hundred images [40]. Since 2010, a competition called the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) is held every year. On this occasion, software programs are made to compete to classify and correctly detect objects and scenes contained in the images. As part of the competition, a reduced list of images with objects belonging to a thousand non-overlapping categories is used [41].

4.2. CNN Models

The *Convolutional Neural Network* (CNN) or ConvNet [42] is one of the most common algorithms for DL, a type of ML in which a computer model learns to perform classification tasks directly from images, video, text or sound. CNNs are particularly useful for finding patterns in images to recognize objects, faces and scenes. They learn directly from image data, using patterns to classify images and eliminating the need for manual feature

extraction. CNNs offer an alternative approach that automates feature learning using large databases of samples, called training sets, which represent an application domain of interest. A CNN can have tens or hundreds of layers capable of learning to detect the different features of an image, see Figure 3.

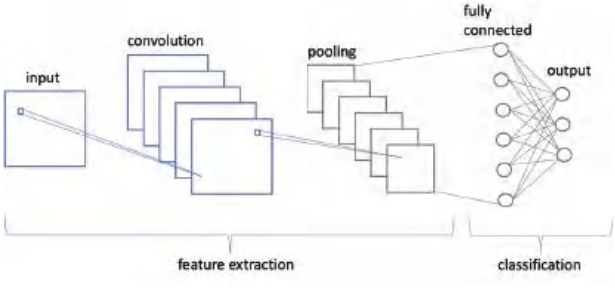


Figure 3. Example of CNN architecture.

On the basis of the literature works analysis and of the main CNN networks proposed and used by researchers to perform their experiments, the choice of the CNNs that we have proposed in our Two-Stage Hybrid Model fell back on the following neural networks: VGG-16, VGG-19, Inception-V3 and ResNet50. In this section, these four CNNs, used for our Two-Stage Hybrid Model, are briefly introduced.

VGG-16 is a neural network architecture designed by the Visual Geometry Group, the department of engineering sciences of the University of Oxford, with 13 convolutional layers and three fully connected layers for classification and detection tasks, as shown in Figure 4. It accepts, as input, images with a resolution of 224 × 224 pixels in RGB, has an output of 4096 features, and input of a classification layer.

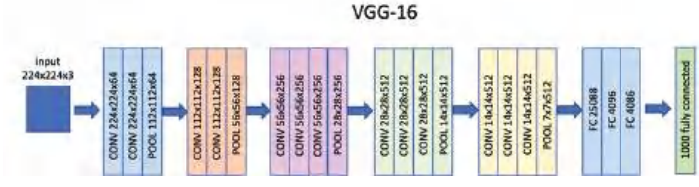


Figure 4. VGG-16 architectural model.

The VGG-19 is a CNN with 19 layers among convolutional, pooling and fully connected layers trained on the ImageNet database. It has an architecture very similar to the previous VGG-16 version as it is possible to view in Figure 5 where the output of the network is a 4096 feature vector which is then used as input of a classification layer.

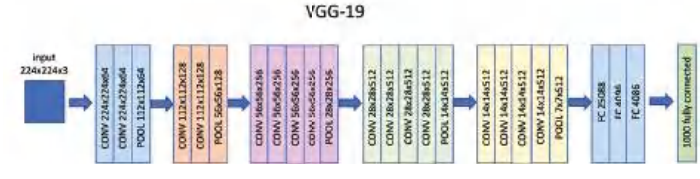


Figure 5. VGG-19 architectural model.

Inception-V3 was developed by Google and trained on the ImageNet database composed of 1000 different classes [43,44]. This model uses the inception modules which take several convolutional kernels of different sizes and stack their outputs along the depth dimension in order to capture features at different scales, see Figure 6.

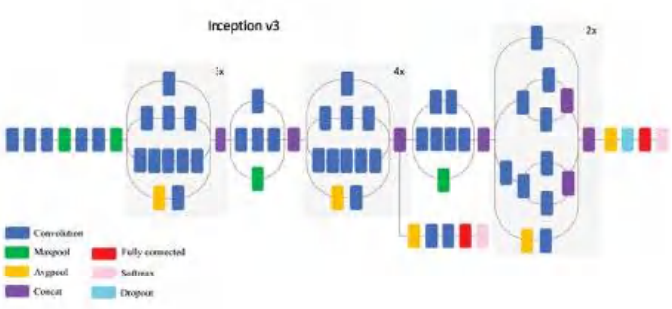


Figure 6. Inception-V3 architectural model.

ResNet won ILSVRC in 2015, the ImageNet Large Scale Visual Recognition Challenge [45]. Five different versions of ResNet exist, with a number of layers from 18 to 152 and with a consequent explosion of complexity. ResNet50 is the version with 50 layers, see Figure 7.

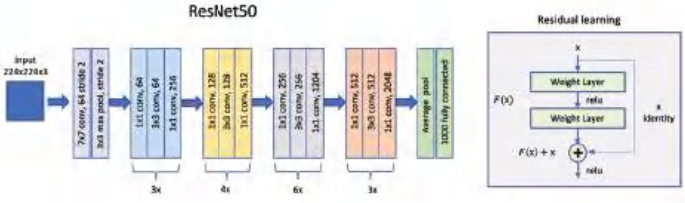


Figure 7. ResNet50 architectural model.

4.3. Classification Techniques

Once the features are extracted by the CNN network in the first stage of our Two-Stage Hybrid Model, these data are passed in input to the second stage where a set of ML classifiers are used for performing stone classification on the 25 different stone classes. The choice of these classifiers is, also in this case, due to the previous analysis of literature manuscripts where a lot of authors proposed mechanisms based on different ML algorithms. We have selected the most used and proposed different machine learning methods in order to perform our tests on the considered stones belonging to Calabrian quarries. A plethora of classifiers could be used but these five are used for the most part of the considered works. So, in this paper, the following ML classification methods are used:

- *Softmax* or *Multinomial Logistic Regression* (MLR) [2], representative of regression models;
- *Support Vector Machine* (SVM) [3,46], an example of linear models;
- *k-Nearest Neighbors* (kNN) [4], representative of density or instance based models;
- *Random Forest* (RF) [5,47], representative of ensemble models;
- *Gaussian Naive Bayes* (GNB) [6], an example of probabilistic models.

For more details on these classifier algorithms please refers to [48].

5. Two-Stage Hybrid Model

With the term *hybrid model* we mean an approach that makes use of a *Deep Learning* (DL) network together with a classical *Machine Learning* (ML) algorithm. The joined use of these two different approaches can give many advantages to the classification purpose.

The proposed hybrid model consists of two stages as it is possible to view in Figure 8. The first stage of the model guarantees an automatic feature extraction phase that is used as input for the second stage, which has the task of performing the classification phase.

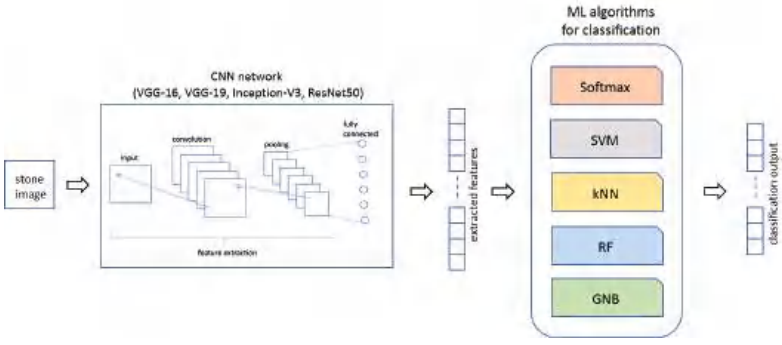


Figure 8. Two-Stage Hybrid Model used for stone classification.

Moreover, in order to make more efficient the phase of the feature extraction we have used in our Two-Stage Hybrid Model the well-known *Transfer Learning* (TL) approach that, throughout models of CNN networks pre-trained on a large image database, guarantees a more rapid feature extraction phase avoiding building the neural network from scratch. This approach, as proved by the literature, guarantees optimal results and allows for efficient creation, on the basis of a set of filters used in the CNN, a feature vector composed of numerical values representing the main information of the image in a very short time period.

The *Two-Stage Hybrid Model* that we propose in this work is shown in Figure 8. We use four different pre-trained CNNs from the main networks provided in the literature and briefly explained in Section 4.2 in the first stage of the model and five different ML algorithms in the second stage, see Section 4.3, in order to accomplish the image classification on our specific dataset. So, each CNN model has been used as a feature extractor for the five different classifiers. The main output parameters are reported in the next section in order to evaluate how each model performs providing a useful comparison of an original context represented by stones of Calabrian quarries.

6. Experiments: Results and Discussion

In this section, we give a detailed description of all the experiments we performed with our proposed two-stage architecture presented in the previous Section (Section 5). The heart of this project is to perform the right class prediction for the considered stones' images. We have several images for the input, subdivided between *training dataset* (80% of the total dataset) and *test dataset* (20% of the total dataset). It might be interesting to see what the Two-Stage Hybrid Model guesses for classes of images it never saw during training.

In order to determine the performance of a neural network, it is important to take into account some characteristic parameters that can help to indicate the goodness of the approach. In the context of AI, the confusion matrix, also called the misclassification table, returns a representation of the accuracy of statistical classification. Each column of the matrix represents the predicted values, while each row represents the real values. The element on row i-th and column j-th is the number of cases in which the classifier has classified the "true" class i-th as class j-th. Through this matrix, it is observable if there is "confusion" in the classification of different classes. The confusion matrix provides a lot of information. However, more concise metrics are often useful, such as: accuracy, precision, recall and F1-score.

6.1. Experimental Environment

In order to perform classification experiments, a workstation equipped with an Intel i9 10900K CPU with 32 GB RAM DDR4, an Nvidia 3060Ti graphic card and a 512 GB SSD and a Python 3.8.10 was used. Moreover, we have installed some additional libraries such as

Matplotlib, Pandas, Tensorflow, and NumPi in order to analyze and perform classification on our dataset.

6.2. Our Dataset

In this study, the image classes, that represent the different object categories, are the 25 different stone types of Calabrian quarries. We need to ensure that the images have the right size for each considered CNN. In particular, a resolution of $224 \times 224 \times 3$ is used for VGG-16, VGG-19 and ResNet50 models and a resolution of $299 \times 299 \times 3$ for the input of Inception-V3 one. So, a little pre-processing operation was made on the input image dataset in order to match the specific image resolution requirements of the considered CNN architecture.

6.3. Augmentation of Our Dataset

In order to make our experiments with a sufficient number of images, we have increased our dataset by creating new images through a simple elaboration of our data samples by performing the so-called *Data Augmentation* technique [49].

In order to augment the training dataset for our experiment and consider the large image resolution of our dataset, we have manipulated our input images in a very simple way. We have cropped the original image in five different parts both in horizontal and vertical axes creating so 25 different images from the original one as it is possible to view in Figure 9. This has allowed us to increase the number of images by a 25 factor, creating a consistent dataset of stone images.

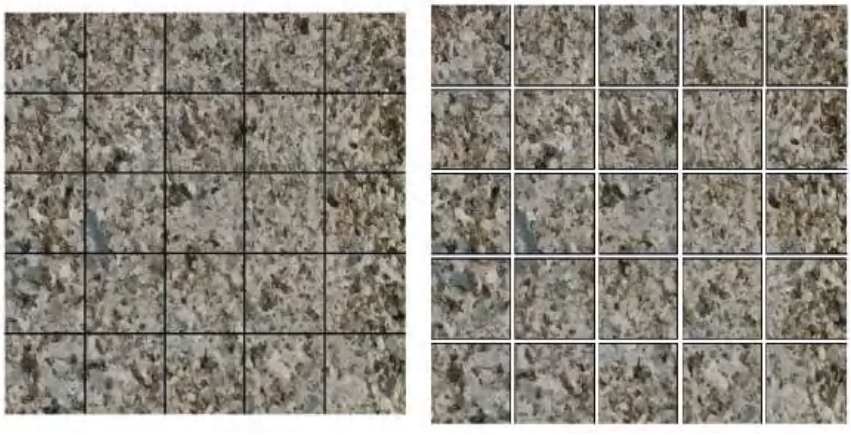


Figure 9. Example of data augmentation used in our experiments.

6.4. Classification Results

As described in Section 3, for our experiments, images from three different typologies of acquisition have been used. For each type of stone, 10 different images with each type of acquisition technique have been created. Then, for each type of stone, such as ASL, CAG, CAP, etc., we have used $10 \times 3 = 30$ different images on which we have performed the technique of data augmentation, as previously described in Section 6.3, in order to increase the input dataset both for training and test campaigns. As we said in the previous section, we extracted 25 different images from each one, then, we obtained $25 \times 30 = 750$ different images for each stone typology. From these images, 600 were used for conducting the training of the neural network and 150 were used for performing the test campaigns in order to evaluate the goodness of the network. This corresponds to having, respectively, a training set and a test set composed of 80% and 20% of the overall dataset.

Different experiments were conducted using the image database created by the stones acquisition process. We have performed experiments on every single typology of stone

acquisition, that is, using first only the database comes from smartphone acquisition, second, the database comes from the acquisition with the flatbed scanner with reflected light and, finally, the database created using the flatbed scanner with reflected light on the wet surface of the samples. These experiments have shown the advantages of the automatic classification through the CNN approach using the different *Classification Algorithm*s (CLFs)

The total number of CNN parameters is reported in Figure 10 (left) in order to show how complex is the CNN network and provide a comparison between the four CNN models used in our experimentation. How it is possible to view in Figure 10 (left), the Inception-V3 and ResNet50 CNN models have a lower number of parameters. Moreover, Figure 10 (right) reports the inference time comparison between the four CNN models, both to infer on training and on test samples. How it is possible to note that the ResNet50 model, due to its small parameter number, is able to infer in a shorter amount of time in both sample sets.

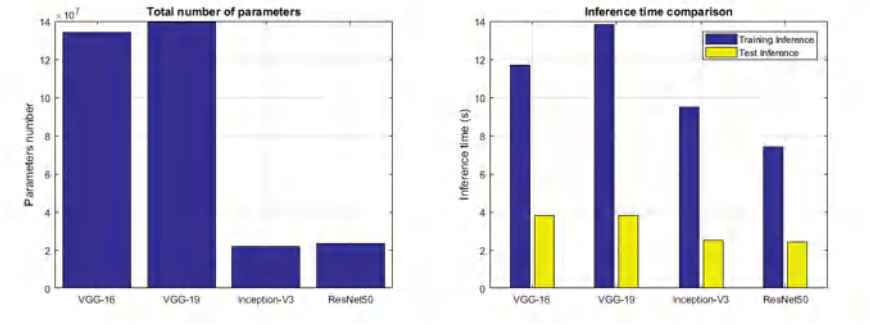


Figure 10. Total number of CNN parameters (**left**) and Inference time (seconds) (**right**) of each CNN model used in the hybrid architecture.

In Figure 11 we show the confusion matrix of the first two used classifiers, Softmax and SVM, with the ResNet50 CNN model that results in the best CNN to be used together with the classifiers in order to have the best results in terms of accuracy. In this section, only the confusion matrices of the ResNet50+CLF hybrid model are reported. It is possible to observe that both approaches reached similar results on the test set used for experiments. The network, both with Softmax and SVM classifier, is able to perform prediction with high accuracy. It is possible to observe that the classes that present some recognition problems in all Two-Stage Hybrid Models are those related to the family of granite stones, that is GRB, GRD, GRS1, GRS2 and GRS3 families. This is due to a very similar texture of these types of stones that also makes it difficult for an expert eye to capture the differences.

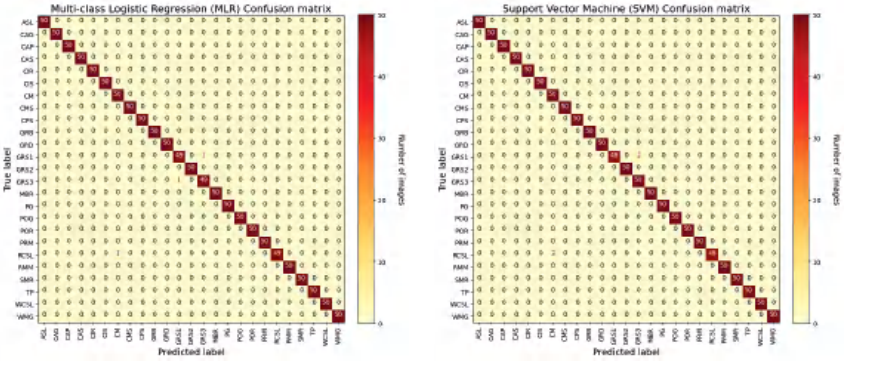


Figure 11. Confusion matrix for (**left**) Softmax (MLR) and (**right**) SVM classifiers with ResNet50 CNN model.

In a similar way, we have performed experiments using *k-Nearest Neighbors* (kNN) and *Random Forest* (RF) classifiers. In the following, the confusion matrices are shown for both classification algorithms. As it is possible to observe in Figure 12, also in this case, the recognition that presents more issues regards the granite stone classification. It is possible to view that these two types of classifiers have similar accuracy to the first two considered in the previous experiments. This means that Softmax and SVM have similar performance to kNN and RF algorithms.

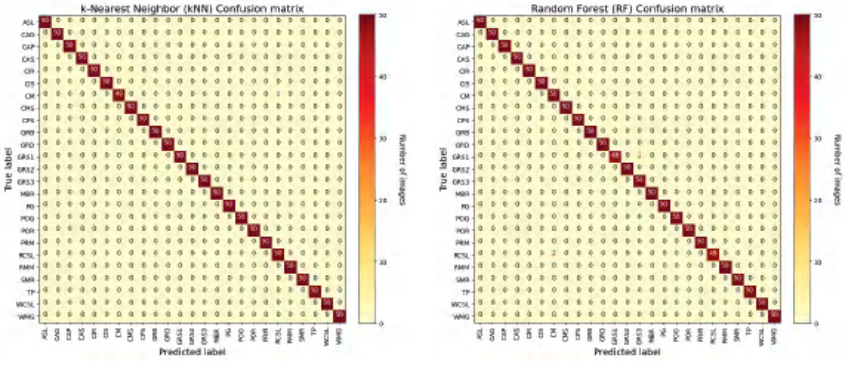


Figure 12. Confusion matrix for (**left**) kNN and (**right**) RF classifiers with ResNet50 CNN model.

The last used classification algorithm is the GNB and, in Figure 13 the confusion matrix extracted by image recognition tests is reported. As shown in the figure, this last method presents the worst results in terms of classification accuracy.

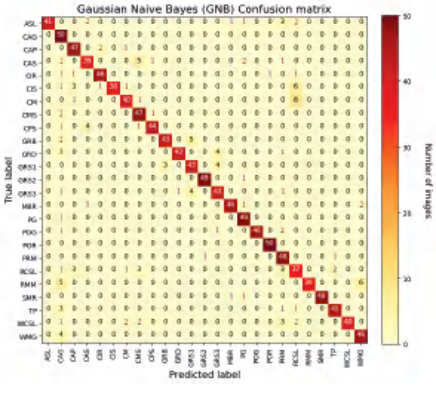


Figure 13. Confusion matrix for GNB classifier with ResNet50 CNN model.

The accuracy that we have calculated for all the conducted experiments is shown in Table 2 and Figure 14 (left). Then, the experiments showed that these algorithms have optimum performance in pattern recognition purposes: all algorithms are able to recognize the most part of the input images but, between all, the Softmax, SVM, RF and kNN approaches allow to reach very high accuracy values representing the best candidates for image classification in this applicative domain.

Table 2. Accuracy of CNNs plus CLFs.

	MLR	SVM	kNN	RF	GNB
VGG-16 (%)	99.0	99.3	97.8	98.3	91.4
VGG-19 (%)	99.0	99.1	98.4	98.2	93.0
Inception-V3 (%)	96.0	91.4	93.8	91.4	78.9
ResNet50 (%)	99.7	99.8	99.9	99.7	88.5

Other than the accuracy, in this section, we also show other metrics: precision, recall and F1-score in order to prove the goodness of the classification. With precision metric, the system shows how on the true classification the most part is correct; with the recall instead, the system is able to cover the most part of the true positive; F1-score gives a joined metrics between precision and recall. Figures 14 and 15 provide the metric values for each Two-Stage Hybrid Model considered in our tests. As it is possible to observe, the two-stage hybrid approach provides a very good performance in almost all tests reaching optimal results using a ResNet50 CNN in the first stage and a kNN ML algorithm in the second stage of the hybrid model.

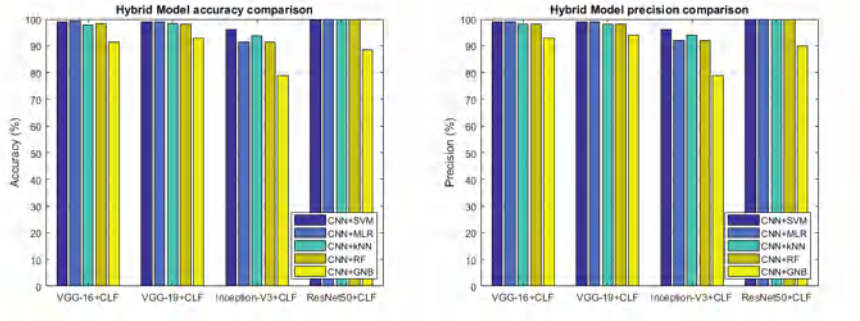


Figure 14. Accuracy (**left**) and precision (**right**) comparison.

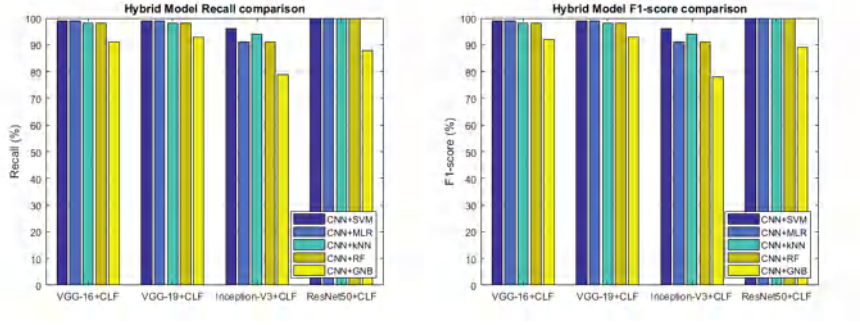


Figure 15. Recall (**left**) and F1-score (**right**) comparison.

So, after this set of experiments that have used the Two-Stage Hybrid Model combining each DL technique with each ML algorithm, it is possible to make some considerations in order to conclude our work. The proposed hybrid model consists of a two-stage approach that makes use of DL techniques in the first stage with the task of performing feature extraction and ML algorithms in the second stage with the task of performing classification and, then operating the stone recognition so as to attribute the right class to the specific stone. The use of the combination of DL and ML approaches resulted in a very high-performing system able to recognize the belonging class very well. The most accurate

combination that emerged from the results was based on the ResNet50 CNN model in joining with the kNN classifier (ResNet50 + kNN). This combination is able to guarantee a high accuracy in the stone recognition as proved by Table 2 and Figures 14 and 15. The ResNet50 network also resulted in the best in terms of the number of CNN parameters and inference time as reported in Figure 10.

7. Conclusions

In this paper, an automatic stones classification approach based on a two-stage hybrid architecture able to classify different stone classes in the Calabria area (Southern Italy) is presented. The obtained results are pretty impressive. The neural network models are able to reach amazing results in the prediction process on the input images provided to the network. The proposed Two-Stage Hybrid Model based on DL and ML techniques results in being a more promising approach for stone recognition issues. From the conducted analysis, it emerged that the only classes that present some minor issues are those related to granite typologies that result quite complex also for a careful eye of an expert in this field. However, this two-stage hybrid approach, which uses *Deep Learning* (DL) CNN models together with different *Machine Learning* (ML) *Classification Algorithm*s (CLFs), permits to create a system that is very powerful and able to reach optimal performance in terms of image recognition. It exploits the power of DL for the phase of feature extraction, which represents the more complex phase, and leverages the classical ML algorithms to perform the classification phase. Moreover, in order to avoid creating the CNNs from scratch, the proposed Two-Stage Hybrid Model is based on the *Transfer Learning* (TL) paradigm that is able to exploit pre-trained networks on large datasets such as ImageNet for reducing the phase of feature extraction. In fact, the CNN in the TL mode is able to infer on both training and test sets in a very quick manner as shown by provided results. The most promising combination that emerged from tests is based on the ResNet50 CNN model together with a kNN classifier. It guarantees high accuracy and allows us to obtain the best results in terms of CNN parameter number and inference time. Furthermore, this type of approach shows that there is a concrete possibility to build tools that are easy to use even for people who do not have geological knowledge; the applications could be numerous and range from the field of archaeometry and diagnostics, up to applications of automatic recognition in the field of the materials sciences.

Author Contributions: Conceptualization, G.F. and M.T.; methodology, G.F. and D.M.; software, M.T.; validation, G.F., F.D.R.; investigation and resource, R.D.L.; data curation, G.F. and M.T.; writing—original draft preparation, G.F. and M.T.; supervision, F.D.R.; funding acquisition, D.M and G.F. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the Calabria Region by "POR Calabria FESR-FSE 2014-2020, finanziamento di Progetti di Ricerca e Sviluppo, obiettivo specifico 1.2—Rafforzamento del Sistema Innovativo Regionale e Nazionale, Azione 1.2.2—Supporto alla realizzazione di Progetti Complessi di Attività di Ricerca e Sviluppo su poche Aree Tematiche di Rilievo e all'Applicazione di Soluzioni Tecnologiche Funzionali alla Realizzazione delle Strategie di S3".

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Acknowledgments: We would like to thank Eng. Francesco Buffone for his precious help in the studying and understanding CNN networks and ML algorithms and developing python script for image classification testing.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	*Artificial Intelligence*
CNN	*Convolutional Neural Network*
DL	*Deep Learning*
DNN	*Deep Neural Network*
FC	*Fully Connected*
GNB	*Gaussian Naive Bayes*
kNN	*k-Nearest Neighbors*
LR	*Logistic Regression*
ML	*Machine Learning*
MLP	*Multi-Layer Perceptron*
MLR	*Multinomial Logistic Regression*
PR	*Pattern Recognition*
RF	*Random Forest*
SVM	*Support Vector Machine*
TL	*Transfer Learning*

References

1. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [CrossRef]
2. Mohammed, A.A.; Umaashankar, V. Effectiveness of Hierarchical Softmax in Large Scale Classification Tasks. In Proceedings of the 2018 IEEE International Conference on Advances in Computing, Communications and Informatics (ICACCI), Bangalore, India, 19–22 September 2018; pp. 1090–1094.
3. Hsu, C.W.; Lin, C.J. A comparison of methods for multiclass support vector machines. *IEEE Trans. Neural Netw.* **2002**, *13*, 415–425. [PubMed]
4. Ma, H.; Gou, J.; Wang, X.; Ke, J.; Zeng, S. Sparse coefficient-based k-nearest neighbor classification. *IEEE Access* **2017**, *5*, 16618–16634. [CrossRef]
5. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
6. Jahromi, A.H.; Taheri, M. A non-parametric mixture of Gaussian naive Bayes classifiers based on local independent features. In Proceedings of the 2017 IEEE Artificial Intelligence and Signal Processing Conference (AISP), Shiraz, Iran, 25–27 October 2017; pp. 209–212.
7. De Luca, R.; Barca, D.; Bloise, A.; Cappa, M.; De Angelis, D.; Fedele, G.; Mariottini, S.; Micieli, D.; Muto, F.; Piluso, E.; et al. RecoStones: A New Tool to Identify Calabrian Stone Materials Through Image Processing. *Geoheritage* **2021**, *13*, 1–15. [CrossRef]
8. Penta, F. Marmi graniti e porfidi della Calabria. In *Marmi Pietre e Graniti Nell'Arte Nell'Industria nel Commercio, Rassegna Bimestrale Ufficiale Della Federazione Nazionale Fascista dell'Industria del Marmo Graniti e Pietre*; Università della Sapienza: Roma, Italy, 1932; Volume 2, pp. 30–39.
9. Dumon, P. Les materiaux naturels de decoration en Italie depuis un siecle. In *Edité par Givors: Le Mausolee*; CNRS: Givors, France, 1975.
10. Cuteri, F.; Iannelli, M.; Mariottini, S. Cave costiere in Calabria tra Ionio e Tirreno. Montagne incise. Pietre incise. In *Atti del convegno. Cave: Censimenti, Indagini di superficie, Valorizzazione*; All'Insegna del Giglio" Sesto Fiorentino: Firenze, Italy, 2011; p. 25.
11. Chen, L.; Li, S.; Bai, Q.; Yang, J.; Jiang, S.; Miao, Y. Review of image classification algorithms based on convolutional neural networks. *Remote Sens.* **2021**, *13*, 4712. [CrossRef]
12. Lorente, Ò.; Riera, I.; Rana, A. Image classification with classic and deep learning techniques. *arXiv* **2021**, arXiv:2105.04895.
13. Tiwari, V.; Pandey, C.; Dwivedi, A.; Yadav, V. Image classification using deep neural network. In Proceedings of the 2020 2nd IEEE International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Greater Noida, India, 18–19 December 2020; pp. 730–733.
14. Gao, F.; Hsieh, J.G.; Jeng, J.H. A Study on Combined CNN-SVM Model for Visual Object Recognition. *J. Inf. Hiding Multim. Signal Process.* **2019**, *10*, 479–487.
15. Tang, Y. Deep learning using linear support vector machines. *arXiv* **2013**, arXiv:1306.0239.
16. He, Y.; Qin, Q.; Vychodil, J. A Pedestrian Detection Method Using SVM and CNN Multistage Classification. *J. Inf. Hiding Multim. Signal Process.* **2018**, *9*, 51–60.
17. Vo, A.T.; Tran, H.S.; Le, T.H. Advertisement image classification using convolutional neural network. In Proceedings of the 2017 9th IEEE International Conference on Knowledge and Systems Engineering (KSE), Hue, Vietnam, 19–21 October 2017; pp. 197–202.
18. Cheng, G.; Ma, C.; Zhou, P.; Yao, X.; Han, J. Scene classification of high resolution remote sensing images using convolutional neural networks. In Proceedings of the 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Beijing, China, 10–15 July 2016; pp. 767–770.

19. Mujawar, S.; Kiran, D.; Ramasangu, H. An Efficient CNN Architecture for Image Classification on FPGA Accelerator. In Proceedings of the 2018 Second IEEE International Conference on Advances in Electronics, Computers and Communications (ICAECC), Bangalore, India, 9–10 February 2018; pp. 1–4.
20. Han, S.H.; Lee, K.Y. Implemetation of image classification cnn using multi thread gpu. In Proceedings of the 2017 IEEE International SoC Design Conference (ISOCC), Seoul, Korea, 5–8 November 2017; pp. 296–297.
21. Zhang, Y.; Sun, X.; Sun, H.; Zhang, Z.; Diao, W.; Fu, K. High Resolution SAR Image Classification with Deeper Convolutional Neural Network. In Proceedings of the IGARSS 2018—2018 IEEE International Geoscience and Remote Sensing Symposium, Valencia, Spain, 22–27 July 2018; pp. 2374–2377.
22. Abdullah; Hasan, M.S. An application of pre-trained CNN for image classification. In Proceedings of the 2017 20th IEEE International Conference of Computer and Information Technology (ICCIT), Dhaka, Bangladesh, 22–24 December 2017; pp. 1–6.
23. Suganthi, M.; Sathiaseelan, J. An exploratory of hybrid techniques on deep learning for image classification. In Proceedings of the 2020 4th IEEE International Conference on Computer, Communication and Signal Processing (ICCCSP), Chennai, India, 22–23 April 2020; 2020; pp. 1–4.
24. Fauzi, F.; Permanasari, A.E.; Setiawan, N.A. Butterfly Image Classification Using Convolutional Neural Network (CNN). In Proceedings of the 2021 3rd IEEE International Conference on Electronics Representation and Algorithm (ICERA), Yogyakarta, Indonesia, 29–30 July 2021; pp. 66–70.
25. Ghosh, S.; Singh, A.; Kavita; Jhanjhi, N.; Masud, M.; Aljahdali, S. SVM and KNN Based CNN Architectures for Plant Classification. *CMC-Comput. Mater. Contin.* **2022**, *71*, 4257–4274. [CrossRef]
26. Bianconi, F.; González, E.; Fernández, A.; Saetta, S.A. Automatic classification of granite tiles through colour and texture features. *Expert Syst. Appl.* **2012**, *39*, 11212–11218. [CrossRef]
27. Bianconi, F.; Bello, R.; Fernández, A.; González, E. On comparing colour spaces from a performance perspective: Application to automated classification of polished natural stones. In Proceedings of the International Conference on Image Analysis and Processing, Genoa, Italy, 7–11 September 2015; pp. 71–78.
28. Araújo, M.; Martínez, J.; Ordóñez, C.; Vilán, J.A. Identification of granite varieties from colour spectrum data. *Sensors* **2010**, *10*, 8572–8584. [CrossRef] [PubMed]
29. Ershad, S.F. Color texture classification approach based on combination of primitive pattern units and statistical features. *arXiv* **2011**, arXiv:1109.1133.
30. Riaz, F.; Hassan, A.; Rehman, S.; Qamar, U. Texture classification using rotation-and scale-invariant gabor texture features. *IEEE Signal Process. Lett.* **2013**, *20*, 607–610. [CrossRef]
31. Zand, M.; Doraisamy, S.; Halin, A.A.; Mustaffa, M.R. Texture classification and discrimination for region-based image retrieval. *J. Vis. Commun. Image Represent.* **2015**, *26*, 305–316. [CrossRef]
32. Chow, B.H.Y.; Reyes-Aldasoro, C.C. Automatic Gemstone Classification Using Computer Vision. *Minerals* **2022**, *12*, 60. [CrossRef]
33. Ather, M.; Khan, B.; Wang, Z.; Song, G. Automatic recognition and classification of granite tiles using convolutional neural networks (CNN). In Proceedings of the 2019 3rd International Conference on Advances in Artificial Intelligence, Istanbul, Turkey, 26–28 October 2019; pp. 193–197.
34. Tereso, M.; Rato, L.; Gonçalves, T. Automatic classification of ornamental stones using Machine Learning techniques A study applied to limestone. In Proceedings of the 2020 15th IEEE Iberian Conference on Information Systems and Technologies (CISTI), Sevilla, Spain, 24–27 June 2020; pp. 1–6.
35. Zhang, Y.; Li, M.; Han, S.; Ren, Q.; Shi, J. Intelligent identification for rock-mineral microscopic images using ensemble machine learning algorithms. *Sensors* **2019**, *19*, 3914. [CrossRef]
36. Iannelli, M.; Mariottini, S.; Vivacqua, P. Indagini geoarcheologiche nel tratto della costa tirrenica calabrese compreso tra Nicotera e Pizzo Calabro. In *I° Convegno Regionale di Geoarcheologia, Geologia e Geoarcheologia: La Calabria, la Protezione dei Beni Culturali, il Turismo*; Sala congressi di Palazzo Sersale: Cerisano, Italy, 2015.
37. Gonzalez, R.C. Deep Convolutional Neural Networks. *IEEE Signal Process. Mag.* **2018**, *35*, 79–87. [CrossRef]
38. Tensorflow Website. 2018. Available online: https://www.tensorflow.org/ (accessed on 10 April 2022).
39. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
40. ImageNet Website. 2022. Available online: http://image-net.org/about-overview/ (accessed on 10 April 2022).
41. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]
42. Nielsen, M.A. *Neural Networks and Deep Learning*; Determination Press: San Francisco, CA, USA, 2015; Volume 25.
43. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
44. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
45. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
46. Cortes, C.; Vapnik, V. Support-vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [CrossRef]

47. Breiman, L. *Classification and Regression Trees*; Routledge: London, UK, 2017.
48. Tropea, M.; Fedele, G. Classifiers comparison for convolutional neural networks (CNNs) in image classification. In Proceedings of the 2019 IEEE/ACM 23rd International Symposium on Distributed Simulation and Real Time Applications (DS-RT), Cosenza, Italy, 7–9 October 2019; pp. 1–4.
49. Cui, Y.; Song, Y.; Sun, C.; Howard, A.; Belongie, S. Large scale fine-grained categorization and domain-specific transfer learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4109–4118.

sensors

MDPI

Article

A Multi-Sensor Data-Fusion Method Based on Cloud Model and Improved Evidence Theory

Xinjian Xiang *, Kehan Li, Bingqiang Huang and Ying Cao

School of Automation and Electrical Engineering, Zhejiang University of Science and Technology, Hangzhou 310023, China
* Correspondence: 188002@zust.edu.cn; Tel.: +86-198-8311-0097

Abstract: The essential factors of information-aware systems are heterogeneous multi-sensory devices. Because of the ambiguity and contradicting nature of multi-sensor data, a data-fusion method based on the cloud model and improved evidence theory is proposed. To complete the conversion from quantitative to qualitative data, the cloud model is employed to construct the basic probability assignment (BPA) function of the evidence corresponding to each data source. To address the issue that traditional evidence theory produces results that do not correspond to the facts when fusing conflicting evidence, the three measures of the Jousselme distance, cosine similarity, and the Jaccard coefficient are combined to measure the similarity of the evidence. The Hellinger distance of the interval is used to calculate the credibility of the evidence. The similarity and credibility are combined to improve the evidence, and the fusion is performed according to Dempster's rule to finally obtain the results. The numerical example results show that the proposed improved evidence theory method has better convergence and focus, and the confidence in the correct proposition is up to 100%. Applying the proposed multi-sensor data-fusion method to early indoor fire detection, the method improves the accuracy by 0.9–6.4% and reduces the false alarm rate by 0.7–10.2% compared with traditional and other improved evidence theories, proving its validity and feasibility, which provides a certain reference value for multi-sensor information fusion.

Keywords: sensor data fusion; cloud model; Dempster–Shafer evidence theory; cosine similarity; Hellinger distance

Citation: Xiang, X.; Li, K.; Huang, B.; Cao, Y. A Multi-Sensor Data-Fusion Method Based on Cloud Model and Improved Evidence Theory. *Sensors* **2022**, *22*, 5902. https://doi.org/10.3390/s22155902

Academic Editor: Jose Manuel Molina López

Received: 7 July 2022
Accepted: 4 August 2022
Published: 7 August 2022

1. Introduction

Heterogeneous multi-sensors play an important role in information perception, the acquired data may contain some ambiguous and conflicting information due to the limitations of multi-sensor devices' measurement accuracy and the complexity of the working environment, which may result in inaccurate data-fusion decisions [1]. Consequently, the way to better handle multi-sensor data and improve data-fusion accuracy is a popular research direction in the field of data-fusion technology. Common data-fusion algorithms currently include Kalman filtering [2], Bayesian estimation [3], Dempster–Shafer (D-S) evidence theory [4], and artificial neural networks [5], etc. Bayesian networks and D-S evidence theory are commonly used to deal with the uncertainty in multi-sensor data, which frequently results in anomalous data. However, the Bayesian estimation fusion method requires access to prior data to generate new probability estimates, which is not always possible [6]. Dempster–Shafer (D-S) evidence theory is a theory of fuzzy reasoning proposed by Dempster in 1967 [7] and subsequently refined by Shafer [8]. It has been widely employed in areas such as target identification [9], multi-attribute decision analysis [10], fault diagnostics [11], and robotics research [12] due to its capacity to better handle uncertain and unknown situations with unknown prior probabilities. Although the D-S evidence theory has been applied in a number of fields, it has certain problems. One is that there is no unified method for determining the BPA function, and the other is that the

evidence theory is prone to produce results that contradict the facts when dealing with highly conflicting evidence, and there is no unified method for solving this problem. Most scholars have done some research on the above two problems.

Determining the BPA function is an important step in evidence theory, which influences the accuracy of fusion results to some extent. Many researchers have proposed various methods for determining BPA functions [13–15]. The cloud model [16] is a concept proposed by Professor Li in 1995, which is a cognitive model based on probability statistics and fuzzy set theory. It can well portray the fuzziness and randomness of information and is applicable to the field of multi-sensor information fusion. Peng et al. [17] improved the multi-criteria group decision method by using a cloud-model method to deal with uncertain information based on information fusion and information measurement, Liu et al. [18] used the cloud model to describe the load direction in topology optimization with uncertainty, and Peng et al. [19] proposed an uncertain pure linguistic information multicriteria group decision-making method based on the cloud model, demonstrating the advantage of the cloud model in dealing with uncertain information. In this paper, the cloud model is used to determine the BPA function to convert measured quantitative data to qualitative concepts.

The directions for improving the accuracy of traditional evidence theory fusion can be divided into two major areas: improvement of combination rules [20,21] and improvement of the body of evidence. The former blames the D-S rule for producing results that contradict the facts, achieving certain results but destroying the D-S rule's own advantages, such as the law of exchange and the law of union. The latter believes that the problem stems from the unreliability of the information source and uses an improved approach to the body of evidence to deal with the conflict, which retains the good characteristics of Dempster's rule and weakens the influence of conflicting evidence on the fusion result. As Haenni [22] points out, the improvement of the body of evidence is more reasonable both from an engineering and mathematical standpoint. The calculation and assignment of weights to the body of evidence is critical to improving the body of evidence, and some scholars have conducted a series of studies on how to evaluate the body of evidence's weights. Murphy [23] proposed a simple averaging method to assign the same weight to each piece of evidence, but it ignores the relationship between the evidence and is therefore unreasonable. Deng et al. [24] proposed a more convergent method based on the rules of evidence theory after weighted average processing of evidence based on trust degree, but it does not take into account the characteristics of the evidence itself. There are two methods for determining the weight of the body of evidence: according to the relationship between the evidence and according to the characteristics of the evidence itself. For the former, Wang et al. [25], Jousselme et al. [26], and Dong et al. [27] measure the relationship between evidence by using the Pignistic probability distance, the Jousselme distance, and cosine similarity, respectively; however, using a single measure of evidence relationship to find the weight of evidence does not accurately describe the relationship between evidence in certain cases. For the latter, scholars have proposed various uncertainty measures based on information entropy, such as Yager's [28] dissonance measure based on the likelihood function and Deng's [29] evidence uncertainty measure based on Shannon entropy, but such methods deal with evidence in a one-sided manner, replacing the entire uncertainty interval with only part of the evidence information. Deng et al. [30] developed a method for evaluating evidence uncertainty based on the Hellinger distance of the uncertainty interval, which is simple to compute and measures uncertainty well for a better integration effect. The relationship between evidence and the characteristics of the evidence itself do not affect each other and are both valid information available within the evidence, yet some current scholarly approaches to improving evidence theory consider only one of them to deal with the evidence, undermining the integrity of the evidentiary information. Some scholars have proposed ways to improve the evidence theory based on both, but they both have some room for improvement. For example, Tao et al. [31] proposed a multi-sensor data-fusion method based on the Pearson correlation coefficient and information entropy.

Xiao et al. [32] proposed a multi-sensor data-fusion method based on belief dispersion of evidence and Deng entropy [29]. Wang et al. [33] combined the Jaccard coefficient and cosine similarity to calculate evidence similarity, combined with evidence-based precision and entropy of evidence to calculate evidence certainty. Although these methods combine the relationship between evidence and the characteristics of evidence itself, they all have certain disadvantages. The Pearson correlation coefficient is only used to portray the linear correlation between normally distributed attributes, which is more demanding on evidence. The Jaccard coefficient and cosine similarity sometimes cannot measure the relationship between evidence correctly. Using information entropy cannot measure the characteristics of evidence itself comprehensively, etc.

In order to more accurately measure the relationship between evidence and the characteristics of evidence itself, and improve the accuracy of data fusion, this paper proposes an improved evidence-theory method based on multiple relationship measures and focal element interval distance. We combine the Jousselme distance, cosine similarity and the Jaccard coefficient to calculate the similarity between the evidence, and we use the Hellinger distance between the evidence determination intervals to measure the certainty of the evidence. Based on these calculations, the evidence weight coefficients are then jointly improved. Finally, the original evidence is average-weighted and fused by using the Dempster rule to produce the result. In addition, we analyze the results of the arithmetic examples to demonstrate the validity of the proposed improved evidence theory. By using the aforementioned improved evidence theory along with cloud model, we developed a multi-sensor data-fusion method. The BPA functions corresponding to each data source are determined based on the cloud model, which converts the collected quantitative data into stereotypical concepts. The fusion results are obtained by fusing each BPA function by using the improved evidence theory mentioned above.

Multi-sensor data-fusion technology can combine relevant information within multiple sensors, thereby increasing the safety and reliability of the overall system. The proposed multi-sensor data-fusion method can be utilized in multi-sensor systems in various fields, such as fault-determination systems, target identification systems, environmental monitoring systems, and intelligent firefighting systems, among others. Due to external factors or their own aging faults, one or more sensors may acquire incorrect information, causing the fusion results to be contradictory to the facts. The proposed method overcomes the problem, improves the handling of ambiguity in sensor data, increases the reliability of data fusion results, and makes it easier for people to make appropriate decisions. We establish an early indoor fire detection model to test the efficacy of the proposed strategy. The proposed method improves accuracy by 0.7–10.2% and reduces false alarm rate by 0.9–6.4% when compared to the traditional evidence theory and other improved evidence theories. It has better fusion performance, which provides some reference value for multi-sensor data fusion.

2. Preliminaries

This section provides a brief overview of D-S evidence theory and the cloud model.

2.1. Cloud Model

Let X be a quantitative domain ($X = \{x\}$) and U be a qualitative concept on the domain X. For any element $x(x \in X)$ and x is a single random realization on U, the certainty of x to U is $y(x) \in [0,1]$, which is a random number with stable tendency, the distribution of x over the domain X is called a cloud model and each (x $y(x)$) becomes a cloud drop [34].

The cloud model completes the conversion of quantitative data to qualitative concepts through numerical characteristic expectation (E_x), entropy (E_n), and hyperentropy (H_e), where expectation is the expected value of the distribution of cloud droplets in the theoretical domain, entropy reflects the dispersion of cloud droplets, and hyperentropy reflects the dispersion of entropy. Because the values of the characteristics corresponding to the

evaluation indices have some stability across the multi-sensor domain and the interval distributions generally follow a normal distribution that is more realistic, the normal cloud model is used in this research. Each parameter's computation formula is presented in Equation (1),

$$\begin{cases} E_{xij} = \frac{C_{ij,max}+C_{ij,min}}{2} \\ E_{nij} = \frac{C_{ij,max}-C_{ij,min}}{2.355} \\ He = \lambda_i \end{cases}, \quad (1)$$

where $[C_{ij,min}, C_{ij,max}]$ are the range of values of the evaluation interval corresponding to the jth certain evaluation index inside the ith data type of the multi-sensor system, and λ_i is a value determined by experts based on experience and is typically 0.01. It is worth noting that the maximum and lower bounds of each data source's evaluation value are the expectation of both cloud E_x values. The entropy of the traditional cloud model is $E_{nij} = (C_{ij,max} - C_{ij,min})/6$, when the data is near the endpoint value, and the corresponding degree of certainty tends to 0. However, the endpoint value of the interval divided by each level is the transition boundary value of the two adjacent levels, and the edge value should belong to the upper and lower intervals at the same time. Therefore, in order to represent the boundary ambiguity of adjacent ranks, the divisor for finding the entropy is determined to be 2.355.

Let (E_{xij}, E_{nij}, H_e) be the three numerical properties of a cloud for a given one-dimensional domain, and the procedure for this one-dimensional normal cloud generator is:

1. Generate a normal random number E'_{nij} with E_{nij} as the expected value and ${H_e}^2$ as the variance.
2. Generate a normal random number x_{ij} with E_{xij} as the expected value and ${E'_{nij}}^2$ as the variance.
3. Calculate $y_{ij} = \exp\left(-\frac{\left(x_{ij}-E_{xij}\right)^2}{2{E'_{nij}}^2}\right)$, where x_{ij} is a specific quantified value, y_{ij} is the degree of determination of x_{ij} on qualitative index U, and $(x_{ij},\ y_{ij})$ is the cloud drop.
4. Repeat the above steps until N cloud drops are generated.

2.2. Dempster–Shafer Evidence Theory

Let $\Theta = \theta_1, \theta_2, \ldots, \theta_n$ be a finite identification framework in the D-S evidence theory, where $\Theta = \theta_1, \theta_2, \ldots, \theta_n$ are all possible events and $\theta_i (i = [1, n])$ is a subset of the recognition frame Θ. The underlying trust function m be a mapping from the set $2^{\Theta} \to [0, 1]$, with A being any subset of Θ and it satisfies

$$\begin{cases} m(\varnothing) = 0 \\ \sum_{A \subset \Theta} m(A) = 1 \end{cases} \quad (2)$$

We call m the basic probability assignment function (BPA function for short) of Θ [35], where $m(\varnothing)$ denotes the degree of confidence of the evidence in the empty set. If $m(A) > 0$, then A is called a focal element within the identification framework Θ, and $m(A)$ reflects the degree of trust of the evidence in A. In particular, the condition $m(\varnothing) = 0$ is not necessarily satisfied. For the open evaluation set space, $m(\varnothing)$ is not necessarily equal to 0. In this paper, we only consider the case in the closed evaluation set space.

For recognition framework $\Theta = \theta_1, \theta_2, \ldots, \theta_n$ and BPA function $m(A)$, $Bel(A)$ is defined as the confidence function, which is the sum of the potential probability assignments of all subsets of A, indicating the degree of certainty of the proposition A, as shown in Equation (3):

$$Bel(A) = \sum_{B \subset \mathrm{A}} m(B), \forall A \subset \Theta\,. \quad (3)$$

$Pl(A)$ is the likelihood function of A, as defined in Equation (4), indicates the degree of trust that does not deny A,

$$Pl(A) = 1 - Bel\left(\overline{A}\right) = \sum_{B \cap\ A \neq \varnothing} m(B). \quad (4)$$

The intervals of evidence are shown in Figure 1, where $[0, Bel(A)]$ is the support interval of proposition A, $[Bel(A), Pl(A)]$ is the uncertainty interval of proposition A, and $[Pl(A), 1]$ is the rejection interval of evidence. Among them, support interval and rejection interval together constitute the definite interval of evidence, which can represent the certainty of evidence.

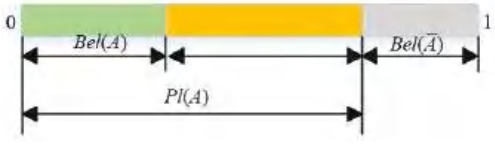


Figure 1. Diagram of evidence intervals.

Let m_1 and m_2 be two BPA functions on the same finite identification frame Θ, with focal elements $B_1, B_2, \ldots, B_n$ and $C_1, C_2, \ldots, C_n$. Then the D-S evidence theory combination rule rules are as follows in Equations (5) and (6):

$$m(A) = \begin{cases} \frac{1}{1-K} \sum_{B_i \cap C_j = A} m_1(B_i) m_2(C_j) , & A \neq \varnothing \\ 0 , & A = \varnothing \end{cases} \tag{5}$$

$$K = \sum_{B_i \cap C_j = \varnothing} m_1(B_i) m_2(C_j), \tag{6}$$

where K is the coefficient of evidence conflict between m_1 and m_2, the higher K value indicates the greater the degree of evidence conflict, and the values of K range from 0 to 1.

3. The Proposed Method

Based on the above theoretical knowledge, this paper proposes a heterogeneous data-fusion method based on a cloud model and improved evidence theory. In order to obtain the BPA function of evidence more accurately, we consider the ambiguity of multi-sensor data when completing data transformation by using the cloud model. To improve the reliability of the fusion results, we propose a new method for measuring the similarity of evidence and improve the evidence by combining the similarity and certainty of evidence together. The specific method for determining the BPA function and calculating the similarity of evidence and the certainty of evidence are described in this section, and finally the overall steps of the method are proposed.

3.1. Determination Method of BPA Function

It is assumed that the multi-sensor system's data information is pre-processed to extract n classes of data, forming n bodies of evidence, i.e., $X = (x_1, x_2, x_3, \ldots, x_n)$, where $x_i (i = [1, n])$ is the ith class of data measured by the system. Based on the knowledge gained from the cloud model, the membership degree $\mu_{ij}(k)$ for the values of discrete feature variables is calculated as follows in Equation (7):

$$\mu_{ij}(k) = e^{-\frac{(x_i - E_{xij})^2}{2E'^2_{nij}}}, \tag{7}$$

where $\mu_{ij}(k)$ is the membership of the ith class of data relative to the jth evaluation index under the kth judgment within the same acquisition cycle of the multi-sensor system, E_{xij} is the expectation value of class i data relative to the jth evaluation index obtained in Equation (1), and E'_{nij} is a normal random number generated with E_{nij} as the expectation and H_e as the standard deviation obtained in Equation (1).

k is the number of times the multi-sensor acquires data in the same acquisition cycle, when k is greater than 1, the membership of class i data with respect to the jth evaluation

index can be determined by the maximum of the k affiliation values when the feature parameters have multiple values:

$$\mu_{ij} = \max\left(\mu_{ij}(a)\right), a = 1, 2, \ldots, k. \tag{8}$$

The multi-sensor membership matrix can be calculated based on the membership degree μ_{ij}:

$$\boldsymbol{R}_{\boldsymbol{n}\times\boldsymbol{m}} = \begin{pmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1m} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ \mu_{n1} & \mu_{n2} & \cdots & \mu_{nm} \end{pmatrix}. \tag{9}$$

The elements in each row in Equation (9) represent the membership of the ith (i = 1, 2, $\cdots$, n) class of data of the multi-sensor for the jth (j = 1, 2, $\cdots$, m) evaluation index, and the elements in each column represent the membership of all data information collected by the multi-sensor system at a certain time for the jth (j = 1, 2, $\cdots$, m) evaluation index.

The obtained membership matrix $\boldsymbol{R}_{\boldsymbol{n}\times\boldsymbol{m}}$ basically satisfies the definition of probability assignment but does not satisfy $\sum_{j=1}^{m} \mu_{ij} = 1$. Considering that the actual use of the sensor will produce a certain measurement error, the following definition is added to transform the membership of each evaluation index into a BPA function:

$$\begin{cases} \gamma_i = 1 - \max(\mu_{i1}, \mu_{i2}, \ldots, \mu_{im}) \\ m_i(\Theta) = \gamma_i \\ m_i(A_j) = (1 - \gamma_i)\frac{\mu_{ij}}{\sum_{j=1}^{m} \mu_{ij}} \end{cases}, \tag{10}$$

where γ_i denotes the uncertainty of the ith characteristic parameter, $m_i(\Theta)$ is the basic probability assignment value of the uncertainty of the ith piece of evidence, and $m_i(A_j)$ is the basic probability assignment value of the jth evaluation index of the ith piece of evidence.

3.2. Similarity of Evidence

Classical measures for describing the relationship between evidence include: conflict coefficient K, Pignistic probability distance, Jousselme distance and cosine similarity, and so on. The computation of the conflict coefficient K is given in (6), and assuming that the evidence bodies m_1 and m_2 are BPA functions of the identification framework $\Theta = \theta_1, \theta_2, \ldots, \theta_n$, the calculation of the Pignistic probability distance, Jousselme distance, and cosine similarity is given below.

1. Pignistic probability distance [25]

Pignistic probability distance is a measure of conflicting relationships between evidence. Let the recognition frame $\Theta = \theta_1, \theta_2, \ldots, \theta_n$, m is the BPA function of Θ, and if $A \subseteq \Theta$, then

$$BetP_m(A) = \sum_{B\subseteq\Theta} \frac{|A \cap B|}{|B|} m(B) \tag{11}$$

is said to be the Pignistic probability of the focal element A.

Assuming that $BetP_{m_1}$ and $BetP_{m_2}$ are the corresponding Pignistic probability functions, the Pignistic probability distances are calculated as follows:

$$difBetP_{m_2}^{m_1} = max_{A\subseteq\Theta}(|BetP_{m_1}(A) - BetP_{m_2}(A)|). \tag{12}$$

2. Jousselme distance [26]

$$d_{BPA}(m_1, m_2) = \sqrt{\frac{1}{2}(\boldsymbol{m_1} - \boldsymbol{m_2})^T \boldsymbol{D}(\boldsymbol{m_1} - \boldsymbol{m_2})}, \tag{13}$$

where $\boldsymbol{m_1}$ and $\boldsymbol{m_2}$ are the vector forms of the evidences m_1 and m_2, and $\boldsymbol{D}$ is a $2^\Theta \times 2^\Theta$ positive definite matrix, its mathematical expression is: $D = \begin{pmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ d_{n1} & d_{n2} & \dots & d_{nn} \end{pmatrix}$, where the element $d_{ij} = J(\theta_i, \theta_j) = \frac{|\theta_i \cap \theta_i|}{|\theta_i \cup \theta_j|}$, θ_i is any focal element in evidence m_1 and θ_j is any focal element in evidence m_2, which can also be called the Jaccard coefficient and can be used to reveal the relationship between unifocal and multifocal elements of the evidence.

The Jousselme distance is a measure of the conflicting relationships of the evidence, and the higher its value, the greater the conflict between the evidence.

3. Cosine similarity [27]

The cosine similarity can be used to calculate the similarity of the evidence. The greater the cosine similarity, the greater the confidence between the evidence.

$$\cos(\boldsymbol{m_1}, \boldsymbol{m_2}) = \frac{\boldsymbol{m_1} \cdot \boldsymbol{m_2^T}}{||\boldsymbol{m_1}|| \cdot ||\boldsymbol{m_2}||}, \tag{14}$$

where $||\boldsymbol{m_i}|| = \sqrt{\boldsymbol{m_i} \cdot \boldsymbol{m_i^T}}$.

The accuracy of the various measurements is examined based on the above computation by calculating the measures under different conditions in conjunction with Example 1.

Example 1. Suppose there are identification frames $\Theta = \{a, b, c, d\}$ with different distributions of evidence bodies under different conditions, as shown in Table 1.

Table 1. Distribution of different bodies of evidence in different situations.

Situation	The Distribution of Evidence Body
Situation 1	$m_1(a) = m_1(b) = m_1(c) = m_1(d) = 0.25$ $m_2(a) = m_2(b) = m_2(c) = m_2(d) = 0.25$
Situation 2	$m_1(a) = m_1(b) = 0.5$ $m_2(c) = m_2(d) = 0.5$
Situation 3	$m_1(a) = m_1(b) = m_1(c) = 1/3$ $m_2(a, b, c) = 1$
Situation 4	$m_1(a) = 0.25, m_1(b) = 0.65, m_1(abc) = 0.1$ $m_2(a) = 0.65, m_2(b) = 0.25, m_2(abc) = 0.1$

The body of evidence under Situation 1 is identical, and its conflict coefficient K is calculated by using Equation (6), yielding 0.75, which contradicts the fact, whereas cosine similarity and the Jousselme distance yield 1, which is consistent with the fact. Situation 2's evidence is radically different, and the Jousselme distance metric produces 0.707, which is inconsistent with the facts, whereas the cosine similarity computation yields 0, which is consistent with the facts. Because it is impossible to determine whether the body of evidence m_2 under Situation 3 supports each focal element on average, the body of evidence under Situation 3 is somewhat conflicting, and the results of the Pignistic probability distance and cosine similarity are both 0, which contradict the facts, the result of the Jousselme distance is 0.577, which is more consistent with the facts.

From the above analysis, the cosine similarity measure is more accurate when measuring evidence with only a subset of single focal elements, and less accurate when faced with evidence containing a subset of multiple focal elements. Wang et al. [33] combined cosine similarity and the Jaccard coefficient to measure the relationship between evidence. But both measures are similarity measures, and the analysis of how the evidence relates to each other is not thorough enough. This can lead to inaccurate measurements in some situations, such as when evidence m_1 and m_2 in Situation 4 point to different correct propositions and

there is a big disagreement. However, Wang's method gives a similarity of 0.80, which is less consistent with the facts. Therefore, this paper proposes to combine conflicting evidence and similarity to jointly measure the relationship between evidence. Because the Jousselme distance can measure the relationship between evidence more accurately in most cases, and it is introduced to jointly measure the relationship between evidence.

Assuming the identification framework $\Theta = A_1, A_2, \ldots, A_n$, we define the local similarity of evidence s_{ij} as:

$$\begin{cases} J(A_a, A_b) = \frac{|A_a \cap A_b|}{|A_a \cup A_b|}, \forall A_a, A_b \subseteq \Theta \\ s_{ij} = (1 - d_{BPA}(m_1, m_2)) \times \frac{\sum_{a=1}^{n} \sum_{b=1}^{n} m_i(A_a) m_j(A_b) \times J(A_a, A_b)}{\sqrt{\sum_{c=1}^{n} m_i(A_c)^2} \sqrt{\sum_{c=1}^{n} m_j(A_c)^2}} \end{cases} \tag{15}$$

According to Equation (15), the local similarities of the evidence under different situations in Example 1 are: 1, 0, 0.244, and 0.470, all of which are more consistent with the facts. Based on the local similarity s_{ij}, the global similarity s_i can be derived for each piece of evidence, and its normalization can lead to the similarity-based weight coefficient α_i, which is calculated as follows:

$$\begin{cases} s_i = \sum_{j=1, i \neq j}^{n} s_{ij} \\ \alpha_i = \frac{s_i}{\sum_{j=1}^{n} s_j} \end{cases}. \tag{16}$$

3.3. Certainty of Evidence

The properties of the evidence itself can be measured based on the degree of certainty of the evidence. In probability theory, the Hellinger distance is a complete distance metric defined in the space of probability distributions and can be used to measure the similarity between two probability distributions. It has the advantage of stability and reliability compared to other metrics. Deng et al. [30] measured the uncertainty of the evidence itself by calculating the uncertainty interval distance of the evidence focal elements. However, finding the weight of the evidence based on uncertainty involves more steps and is more tedious than finding the weight based on certainty, so this paper proposes a method by which to combine the Hellinger distance of the evidence support interval and rejection interval to jointly measure the certainty of the evidence.

Suppose $X = \{x_1, x_2, \ldots x_n\}$ *and* $Y = \{y_1, y_2, \ldots y_n\}$ are two probability distribution vectors of the random variable Z, and the Hellinger distance is

$$Hel(X||Y) = \sqrt{\frac{1}{2} \sum_{i=1}^{n} (\sqrt{x_i} - \sqrt{y_i})^2}. \tag{17}$$

Assuming the identification framework $\Theta = A_1, A_2, \ldots, A_n$ and defining $DU(m_i)$ as the evidence certainty, the calculation of $DU(m_i)$ is as follows:

$$DU(m_i) = \sum_{j=1}^{n} \sqrt{2} \times \left(\sqrt{\frac{1}{2} \times \left[\left(\sqrt{Bel\left(m_i\left(A_j\right)\right)} - 0 \right)^2 + \left(1 - \sqrt{Pl\left(m_i\left(A_j\right)\right)} \right)^2 \right]} \right), \tag{18}$$

where $\sqrt{2}$ is the normalization factor. The Hellinger distance reaches its maximum when the evidence determines that the interval is [1,1] or [0,0], which leads to the calculation of the normalization factor: $\frac{1}{Hel[[1,1],[0,1]]} = \sqrt{2}$.

Normalizing the resulting determinacy $DU(m_i)$ obtains the weight of the evidence based on the determinacy:

$$\beta_i = \frac{DU(m_i)}{\sum_{j=1}^{n} DUm_j}. \tag{19}$$

3.4. Steps of the Proposed Method

Based on the above study, the specific steps of the proposed method in this paper are given as follows, and the flow chart is shown in Figure 2.

Step 1: After pre-processing the data from sensors, the BPA function of each data source related to the body of evidence is calculated by integrating the cloud model and each data evaluation index.

Step 2: With the obtained BPA function of each evidence, the weight α_i based on the similarity of evidence is calculated by combining Equations (15) and (16), and the weight β_i based on the certainty of evidence is calculated by combining Equations (18) and (19).

Step 3: With the weights α_i and β_i, the total weight of the evidence body is calculated and normalized to obtain the final weight ω_i, which is calculated as follows:

$$\begin{cases} \omega_i' = \alpha_i \times \beta_i \\ \omega_i = \frac{\omega_i'}{\sum_{j=1}^{n} \omega_j'} \end{cases}. \tag{20}$$

Step 4: Based on the weights ω_i, the original evidence is averaged and weighted to obtain the processed body of evidence m,

$$m(A) = \sum_{i=1}^{n} \omega_i \times m_i(A). \tag{21}$$

Step 5: Use Dempster's fusion rule to perform $n-1$ fusion for evidence body m.

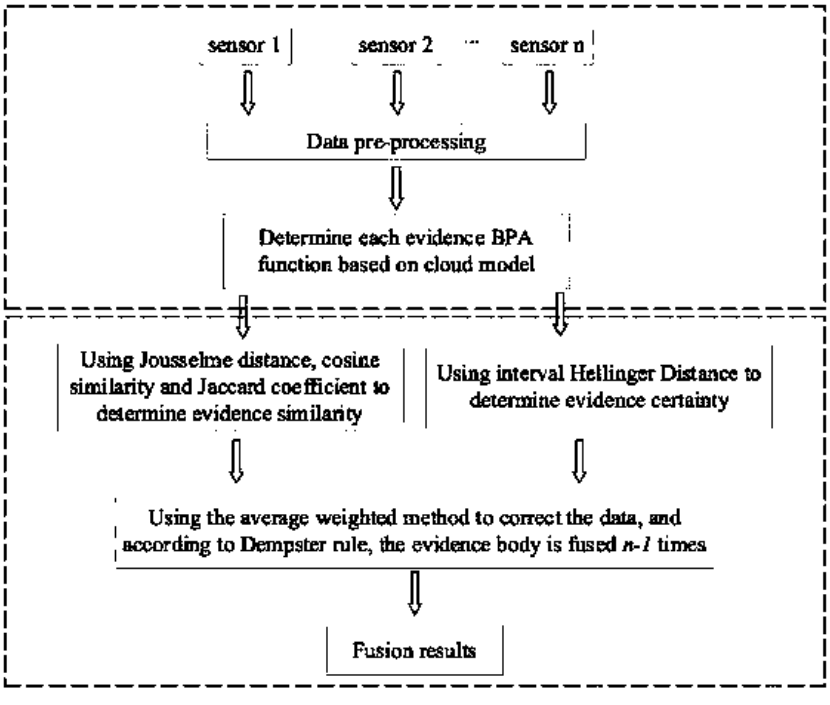


Figure 2. Flow chart of the proposed method.

4. Numerical Example and Simulation Results

In this section, the proposed improved D-S evidence theory method based on similarity and certainty, as well as the proposed overall method of heterogeneous data fusion based on cloud model and evidence theory, are evaluated and simulated to demonstrate the feasibility and effectiveness of the proposed method in this paper.

4.1. The method for Improving D-S Evidence Theory

In this section, four common conflicting, normal, and multi-quantity single-focal and multi-focal element evidences are fused based on the proposed improvement method, comparing traditional evidence theory, classical improvement methods, and similar improvement methods, and demonstrating the effectiveness of the proposed methods in this paper through Examples 2–4. We take the methods proposed by Deng Z. [30] and Wang [33] as similar improvement methods.

Example 2. In evidence theory, there are four common sorts of conflicts: complete conflict, 0 trust conflict, 1 trust conflict, and severe conflict [36], and the BPA functions for the four typical conflicts are provided in Table 2.

Table 2. Four common conflicting BPA functions.

Types of Conflict	Evidences	Proposition BPA				
		A	***B***	***C***	***D***	***E***
Complete conflict (k = 1)	m_1	1	0	0	\	\
	m_2	0	1	0	\	\
	m_3	0.8	0.1	0.1	\	\
	m_4	0.8	0.1	0.1	\	\
0 trust conflict (k = 0.99)	m_1	0.5	0.2	0.3	\	\
	m_2	0.5	0.2	0.3	\	\
	m_3	0	0.9	0.1	\	\
	m_4	0.5	0.2	0.3	\	\
1 trust conflict (k = 0.9998)	m_1	0.9	0.1	0	\	\
	m_2	0	0.1	0.9	\	\
	m_3	0.1	0.15	0.75	\	\
	m_4	0.1	0.15	0.75	\	\
High conflict (k = 0.9999)	m_1	0.7	0.1	0.1	0	0.1
	m_2	0	0.5	0.2	0.1	0.2
	m_3	0.6	0.1	0.15	0	0.15
	m_4	0.55	0.1	0.1	0.15	0.1
	m_5	0.6	0.1	0.2	0	0.1

The global similarity s_i and own determination $DU(m_i)$ of each evidence under the four conflict types are shown in Table 3. The weights α_i and β_i of the evidence can be calculated based on the degree of similarity s_i and the degree of certainty $DU(m_i)$, and the overall weight ω_i of the evidence can be obtained by combining the weights α_i and β_i. Figure 3 displays the distribution chart for each weight. Figure 3 shows that the weights of conflicting evidence are lower than those of normal evidence, and the distribution of each weight is consistent with the facts. We combined similarity and certainty to improve the body of evidence in order to improve the science of data fusion, and it should be noted that because the certainty of evidence describes the characteristics of the evidence itself, which includes the interval information of all focal elements within the evidence and is independent of the relationship between the evidence, the weights α_i and β_i are not always positively correlated.

Table 3. Similarity and certainty of each evidence under four conflicts.

Types of Conflict	Evidences	Global Similarity s_i	Determinacy $DU(m_i)$
Complete conflict	m_1	1.628	3
	m_2	0.036	3
	m_3	1.832	4.527
	m_4	1.832	4.527
0 trust conflict	m_1	2.141	6.009
	m_2	2.141	6.009
	m_3	0.423	3.785
	m_4	2.141	6.009
1 trust conflict	m_1	0.068	3.785
	m_2	1.715	3.785
	m_3	1.891	4.858
	m_4	1.891	4.858
High conflict	m_1	2.651	7.194
	m_2	0.631	8.127
	m_3	2.737	7.733
	m_4	2.779	7.987
	m_5	2.888	7.718

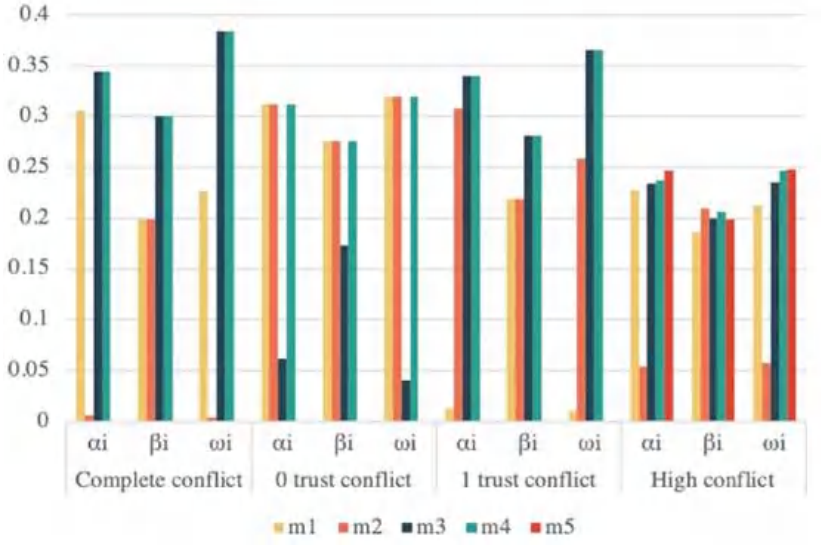


Figure 3. Four common types of conflicting evidence weights.

Table 4 displays the fusion results of the traditional D-S rule, the methods proposed by Sun [20], Murphy [23], Deng Y. [24], Deng Z. [30], and Wang [33], and the improved method proposed in this paper. As seen in Table 4, when confronted with the four conflicting situations listed above, the D-S fusion rule fails or does not match the genuine situation, and Sun's method allocates the uncertainty to the entire set, resulting in high BPA values for the entire set that do not fit the true situation. The larger the value of BPA after fusing, the greater the amount of confidence in the proposition. Although the approaches of Murphy, Deng Y., Deng Z., and Wang yield correct answers, the method proposed in this work yields a higher BPA function value and converges faster, demonstrating that the improved method in this research performs better than the other methods in resolving the four conflicts. The fusion BPA results on the reasonable propositions of each algorithm are shown in Figure 4.

Table 4. Fusion results of four common conflicts.

Types of Conflict	Methods	Proposition					Θ
		A	*B*	*C*	*D*	*E*	
Complete conflict	D-S	\	\	\	\	\	Invalid
	Sun	0.0917	0.0423	0.0071	\	\	0.8589
	Murthy	0.8204	0.1748	0.0048	\	\	0
	Deng Y.	0.8166	0.1164	0.0670	\	\	0
	Deng Z.	0.9792	0.0207	0.0001	\	\	0
	Wang	0.9996	0.0002	0.0002	\	\	0
	This paper	0.9999	0.0001	0.0001	\	\	0
0 trust conflict	D-S	0	0.7270	0.2730	0	0	0
	Sun	0.0525	0.0597	0.0377	\	\	0.8501
	Murthy	0.4091	0.4091	0.1818	\	\	0
	Deng Y.	0.4318	0.2955	0.2727	\	\	0
	Deng Z.	0.6510	0.2384	0.1106	\	\	0
	Wang	0.7628	0.2200	0.0172	\	\	0
	This paper	0.8421	0.0428	0.1151	\	\	0
1 trust conflict	D-S	0	1	0	0	0	0
	Sun	0.0388	0.0179	0.0846	\	\	0.8587
	Murthy	0.1676	0.0346	0.7978	\	\	0
	Deng Y.	0.1388	0.1318	0.7294	\	\	0
	Deng Z.	0.0273	0.0018	0.9709	\	\	0
	Wang	0.0006	0.0015	0.9980	\	\	0
	This paper	0.0001	0.0008	0.9991	\	\	0
High conflict	D-S	0	0.3571	0.4286	0	0.2143	0
	Sun	0.0443	0.0163	0.0163	0.0045	0.0118	0.9094
	Murthy	0.7637	0.1031	0.0716	0.0080	0.0538	0
	Deng Y.	0.5324	0.1521	0.1462	0.0451	0.1241	0
	Deng Z.	0.9846	0.004	0.0055	0.0001	0.0029	0
	Wang	0.9911	0.0025	0.001	0.0	0.0004	0
	This paper	0.9983	0.0002	0.0013	0.0	0.0002	0

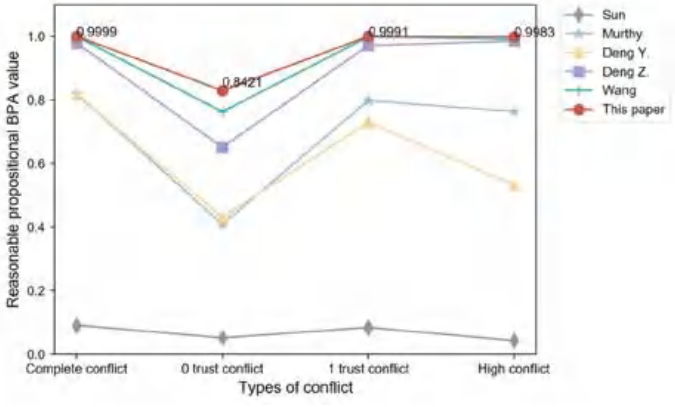


Figure 4. Comparison of reasonable proposition BPA value of different fusion algorithms.

Example 3. Assume the radar identification library contains three radar model data A, B, and C, with identification frame $\Theta = \{A, B, C, AC\}$. Five existing heterogeneous sensors are used separately to identify the radar radiation sources, yielding a total of five conflicting evidences ranging from m_1 to m_5. Tables 5 and 6 show the results of a specific two times, which represent the data distribution of multi-quantity single and multi-focal element conflict evidence, respectively.

Table 5. Single focal element evidence body data distribution.

Evidences	*A*	*B*	*C*
m_1	0.5	0.2	0.3
m_2	0	0.8	0.2
m_3	0.6	0.3	0.1
m_4	0.55	0.25	0.2
m_5	0.65	0.15	0.2

Table 6. Multi-focus evidence body data distribution.

Evidences	*A*	*B*	*C*	*AC*
m_1	0.5	0.2	0.3	0
m_2	0	0.9	0.1	0
m_3	0.55	0.1	0	0.35
m_4	0.55	0.1	0	0.35
m_5	0.6	0.1	0	0.3

The global similarity s_i and certainty $DU(m_i)$ of each evidence under single and multifocal elements are shown in Table 7. The weights of evidence α_i, β_i, and ω_i for a different number of evidence cases are shown in Figure 5. From Figure 5, it can be seen that the weight of conflicting evidence is less than the normal weight, the weight occupied by conflicting evidence gradually decreases as the number of evidence increases, and the distribution of each evidence is consistent with the facts, which proves the rationality of the method proposed in this paper.

Table 7. Similarity and certainty of evidence under single and multifocal elements.

Evidences	Global Similarity s_i		Determinacy $DU(m_i)$	
	Single-Focal Element	Multi-Focal Element	Single-Focal Element	Multi-Focal Element
m_1	2.743	6.009	2.496	6.009
m_2	0.858	4.485	0.345	3.785
m_3	2.756	5.599	3.685	3.221
m_4	2.983	5.868	3.685	3.221
m_5	2.999	5.434	3.730	3.422

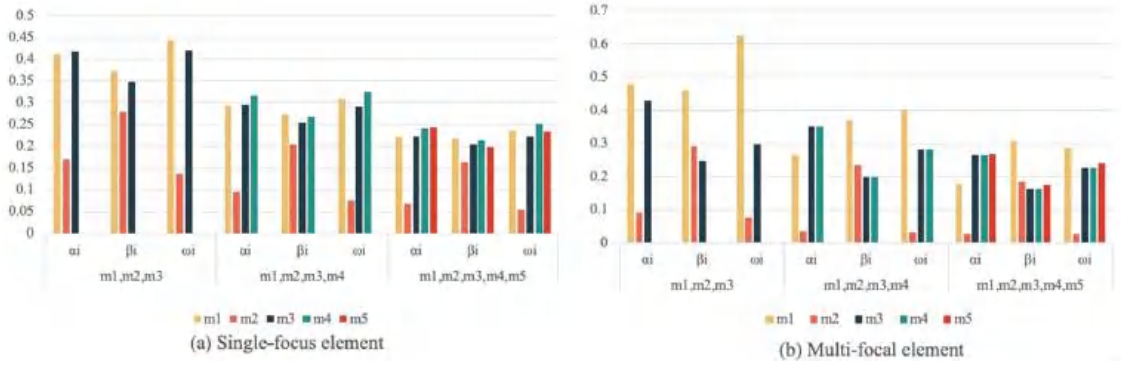


Figure 5. Weight of evidence under different amounts of evidence. (**a**) Single-focal element evidence weights. (**b**) Multi-focal element evidence weights.

To verify the effectiveness of the improved method proposed in this paper, the evidences are fused by using Murthy [23], Deng Y. [24], Deng Z. [30], and Wang [33], and the proposed method are fused respectively. Table 8 shows the fusion results for each method, and Figure 6 shows the comparison of BPA values for reasonable propositions. From the fusion results and comparison results, it can be concluded that when facing different numbers of single and multifocal element conflicting evidence bodies, the traditional D-S fusion results all contradict the facts. Although Murthy [23], Deng Y. [24], Deng Z. [30], and Wang [33] and the proposed method all point to the correct results, the BPA functions of the proposed method are higher than the other two improved methods, and as the number of evidence bodies increases, the improved method converges faster with higher accuracy on the BPA value as the number of evidence bodies increases.

Table 8. Multi-quantity evidence body fusion results.

Methods	m_1–m_3		m_1–m_4		m_1–m_5	
	Single-Focal Element	**Multi-Focal Element**	**Single-Focal Element**	**Multi-Focal Element**	**Single-Focal Element**	**Multi-Focal Element**
D-S	$m(A) = 0$ $m(B) = 0.9132$ $m(C) = 0.0868$	$m(A) = 0$ $m(B) = 0.6315$ $m(C) = 0.3685$ $m(AC) = 0$	$m(A) = 0$ $m(B) = 0.9293$ $m(C) = 0.0707$	$m(A) = 0$ $m(B) = 0.3287$ $m(C) = 0.6713$ $m(AC) = 0$	$m(A) = 0$ $m(B) = 0.9079$ $m(C) = 0.0921$	$m(A) = 0$ $m(B) = 0.1403$ $m(C) = 0.8597$ $m(AC) = 0$
Murthy	$m(A) = 0.3555$ $m(B) = 0.5868$ $m(C) = 0.0577$	$m(A) = 0.5568$ $m(B) = 0.3562$ $m(C) = 0.0782$ $m(AC) = 0.0088$	$m(A) = 0.5453$ $m(B) = 0.4246$ $m(C) = 0.0301$	$m(A) = 0.8656$ $m(B) = 0.0891$ $m(C) = 0.0382$ $m(AC) = 0.0074$	$m(A) = 0.8090$ $m(B) = 0.1785$ $m(C) = 0.0125$	$m(A) = 0.9688$ $m(B) = 0.0156$ $m(C) = 0.0127$ $m(AC) = 0.0029$
Deng Y.	$m(A) = 0.4978$ $m(B) = 0.4434$ $m(C) = 0.0588$	$m(A) = 0.6500$ $m(B) = 0.2547$ $m(C) = 0.0858$ $m(AC) = 0.0095$	$m(A) = 0.7418$ $m(B) = 0.2312$ $m(C) = 0.0270$	$m(A) = 0.9305$ $m(B) = 0.0274$ $m(C) = 0.0339$ $m(AC) = 0.0082$	$m(A) = 0.9277$ $m(B) = 0.0633$ $m(C) = 0.0090$	$m(A) = 0.9846$ $m(B) = 0.0024$ $m(C) = 0.0098$ $m(AC) = 0.0032$
Deng Z.	$m(A) = 0.6367$ $m(B) = 0.2631$ $m(C) = 0.1002$	$m(A) = 0.5669$ $m(B) = 0.3325$ $m(C) = 0.0966$ $m(AC) = 0.0044$	$m(A) = 0.6603$ $m(B) = 0.3095$ $m(C) = 0.0301$	$m(A) = 0.8389$ $m(B) = 0.1068$ $m(C) = 0.0507$ $m(AC) = 0.0036$	$m(A) = 0.8733$ $m(B) = 0.1152$ $m(C) = 0.0115$	$m(A) = 0.9136$ $m(B) = 0.0454$ $m(C) = 0.0357$ $m(AC) = 0.0053$
Wang	$m(A) = 0.6594$ $m(B) = 0.3119$ $m(C) = 0.0286$	$m(A) = 0.6581$ $m(B) = 0.2409$ $m(C) = 0.0937$ $m(AC) = 0.0073$	$m(A) = 0.8142$ $m(B) = 0.1604$ $m(C) = 0.0255$	$m(A) = 0.9391$ $m(B) = 0.0190$ $m(C) = 0.0342$ $m(AC) = 0.0077$	$m(A) = 0.9518$ $m(B) = 0.0401$ $m(C) = 0.0081$	$m(A) = 0.9859$ $m(B) = 0.0014$ $m(C) = 0.0096$ $m(AC) = 0.0031$
This paper	$m(A) = 0.7983$ $m(B) = 0.175$ $m(C) = 0.0267$	$m(A) = 0.8368$ $m(B) = 0.0478$ $m(C) = 0.1105$ $m(AC) = 0.0049$	$m(A) = 0.8842$ $m(B) = 0.0944$ $m(C) = 0.0221$	$m(A) = 0.9597$ $m(B) = 0.0028$ $m(C) = 0.0316$ $m(AC) = 0.0059$	$m(A) = 0.9849$ $m(B) = 0.0109$ $m(C) = 0.0026$	$m(A) = 0.9895$ $m(B) = 0.0003$ $m(C) = 0.0078$ $m(AC) = 0.0024$

Example 4. With the identification frame $\Theta = \{A, B, C\}$, there are five normal evidence bodies from m_1 to m_5, and the distribution is shown in Table 9.

Table 9. Normal evidence body data distribution.

Evidences	*A*	*B*	*C*
m_1	0.85	0.05	0.1
m_2	0.70	0.15	0.15
m_3	0.50	0.20	0.30
m_4	0.50	0.20	0.30
m_5	0.7	0.25	0.05

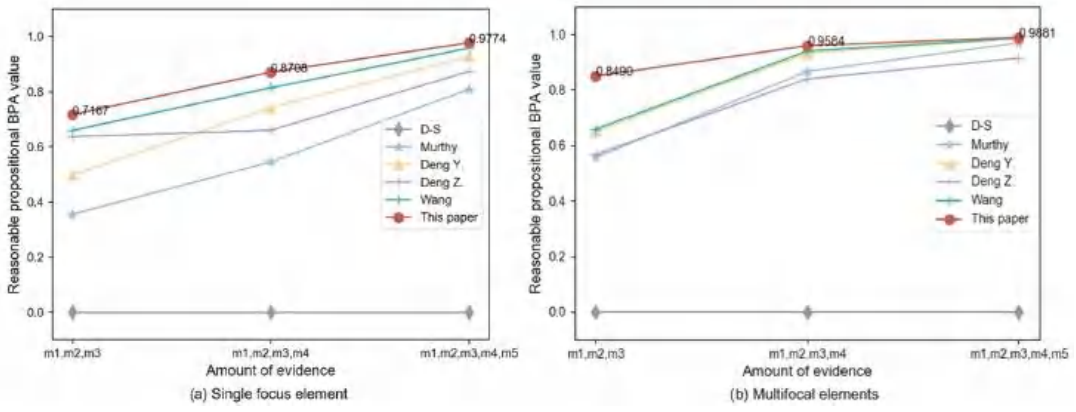


Figure 6. Comparison of the fusion results of multiple evidence. (**a**) Single focus element evidence fusion results. (**b**) Multi-focal element evidence fusion results.

The proposed improved method's fusion of normal evidence is compared to the traditional evidence theory to demonstrate the proposed improved method's superior performance in dealing with normal data, and the fusion results are shown in Table 10. Compared to the traditional evidence theory algorithm, the proposed method can also get correct results when dealing with normal body of evidence and has a higher BPA function with higher credibility.

Table 10. Normal evidence fusion result.

Methods	*m(A)*	*m(B)*	*m(C)*
D-S	0.9985	0.0007	0.0008
This paper	1	0	0

According to the above examples, the similarity and certainty-based evidence theory fusion algorithm proposed in this paper performs better in handling both normal and conflicting evidence bodies, demonstrating the improved method's rationality and effectiveness.

4.2. The Proposed Holistic Approach

To demonstrate the feasibility and effectiveness of the proposed data-fusion method, the heterogeneous data-fusion method combining cloud model and the proposed improved evidence theory in this paper is used for indoor early fire detection in this subsection.

It has been discovered that the combination of temperature, smoke concentration, and CO concentration has superior detection performance in fires [37], and the above information is collected as fire characteristic parameters in this paper. The fire discrimination results are divided into three categories: no fire, smoldering fire, and open fire. In the established fire identification framework $\Theta = \{\theta_1, \theta_2, \theta_3, \theta_1\theta_2\theta_3\}$, $\theta_1, \theta_2, \theta_3$ represent no fire, smoldering fire, and open fire, respectively, and $\theta_1\theta_2\theta_3$ indicates uncertainty of fire. Lin et al. [38] proposed a fire-detection method by using the Jousselme distance to improve the evidence corresponding to the fire characteristic parameters and fusing the evidence according to Dempster's rule to improve the timeliness of detection. However, the method ignored the characteristics of the evidence body itself and did not fully exploit the fire data information. Because the attribute values corresponding to the three fire characteristic parameters of CO concentration, smoke concentration, and temperature have certain stability and the interval distribution obeys normal distribution within a certain value interval [39], the cloud model

of each data source based on the forward cloud generator and the evaluation index of each parameter is built, and the cloud diagram is shown in Figure 7.

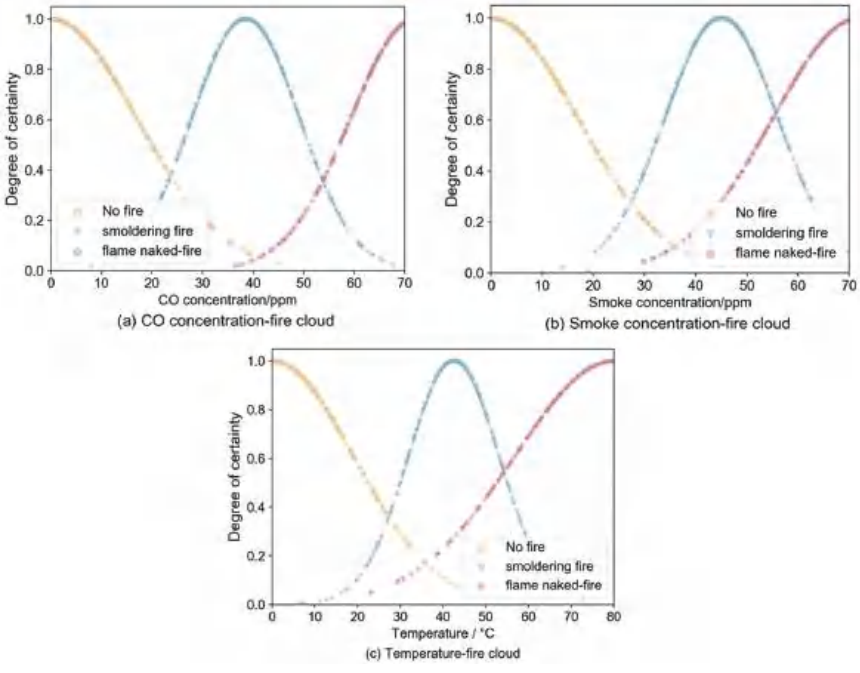


Figure 7. Fire characteristic parameter cloud chart. (**a**) CO concentration-fire cloud chart. (**b**) Smoke concentration-fire cloud chart. (**c**) Temperature-fire cloud chart.

PyroSim fire simulation software provides a visual user interface for fire dynamic simulation (FDS) and can more accurately predict the distribution of characteristic parameters such as fire smoke and temperature [40], so this paper simulates the occurrence of fire to obtain fire characteristic parameters. We build the indoor scenario as follows:

- The length, width and height of the room are 5, 5, 3 m;
- The room has a sofa, wooden bed and wooden table, in the upper left corner of the room from the wall 1 m set CO, temperature and smoke sensor group;
- Set the vent: room left wall with 1 × 1 m window, room directly opposite the sofa with 1.2 × 2 m door;
- The fire burning material is n-Heptane, the center of combustion is the center of the wooden bed, the burning area is 1 m^2.

By setting different heat release rate and heat ramp up time to simulate the occurrence of open fire and negative fire in the room, the starting room temperature is 30 °C, the simulation time is 30 s, and the data acquisition frequency is 2 Hz. Based on the proposed data-fusion method, a fire detection model is built. The initial fire detection probability is estimated by combining CO concentration, smoke concentration, and temperature data. The probability of smoldering fire and open flame within the initial fire detection probability is also added, and if it is greater than 0.75, the fire occurred in the room.

Figure 8a depicts a simulation of an open fire with visible fire and black smoke visible at t = 2.5 s. Figure 8b shows the change of the measured CO concentration, smoke concentration, and temperature data with time. When the probability of an open fire is 1, the values of CO, smoke, and temperature are the thresholds, and the time when each parameter first reached the threshold is shown in Figure 8b. The three characteristic parameters of CO, smoke, and temperature had almost no fluctuation in the first 2 s and increased rapidly

after 2 s. The temperature and smoke reached the threshold value relatively quickly, and all parameters showed an increasing trend in the first 30 s response time.

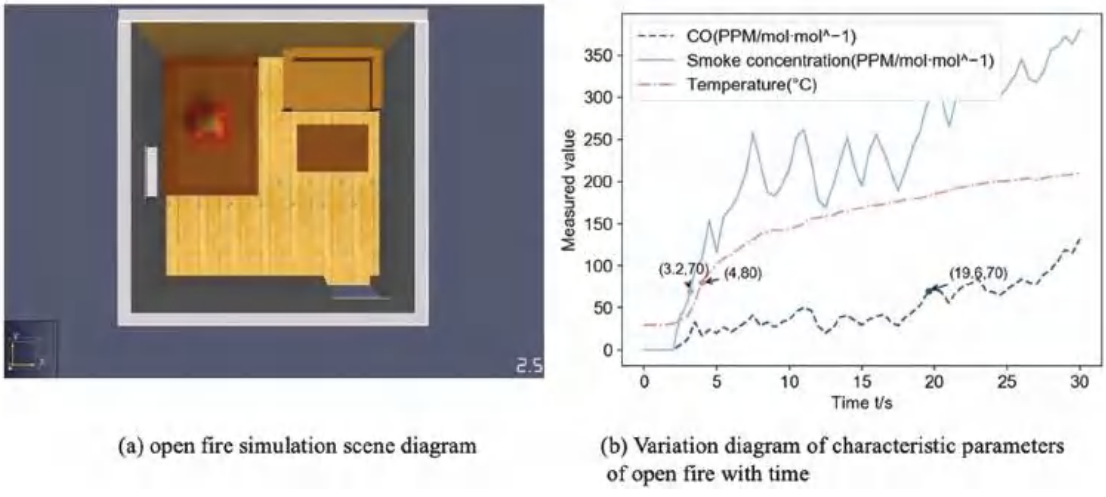


Figure 8. Open fire simulation information diagram.

To determine whether a fire has occurred, the early open fire data from this simulation are fused using the traditional D-S evidence theory, the methods proposed by Murphy [23], Deng Y. [24], Deng. Z. [30], and Wang [33] and this paper, respectively. Because the frequency of data acquisition in the simulation is 2 Hz, the period of data fusion is 0.5 s. The traditional evidence theory, Murphy's Deng Y's., and Deng. Z's methods all detect fire at t = 3.5 s, Wang's method detects fire at t = 3 s, and the proposed method detects fire at t = 2.5 s, proving the method's effectiveness. Figure 9 depicts the probability comparison of fire occurrence in this open fire scenario.

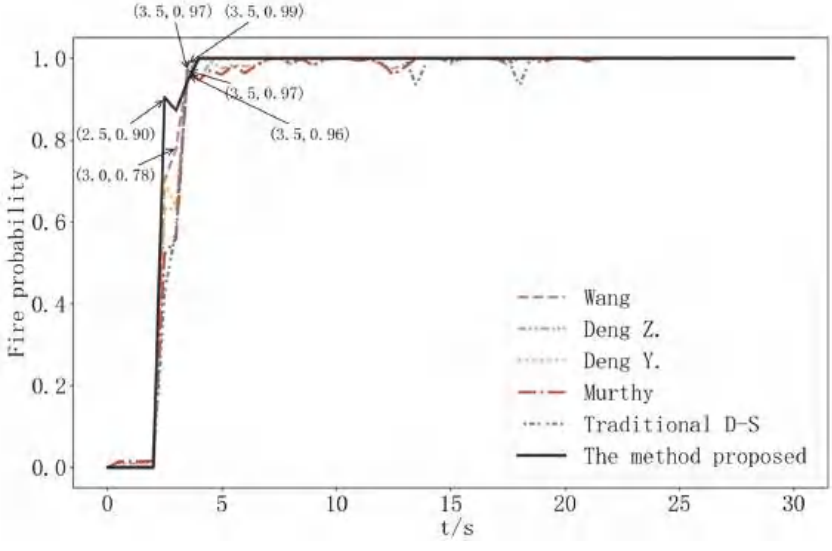


Figure 9. Fire occurrence probability comparison in open fire scene.

A smoldering fire's combustion features include the emission of a significant amount of black smoke from the combustion point prior to the appearance of the evident fire. Figure 10a depicts a simulation of a smoldering fire, with a clear fire visible at t = 18 s. Figure 10b displays a time-plot of the data collected by the multi-sensor group during the first 30 s. As shown in Figure 10b, the rising trend of each characteristic parameter in the shaded fire scenario is slower than it is in the open fire scenario, and the parameters only continue to grow after 7 s as a result of the early shaded fire's insufficient combustion.

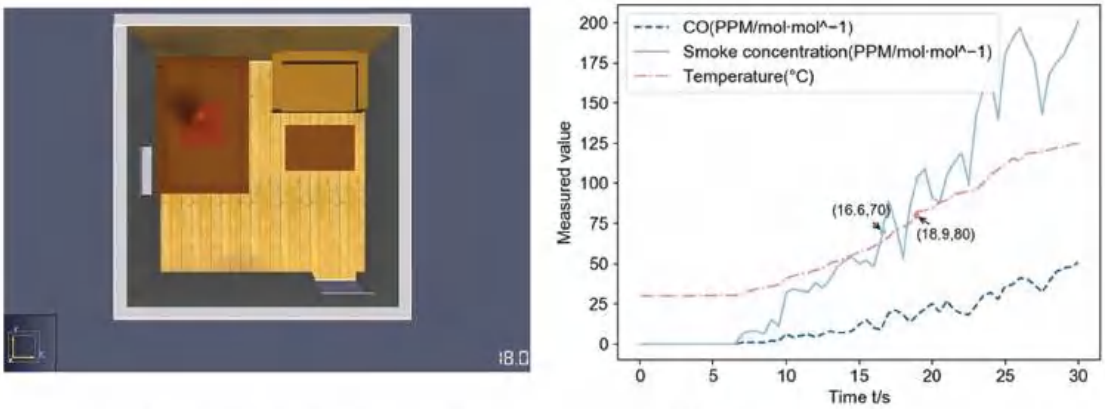


Figure 10. Smoldering fire simulation information diagram.

The determination of whether a fire has occurred is made possible by combining data on smoldering fires based on the traditional D-S evidence theory, the methods proposed by Murphy [23], Deng Y. [24], Deng. Z. [30], and Wang [33] and this paper, respectively. The method proposed in this paper can detect the occurrence of fire at t = 10 s, which is earlier than the 11.5 s of Wang's method, 11.5 s of Deng Y.'s method, 12 s of Deng Z's method, 13.5 s of traditional evidence theory, and 17 s of Murthy's method, as shown in Figure 11. As illustrated in Figure 11, when compared to the traditional evidence theory, classical improvement method, and similar improvement method, the method proposed in this paper not only detects the occurrence of fire in advance, but also has a higher detection accuracy.

To further verify the effectiveness of the proposed fire detection method, we obtained different CO concentration, smoke concentration and temperature data by setting different combustibles, combustion locations, heat release rates, and heat ramp-up times. Then we made our own fire dataset, which included 1000 positive samples and 1000 negative samples. Based on the traditional evidence theory, classical improvement method, similarity improvement method and the proposed method in this paper, the homemade samples are fused to calculate the accuracy rate and false alarm rate of detection. Assuming that TP represents the number of samples correctly judged to be fires, FN represents the number of samples not correctly judged to be fires, FP represents the number of samples misreported to be fires, and TN represents the number of samples correctly judged to be fires that did not occur. The accuracy and false alarm rates (FAR) are calculated as Equation (22):

$$\begin{cases} accurary = \frac{TP+TN}{TP+FN+FP+TN} \\ FAR = \frac{FP}{FP+TN} \end{cases} . \quad (22)$$

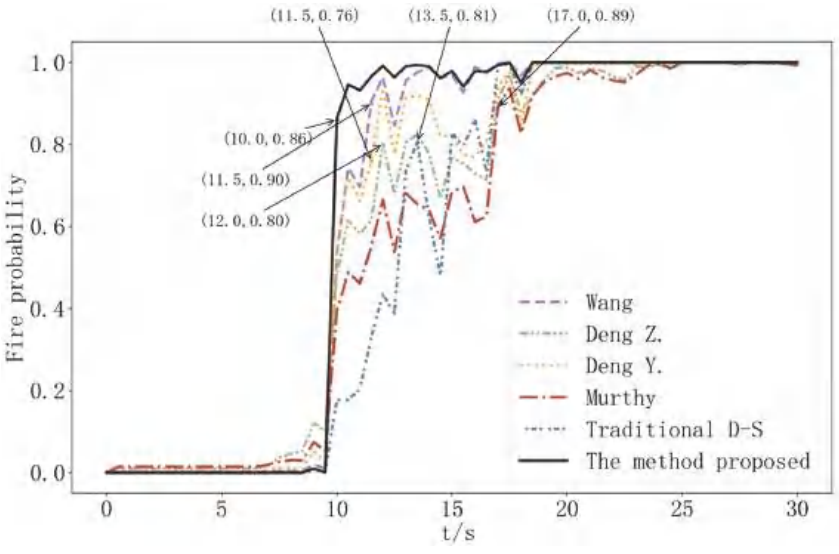


Figure 11. Fire occurrence probability comparison in smoldering fire scene.

Table 11 shows the fire detection accuracy and false alarm rate of various methods. According to Table 11, compared to other methods, the proposed method increased the fire detection rate by 0.7–10.2% and reduces the false alarm rate by 0.9–6.4%, which improves the reliability of fire discrimination obviously.

Table 11. Comparison of fire detection accuracy and false alarm rate of different methods.

Fusion Methods	Accuracy Rate	False Alarm Rates
Traditional D-S	88.6%	7.2%
Murthy	93.4%	5.6%
Deng Y.	96.6%	2.2%
Deng Z.	96.3%	3.1%
Wang	98.1%	1.7%
The method proposed	98.8%	0.8%

It is evident that when applied to indoor fire detection, the proposed heterogeneous data-fusion method has better fire detection performance and can simultaneously improve the timeliness and accuracy of detection, proving its feasibility and effectiveness in multi-sensor data fusion.

5. Conclusions

In this paper, a multi-sensor heterogeneous data fusion strategy based on the cloud model and improved evidence theory is presented, which can better cope with the ambiguity and conflict of heterogeneous multi-sensor gathered data. The cloud model is used to estimate the BPA function of each data source's associated evidence. Evidence similarity is calculated by using multi-relationship measures, evidence certainty is measured by using interval distance, the body of evidence is jointly improved by combining the evidence's similarity and certainty, and the improved body is fused by using Dempster's rule. The usefulness of the improved evidence theory technique is validated in this research, and the results show that the proposed method performs better when dealing with both conflicting and normal evidence. The method is used for indoor fire detection in light of the issues of prolonged duration and low accuracy. Compared to traditional evidence theory, classical improvement method, and similar improvement method, the proposed method improves

detection speed by 0.5–3 s, accuracy by 0.7–10.2%, and reduces the false alarm rate by 0.9–6.4%, which has better detection performance. It also provides a specific reference value for multi-source information fusion.

In future work, we intend to test the feasibility of the proposed method on other multi-sensor acquisition information systems, as well as investigate how to combine homogeneous and heterogeneous data fusion algorithms to fully exploit effective data information and improve data fusion accuracy.

Author Contributions: X.X. planned the study; K.L. designed the experiment and wrote the manuscript; Y.C. analyzed the experimental data; B.H. supervised and checked the completion of the study. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Zhejiang Provincial Natural Science Foundation (LY19F030004) and Zhejiang Provincial Key R&D Program Project (2018C01085).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors greatly appreciate the reviews, the suggestions from reviewers, and the editor's encouragement.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jiang, W.; Zhuang, M.; Xie, C. A Reliability-Based Method to Sensor Data Fusion. *Sensors* **2017**, *17*, 1575. [CrossRef] [PubMed]
2. Yang, Y.; Li, F.; Gao, Y.; Mao, Y. Multi-Sensor Combined measurement while drilling based on the improved adaptive fading square root unscented Kalman filter. *Sensors* **2020**, *20*, 1897. [CrossRef] [PubMed]
3. Zhou, T.; Chen, M.; Yang, C.; Nie, Z. Data fusion using Bayesian theory and reinforcement learning method. *Sci. China Inf. Sci.* **2020**, *63*, 170209:1–170209:3. [CrossRef]
4. Xiao, F. Evidence combination based on prospect theory for multi-sensor data fusion. *ISA Trans.* **2020**, *106*, 253–261. [CrossRef]
5. Liang, Y.; Tian, W. Multi-sensor fusion approach for fire alarm using BP neural network. In Proceedings of the 2016 International Conference on Intelligent Networking and Collaborative Systems (INCoS), Ostrava, Czech Republic, 7–9 September 2016; IEEE: Piscataway Township, NJ, USA; pp. 99–102.
6. Muñoz, J.; Molero-Castillo, G.; Benítez-Guerrero, E.; Bárcenas, E. Data fusion as source for the generation of useful knowledge in context-aware systems. *J. Intell. Fuzzy Syst.* **2018**, *34*, 3165–3176. [CrossRef]
7. Dempster, A.P. Upper and Lower Probabilities Induced by a Multi-valued Mapping. *Ann. Math. Stat.* **1967**, *38*, 325–339. [CrossRef]
8. Shafer, G. *A Mathematical Theory of Evidence*; Princeton University Press: Princeton, NJ, USA, 1976; Volume 24.
9. Zhang, H.; Wang, X.; Wu, X.; Zhou, Y. Airborne multi-sensor target recognition method based on weighted fuzzy reasoning network and improved DS evidence theory. *J. Phys. Conf. Ser.* **2020**, *1550*, 032112. [CrossRef]
10. Koksalmis, E.; Kabak, Ö. Sensor fusion based on Dempster-Shafer theory of evidence using a large scale group decision making approach. *Int. J. Intell. Syst.* **2020**, *35*, 1126–1162. [CrossRef]
11. Li, S.; Liu, G.; Tang, X.; Lu, Y.; Hu, J. An ensemble deep convolutional neural network model with improved DS evidence fusion for bearing fault diagnosis. *Sensors* **2017**, *17*, 1729. [CrossRef]
12. Li, G.; Liu, Z.; Cai, L.; Yan, J. Standing-posture recognition in human–robot collaboration based on deep learning and the dempster–shafer evidence theory. *Sensors* **2020**, *20*, 1158. [CrossRef]
13. Jiang, W.; Yang, Y.; Luo, Y.; Quin, X. Determining basic probability assignment based on the improved similarity measures of generalized fuzzy numbers. *Int. J. Comput. Commun. Control* **2015**, *10*, 333–347. [CrossRef]
14. Xu, P.; Deng, Y.; Su, X.; Mahadevan, S. A new method to determine basic probability assignment from training data. *Knowl. Based Syst.* **2013**, *46*, 69–80. [CrossRef]
15. Tang, Y.; Wu, D.; Liu, Z. A new approach for generation of generalized basic probability assignment in the evidence theory. *Pattern Anal. Appl.* **2021**, *24*, 1007–1023. [CrossRef]
16. Wang, G.; Xu, C.; Li, D. Generic normal cloud model. *Inf. Sci.* **2014**, *280*, 1–15. [CrossRef]
17. Peng, H.; Zhang, H.; Wang, J.; Li, L. An uncertain Z-number multicriteria group decision-making method with cloud models. *Inf. Sci.* **2019**, *501*, 136–154. [CrossRef]
18. Liu, J.; Wen, G.; Xie, Y. Layout optimization of continuum structures considering the probabilistic and fuzzy directional uncertainty of applied loads based on the cloud model. *Struct. Multidiscip. Optim.* **2016**, *53*, 81–100. [CrossRef]
19. Peng, B.; Zhou, J.; Peng, D. Cloud model-based approach to group decision making with uncertain pure linguistic information. *J. Intell. Fuzzy Syst.* **2017**, *32*, 1959–1968. [CrossRef]

20. Sun, Q.; Ye, X.; Gu, W. A new combination rules of evidence theory. *Acta Electonica Sin.* **2000**, *28*, 117.
21. Leung, Y.; Ji, N.; Ma, J. An integrated information fusion approach based on the theory of evidence and group decision-making. *Inf. Fusion* **2013**, *14*, 410–422. [CrossRef]
22. Haenni, R. Are alternatives to dempster's rule of combination real alternative. Comments on "About the belief function combination and the conflict management problem". *Inf. Fusion* **2002**, *3*, 237–239. [CrossRef]
23. Murphy, C.K. Combining belief functions when evidence conflict. *Decis. Support Syst.* **2000**, *29*, 1–9. [CrossRef]
24. Deng, Y.; Shi, W.; Zhu, Z.; Qi, L. Combining belief functions based on distance of evidence. *Decis. Support Syst.* **2004**, *38*, 489–493.
25. Wang, J.; Zhu, J.; Song, Y. A Self-Adaptive Combination Method in Evidence Theory Based on the Power Pignistic Probability Distance. *Symmetry* **2020**, *12*, 526. [CrossRef]
26. Jousselme, A.L.; Grenier, D.; Bossé, É. A new distance between two bodies of evidence. *Inf. Fusion* **2001**, *2*, 91–101. [CrossRef]
27. Dong, Y.; Cheng, X.; Chen, W.; Shi, H.; Gong, K. A cosine similarity measure for multi-criteria group decision making under neutrosophic soft environment. *J. Intell. Fuzzy Syst.* **2020**, *39*, 7863–7880. [CrossRef]
28. Yager, R.R. Entropy and specificity in a mathematical theory of evidence. *Int. J. Gen. Syst.* **2008**, *219*, 291–310.
29. Deng, Y. Deng entropy: A generalized Shannon entropy to measure uncertainty. *Artif. Intell.* **2015**, *6*, 176–188.
30. Deng, Z.; Wang, J. Measuring total uncertainty in evidence theory. *Int. J. Intell. Syst.* **2021**, *36*, 1721–1745. [CrossRef]
31. Tao, Y.; Zhu, X.; Yang, L. Multi-Sensor Data Fusion Based on Pearson Correlation Coefficient and Information Entropy. *Minicomput. Syst.* **2022**, 1–7. Available online: http://kns.cnki.net/kcms/detail/21.1106.tp.20220225.1128.006.html (accessed on 24 July 2022).
32. Xiao, F. Multi-sensor data fusion based on the belief divergence measure of evidences and the belief entropy. *Inf. Fusion* **2019**, *46*, 23–32. [CrossRef]
33. Wang, L.; Xing, Q.; Mao, Y. A weighted combination method of evidence based on trust and certainty. *J. Commun.* **2017**, *38*, 83–88.
34. Wu, H.; Zhen, J.; Zhangz, J. Urban rail transit operation safety evaluation based on an improved CRITIC method and cloud model. *J. Rail Transp. Plan. Manag.* **2020**, *16*, 100206. [CrossRef]
35. Jin, X.; Yang, A.; Su, T.; Kong, J.-L.; Bai, Y. Multi-channel fusion classification method based on time-series data. *Sensors* **2021**, *21*, 4391. [CrossRef] [PubMed]
36. Li, Y.; Chen, J.; Ye, F.; Liu, D. The improvement of DS evidence theory and its application in IR/MMW target recognition. *J. Sensors* **2016**, *2016*, 1903792. [CrossRef]
37. Wu, L.; Chen, L.; Hao, X. Multi-Sensor Data Fusion Algorithm for Indoor Fire Early Warning Based on BP Neural Network. *Information* **2021**, *12*, 59. [CrossRef]
38. Lin, S.; Wei, B.; Yang, H.; Xiong, Y.; Zhu, L.; Yu, L. D-S fusion detection method with new data sources. *J. Univ. Electron. Sci. Technol. China* **2021**, *50*, 861–867.
39. Liu, X.; Ma, W. A multi-sensor fire alarm method based on D-S evidence theory. *J. North China Univ. Technol.* **2017**, *39*, 74–81.
40. Zhang, H.; Yan, G.; Li, M.; Han, J. Analysis of the indoor fire risk based on the Pyrosim simulation. *Conf. Ser. Earth Environ. Sci.* **2021**, *636*, 012002. [CrossRef]

sensors

MDPI

Article

Motion Blur Kernel Rendering Using an Inertial Sensor: Interpreting the Mechanism of a Thermal Detector

Kangil Lee [1,2], Yuseok Ban [3] and Changick Kim [2,*]

[1] Agency for Defense Development, Daejeon 34060, Korea; traveler0802@kaist.ac.kr
[2] School of Electrical Engineering, Korea Advanced Institute of Science and Technology (KAIST), 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Korea
[3] School of Electronics Engineering, Chungbuk National University, 1 Chungdae-ro, Seowon-gu, Cheongju 28644, Korea; ban@cbnu.ac.kr
* Correspondence: changick@kaist.ac.kr

Abstract: Various types of motion blur are frequently observed in the images captured by sensors based on thermal and photon detectors. The difference in mechanisms between thermal and photon detectors directly results in different patterns of motion blur. Motivated by this observation, we propose a novel method to synthesize blurry images from sharp images by analyzing the mechanisms of the thermal detector. Further, we propose a novel blur kernel rendering method, which combines our proposed motion blur model with the inertial sensor in the thermal image domain. The accuracy of the blur kernel rendering method is evaluated by the task of thermal image deblurring. We construct a synthetic blurry image dataset based on acquired thermal images using an infrared camera for evaluation. This dataset is the first blurry thermal image dataset with ground-truth images in the thermal image domain. Qualitative and quantitative experiments are extensively carried out on our dataset, which show that our proposed method outperforms state-of-the-art methods.

Keywords: motion blur model; synthetic blurry thermal image; thermal detector; thermal image deblurring; blur kernel rendering; inertial sensor; gyroscope sensor

Citation: Lee, K.; Ban, Y.; Kim, C. Motion Blur Kernel Rendering Using Inertial Sensor: Interpreting the Mechanism of a Thermal Detector. *Sensors* **2022**, *22*, 1893. https://doi.org/10.3390/s22051893

Academic Editor: Manuel José Cabral dos Santos Reis

Received: 24 January 2022
Accepted: 24 February 2022
Published: 28 February 2022

1. Introduction

Infrared images are increasingly being used in various fields, e.g., commercial, medical, and military applications. Infrared cameras have been mainly used in industrial applications, such as thermal insulation performance measurement and electrical leakage testing [1]. Recently, new applications of infrared imaging are emerging. For instance, drones equipped with infrared cameras have been used to search for missing survivors at nighttime [2,3], and the infrared camera is becoming an essential sensor for autonomous vehicle driving at night to prevent accidents [4]. Furthermore, due to the outbreak of COVID-19, many applications measuring the body temperature of visitors at a building entrance have been widely used.

The infrared image sensor is a device that displays the thermal information of subjects as an image. The wavelength of the infrared band is longer than the visible band, being invisible to human eyes. The infrared band can be categorized into three types according to its wavelength: Short Wavelength Infrared (SWIR) with the wavelength ranging from 1.4 μm to 3 μm, Mid Wavelength Infrared (MWIR) with the wavelength ranging from 3 μm to 8 μm, and Long Wavelength Infrared (LWIR) with the wavelength ranging from 8 μm to 15 μm [5]. Due to the cost issue, most commercial applications use LWIR image sensors. More specifically, since SWIR and MWIR image sensors are fabricated based on compound semiconductors, they are more expensive than silicon-based visible and LWIR image sensors. Further, MWIR image sensors require a cryogenic system to maintain the sensor temperature at precisely 77K, which significantly increases the price, volume, and weight. Therefore, the MWIR image sensors have limitations in being used for commercial

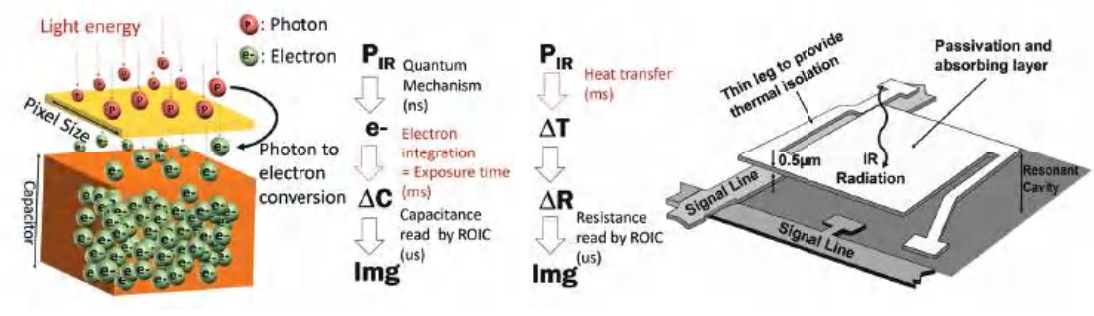
purposes. The cost of LWIR image sensors, on the other hand, is relatively low because they are fabricated based on the MEMS (Micro Electro Mechanical Systems) technology. Further, the LWIR image sensors can be manufactured in a very small since they do not need any cryogenic cooling system. The principle of the LWIR image sensors are different from the ones of CCD and CMOS image sensors which usually are for visible band images. The CCD and CMOS image sensors, so-called photon detectors, have semiconductor materials and structures that directly convert photons into electrons. In contrast, the LWIR sensors have the structure of a microbolometer [6]. This structure absorbs photons and changes them into heat. The LWIR sensors generate an image signal by detecting the temperature change induced by photons. The sensors having the mechanism of a microbolometer are called thermal detectors.

Traditional image processing tasks such as denoising [7–10], contrast enhancement [11], deblocking [12,13], inpainting [14,15], deblurring [16–19], and compressive sensing recovery [20,21] have been intensively studied in the visible image area since it is easy to acquire sufficient test data. However, due to domain dependency, image processing algorithms that properly work on a visible image are not guaranteed to work well on a thermal image. In general, the algorithms developed for the visible images tend to suffer from performance degradation in the thermal image domain. Therefore, it is essential to develop algorithms that directly consider the characteristics of the image domain. For example, in the studies on image quality metric, many efforts have been made to find appropriate metrics for thermal images [22–24]. Further, in the studies on image enhancement, many research proposals have been made to develop methods specialized for thermal images to solve problems such as low signal-to-noise ratio (SNR), halo effect, blurring, and low dynamic range compared to visible images [25–27].

The domain dependency can also be observed in the image deblurring area, where the two types of sensors produce apparently different motion blur patterns. The shape of a motion blur is very strongly related to the principle of image sensors, as shown in Figure 1. Photon detectors such as CCD and CMOS require time to physically collect photons, which is called exposure time (or integration time). If the camera or subject moves during the exposure time, motion blur occurs in the resulting image. In addition, the motion blur is easily observed at nighttime when the camera needs a longer exposure time. In contrast, the main cause of the motion blur in thermal detectors is the heat flow in a microbolometer structure. The microbolometer structure is designed and manufactured to provide good thermal isolation. Due to the thermal isolation of the microbolometer, time is needed for the heat to be transferred from one structure to another. The thermal detector generates images by measuring the temperature change of a microbolometer structure. Therefore, the remaining heat in the previous frame can appear as the motion blur in the next frame. As such, the photon detector and the thermal detector have different mechanisms for motion blur and produce different blur patterns in an image. As shown in Figure 2, the motion blur of the photon detector exhibits a linear blur pattern, whereas the thermal detector shows a blur pattern similar to a comet-tail shape.

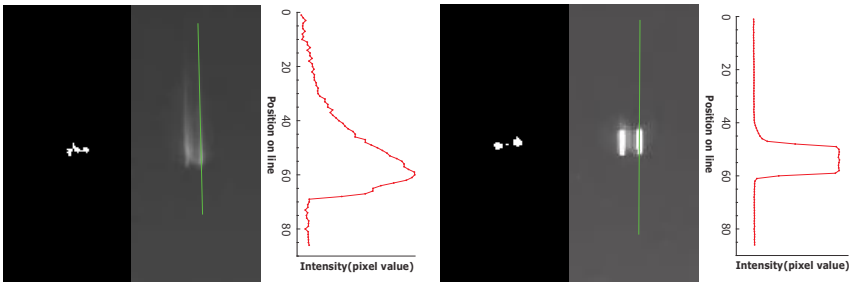
Several algorithms have been proposed to address this issue for thermal image deblurring. Oswald-Tranta [28] and Nihei et al. [29] observed that the motion blur in the LWIR image is different from that of the visible image and then proposed methods for image restoration. However, their image restoration experiments were conducted in limited conditions. The target's velocity was maintained with a constant at a fixed distance from the sensor, or the camera moved at a constant speed with its fixed direction. Consequently, their deblurring methods suffer from performance degradation when the size or orientation of the motion blur changes. Ramanagopal et al. [30] assumed the temporal sparsity of pixel-wise signals and performed motion deblurring on a thermal video using the LASSO (Least Absolute Shrinkage and Selection Operator) algorithm. However, it does not operate in real-time, and the deblurring fails when the temporal sparsity assumption is broken (e.g., fast camera motion). Zhao et al. [31] used the deep learning-based approach, a new GAN (Generative Adversarial Networks) structure for thermal image deblurring. However, the

training dataset was synthesized simply by averaging video frames without considering the characteristics of a motion blur in thermal images. Therefore, their method cannot be applied to thermal images with large motion blur. Batchuluun et al. [32] improved the deblurring performance by converting the one-channel thermal image into a three-channel thermal image. However, their method also did not consider how the motion blur occurs in thermal images when constructing the training dataset.



(**a**) (**b**)

Figure 1. The mechanism of two different sensors and cause of motion blur. (**a**) the cause of motion blur in the photon detector is integration time, (**b**) the cause of motion blur in the thermal detector is the response time of temperature change.



(**a**) (**b**)

Figure 2. Two kinds of cameras simultaneously take an image of the aircraft's twin-jet engine flames. Both images have motion blur, but they have different motion blur patterns. (**a**) LWIR camera using thermal detector, (**b**) MWIR camera using photon detector.

In fact, a number of deblurring methods have been studied based on visible images. Deep-learning-based methods have recently shown state-of-the-art performance in the image deblurring task, outperforming classic handcrafted methods. LSTM and CNNs are combined in SRN-DeblurNet [33] to deblur an image in a multi-scale manner. Pan et al. [34] proposed a method, in which neighboring video frames are warped into the center frame to use latent image information from adjacent frames for deblurring. Kupyn et al. [35] proposed a GAN-based structure, in which the feature pyramid networks balance performance and efficiency. Ye et al. [36] proposed a scale-iterative upscaling network with sharing weights to recover sharp images, and they used the super-resolution architecture for better performance. Zha et al. [18] proposed an effective algorithm for image deblurring by combining an optimization-based model with a deep neural network model. Although the deep learning-based method shows remarkable performance, the deblurring performance can still be significantly improved by incorporating the thermal image characteristics as well as by addressing the issue of the lack of datasets. Except for deep learning-based approaches, the most common and widely used approach for image deblurring is to estimate the blur kernel and sharp image simply using the observed blurry image [16,17,19]. In these

conventional methods, the latent image and blur kernels are obtained by minimizing the energy function with its constraints of statistics information. However, as a typical ill-posed problem, the conventional methods need large computational resources and often fail to deblur when the blur kernel size is large. So as to avoid these problems, the approach using an inertial sensor has been proposed especially for the blurry images caused by camera motions [37–47]. This approach has been evaluated as a method with great advantages over the existing blind deblurring method, in that the computational resources can be reduced by directly rendering the blur kernel with the inertial sensor information. However, all previous studies have proposed blur kernel rendering methods based on a photon detector model, which is generally used for visible images.

This paper proposes a novel motion blur kernel rendering method inspired by the sensing mechanism of a thermal image sensor and the supplementary information from a gyroscope sensor. Rendering the blur kernel by using gyroscope information is both efficient and accurate. It also enables the deblurring task through an efficient deconvolution. In our study, we interpret the microbolometer structure model in the aspect of motion blur, construct the motion blur model of the thermal image, and propose the method to efficiently and accurately render a blur kernel connoting the properties of the physical mechanism.

The main contributions of our study are summarized as follows:

- We propose a novel synthesis method for the blurring effect in the thermal image by interpreting the operating properties of a microbolometer.
- We propose the blur kernel rendering method for a thermal image by combining the gyroscope sensor information with the motion blur model.
- We acquire and publically release both actual thermal images and synthetic blurry thermal images for the construction of a dataset for thermal image deblurring.
- Our method quantitatively and qualitatively outperforms the latest state-of-the-art deblurring methods.

2. Image Generation and Motion Blur Model

There is a fundamental difference between a photon detector and a thermal detector in the principle of image generation. This section describes the mechanism of how the two detectors generate an image. Based on the analysis of detector mechanism, we propose an approach to synthesize the motion blur in a thermal image.

2.1. Photon Detector Model

A photon detector is based on a photodiode structure. When photons are incident on the p–n junction in the photodiode, electron-hole pairs are generated, and the electrical current flows along with the direction of the photodiode bias. The generated electrons are accumulated in a capacitor during the integration time. The integration time means the exposure time of a camera. The read-out integrated circuit (ROIC) outputs an image signal by measuring the charge stored in the capacitor.

$$I(i,j) = \int_0^{T_{int}} \Phi_{i,j}(t)dt. \tag{1}$$

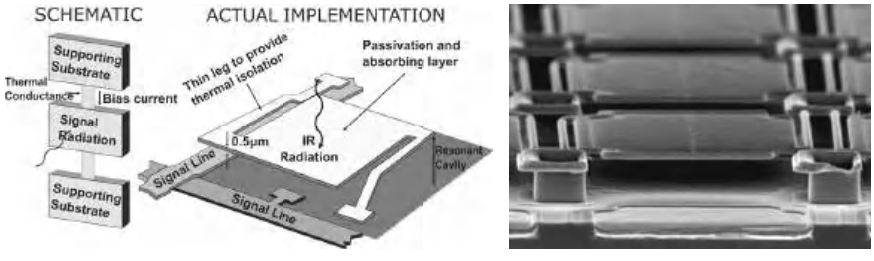
As can be seen in Equation (1), an image is corresponds to the sum of the incident photon energy during the integration time. The incident photon power is $\Phi_{i,j}(t)$, the image signal is $I(i,j)$, and the integration time is T_{int}, where (i,j) is the index of pixels in an image. Previous studies have used Equation (2) to generate a motion blur image from sharp images in the visible image domain [48–51].

$$B[n] = \frac{1}{n}\sum_{k=1}^{n} S[k]. \tag{2}$$

$S[k]$ denote the kth sharp image, which is equal to the incident photon power. n is the number of sampled sharp images during the exposure time.

2.2. Thermal Detector Model

The microbolometer sensor is the most frequently used device structure in a thermal detector. Since the fabrication cost of the microbolometer is relatively cheap than other structures, this structure is predominantly used for the mass-production of the uncooled infrared detector [6]. The operating mechanism of a microbolometer consists of four steps: (i) the incident photon energy is converted into thermal energy, (ii) the heat changes the device resistance, (iii) ROIC measures the amount of change in resistance, (iv) ROIC outputs an image signal proportional to the measuring value. The thermal isolation structure is essential for this four-stage operation to be conducted normally. The microbolometer supports a large sheet area with extremely thin legs for thermal isolation. The large sheet absorbs incident photons, and the generated heat is isolated by thin legs. The conceptual diagram of a microbolometer structure and substantive implementation are shown in Figure 3. The following Equation (3) expresses the heat flow of a microbolometer [52].



(**a**) (**b**)

Figure 3. (**a**) Microbolometer structure and Schematic model, (**b**) Microbolometer scanning electron microscope (SEM) image [53].

$$C_{th} \cdot \frac{d\Delta T}{dt} + \frac{d\Delta T}{R_{th}} = \eta \Phi(t). \tag{3}$$

C_{th}, R_{th}, $\Phi(t)$, ΔT and η denote thermal capacitance $(W \cdot K)$, thermal resistance$(K{\cdot}W^{-1})$, photon power (W), device temperature (K) and photon absorption rate, respectively. $C_{th}R_{th}$ is the thermal time constant value and is expressed as τ. Therefore, Equation (3) becomes Equation (4), and the solution of first-order differential equation is given as Equation (5).

$$\tau \cdot \frac{d\Delta T}{dt} + \Delta T = R_{th}\eta \Phi(t), \tag{4}$$

$$\Delta T(t) = \frac{R_{th}\eta}{\tau} \Phi(t) * e^{\frac{-t}{\tau}}. \tag{5}$$

Let $B(t)$ be a final output image. The temperature difference is converted into an image signal through the element resistance change. As a more specific expression, the temperature difference of the microbolometer and the signal level of an output image are proportional to each other [6]. Therefore, considering the scale factor, Equation (5) is expressed as Equation (6).

$$B(t) = K\Phi(t) * e^{\frac{-t}{\tau}}, \text{where} \quad K = \frac{R_{th}\eta}{\tau}. \tag{6}$$

It is important to note that the image generation models of a thermal detector and a photon detector are different as shown in Equations (6) and (1). In the case of the photon detector, the output signal is formed by accumulating incident photon energy. On the other hand, the output of the thermal detector is the convolutional result of incident photon energy and an exponential decay function. Therefore, the output images of the thermal

detector lose the signal value over time. The theoretical mechanism difference between the two detectors is observed by our experiments. Even though the photon detector and thermal detector acquire a moving subject simultaneously, the blur effects appear differently, as shown in Figure 2. The response time of the thermal detector is related to τ. A high τ value means that the device has a high response time, showing a large amount of motion blur in an image. In contrast, a low τ value indicates less amount of blur effect in an image due to the faster response of the device.

2.3. Generating the Synthetic Blurry Image in a Thermal Image

In order to actually use the thermal detector model, it is necessary to convert the continuous model into a discrete model. Therefore, for the discrete model, we propose a new assumption based on Equation (4). A sampling process is used to replace continuous-time with discrete-time. Through the sampling process, t is converted to t_k. By applying Backward Euler method [54], Equations (7)–(9) can be obtained based on Equation (4) using $\frac{d\Delta T(t_k)}{dt_k} \approx \frac{\Delta T(t_k) - \Delta T(t_{k-1})}{h}$.

$$\tau \cdot \frac{\Delta T(t_k) - \Delta T(t_{k-1})}{h} + \Delta T(t_k) = R_{th}\eta\Phi(t_k), \tag{7}$$

$$\begin{gathered} \Delta T(t_k) = \frac{\tau}{\tau + h}\Delta T(t_{k-1}) + \frac{h}{\tau + h}\Phi^{'}(t_k), \\ \text{where} \quad \Phi^{'}(t_k) = R_{th}\eta\Phi(t_k), \end{gathered} \tag{8}$$

$$\begin{gathered} \Delta T(t_k) = (1 - \alpha)\Delta T(t_{k-1}) + \alpha\Phi^{'}(t_k), \\ \text{where} \quad \alpha = \frac{h}{\tau + h}. \end{gathered} \tag{9}$$

$\Delta T(t_k)$ is proportional to $B(t_k)$, and $\Phi^{'}(t_k)$ is a sharp image, which can be rewritten by using $S(t_k)$. Furthermore, the formula for a single device can be expanded to an image array, and the formula should be as the following Equation (10).

$$B_{i,j}(t_k) = (1 - \alpha)B_{i,j}(t_{k-1}) + \alpha S_{i,j}(t_k). \tag{10}$$

The kth blurry image is expressed as the weighted sum of the blurry image at t_{k-1} and the sharp image at t_k. Equation (10) has the form of the Infinite Impulse Response (IIR) filter, and when the recursive term is eliminated, it becomes Equation (11).

$$B_{i,j}(t_k) = \alpha \sum_{n=1}^{k} (1 - \alpha)^{k-n} S_{i,j}(t_n). \tag{11}$$

The blurry thermal image $B_{i,j}(t_k)$ is expressed as the exponential average of sharp images $S_{i,j}(t_n)$. In a photon detector, sharp images are averaged over a certain exposure time to synthesize a blurry image, as shown in Equation (2). On the other hand, it can be observed that an exponential average is used for a thermal image.

One thing that remains is how many sharp images are needed to synthesize the exact motion blur effect in the thermal detector. To address this problem, we need to look at the assumption taken in Equation (7). In the Backward Euler method, it is assumed that $h = t_k - t_{k-1} \approx 0$, while h is the interval time between t_k and t_{k-1}. If the assumption $t_k \approx t_{k-1}$ is satisfied, then $\Phi(t_k) \approx \Phi(t_{k-1})$ also must be satisfied. Therefore, to satisfy $\Phi(t_k) \approx \Phi(t_{k-1})$, the translation using a sharp image must be less than one pixel during h. In other words, if the subject image focused on the sensor plane moves within one pixel during h, the subject does not change in the image. The assumption can be satisfied if

the shift between adjacent images is within one pixel. For example, if the camera rotation directly causes an image motion blur, the following Equation (12) must be satisfied.

$$h = t_k - t_{k-1} \leq \frac{IFOV}{\omega}. \tag{12}$$

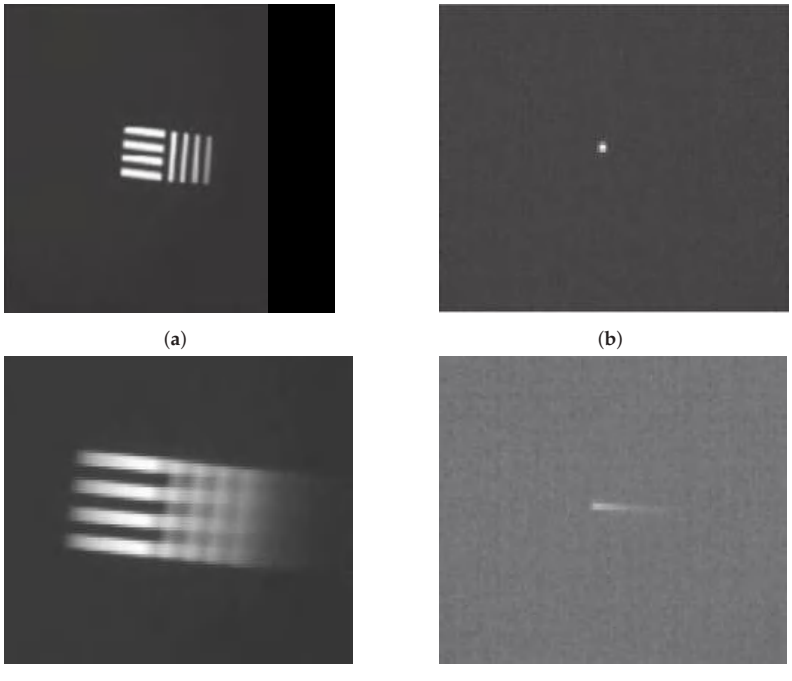
Instantaneous Field of View (IFOV) [55] is the field of view corresponding to a single pixel. ω is the angular velocity, which can be obtained when the camera rotates in the pitch or yaw direction. IFOV/ω is the time for an image to be shifted by one pixel. For example, if IFOV is 0.1° and the angular velocity of a camera is 100°/s, time interval h required for synthesis is 1 ms (where h is 1 ms, having the sharp image frame rate as 1000 Hz).

2.4. Verification of Thermal Detector Blur Model

This section describes the verification of our thermal detector blur model through experiments. Two test patterns are acquired using FLIR *A*655*sc* thermal camera and a collimator. Firstly, *A*655*sc* thermal camera was installed on the pan/tilt mount and rotated to collect real blurry images. Sharp images are obtained when the camera is stopped. The blurry images are synthesized by applying our thermal detector blur model to the sequential frames of sharp images. The model verification is achieved by quantitatively comparing real blurry images with synthetic blurry images.

2.4.1. Acquiring a Real Blurry Image

Real blurry images are acquired by rotating the camera at a certain angular velocity. The infrared camera is installed on a pan/tilt framework to precisely control the rotation speed. The image sensor plane is aligned with the rotation center. The camera rotation speed is 40°/s. Point source and 4-bar patterns are used as simple targets. The test patterns in a sharp image and a real blurry image are shown in Figure 4c,d, respectively.

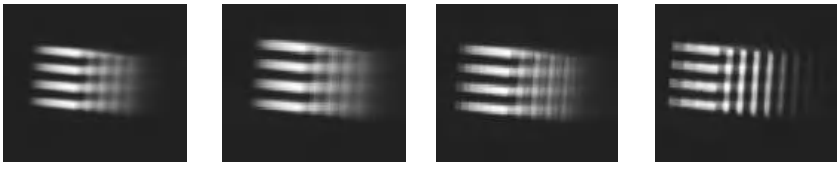


(**a**) (**b**) (**c**) (**d**)

Figure 4. Examples of motionless and moving pattern images. (**a**) 4-bar pattern, (**b**) Point source, (**c**) 4-bar pattern at 40°/s, (**d**) Point source at 40°/s.

2.4.2. Obtaining a Synthetic Blurry Image from Sharp Images

The set of sharp images with a high frame rate is required to generate synthetic blurry images via Equation (10). According to the previous section, a set of sharp images must be shifted by less than one pixel from adjacent frames. As shown in Figure 4a,b, we acquire a sharp image while the camera is stopped, and the set of sharp images is generated by shifting the image. The set of sharp images is used as $S_{i,j}(t_k)$ in Equation (10). If the sharp images are shifted by more than one pixel, the synthetic blurry image suffers from the stepping effect, as shown in Figure 5. The stepping effect makes synthetic blurry images have low similarity with real blurry images and makes them difficult to use either for training or for evaluation. In this experiment, the maximum rotation speed of a camera is 40°/s, and IFOV of FLIR *A*655*sc* is 0.0391°. Hence, the time interval h is 0.978 ms for synthesizing a blurry image without any stepping effect.

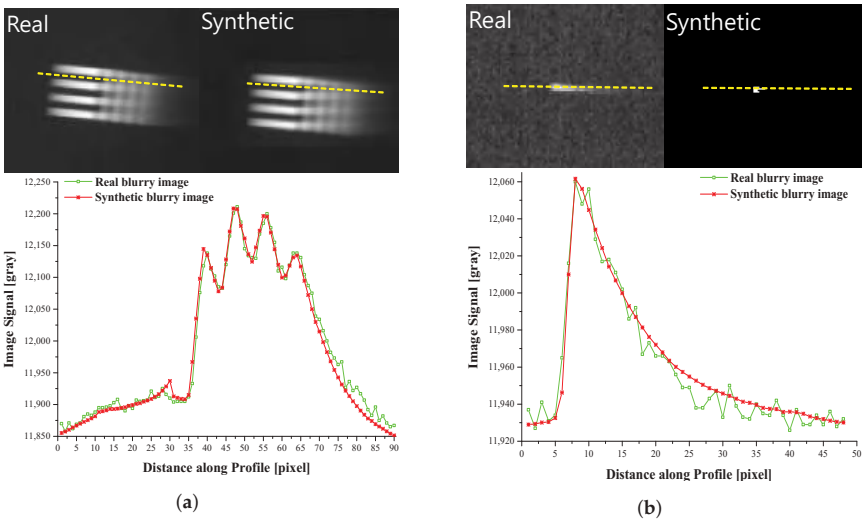


(**a**) (**b**) (**c**) (**d**)

Figure 5. Examples of stepping effects. (**a**) Shifting one pixel between adjacent frames, (**b**) Shifting two pixels between adjacent frames, (**c**) Shifting four pixels between adjacent frames, (**d**) Shifting eight pixels between adjacent frames.

2.4.3. Comparing Real and Synthetic Blurry Images

Figure 6 shows the real and synthetic blurry images when the camera rotation speed is 40°/s. In both test patterns, the comet tail shape appears in the opposite direction of a target movement. Even though the camera is rotating at a constant speed, the asymmetric blur phenomenon occurs. There is no difference in the position and value of the peak point of a signal value between real and synthetic blurry images. Therefore, the two signal profiles show high similarity, which means that our model has the sufficient ability to synthesize a blur effect.



(**a**) (**b**)

Figure 6. The comparison of real blurry images and synthetic blur images. (**a**) 4-bar pattern, (**b**) Point source.

3. Blur Kernel Rendering Using a Gyroscope Sensor for a Thermal Detector

The gyroscope sensor provides reliable information for rendering the blur kernel in the blurry images caused by camera motions. The blur kernel rendering methods with the assistance of an external sensor have been studied in many papers [37–47]. However, all approaches have been conducted in the visible image domain based on a photon detector. We propose the first blur kernel rendering method using an inertial sensor in the thermal image domain, leveraging the physical model of a thermal detector.

3.1. Blur Kernel Rendering and Gyroscope Data Selection

When a camera has motion, the relationship between the real-world scene and the image on a camera sensor plane is expressed as a homography transform [56]. In this case, the camera motion is expressed by translation and rotation. The intrinsic matrix of a camera is expressed in Equation (13), where f is the focal length, (p_{x_0}, p_{y_0}) is the principal point, and s is the skew parameter.

$$\begin{bmatrix} f & s & p_{x_0} \\ 0 & f & p_{y_0} \\ 0 & 0 & 1 \end{bmatrix} \tag{13}$$

We assumed the principle point and skew parameter to be 0. If the distance between a camera and a target is d, the rotation matrix is $R(\theta)$, the translation vector is $\boldsymbol{t}$, and the normal vector of a scene is $\boldsymbol{n}$. Then, the warping matrix and the rotation matrix are expressed by Equations (14) and (15), respectively.

$$H(\boldsymbol{t}, \theta) = K\left(R(\theta) - \frac{\boldsymbol{t}\boldsymbol{n}^{\mathsf{T}}}{d}\right)K^{-1}, \tag{14}$$

$$R(\theta) = \begin{bmatrix} cos\ \theta_x & -sin\ \theta_x & 0 \\ sin\ \theta_x & cos\ \theta_x & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} cos\ \theta_y & 0 & sin\ \theta_y \\ 0 & 1 & 0 \\ -sin\ \theta_y & 0 & cos\ \theta_y \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\ \theta_z & -sin\ \theta_z \\ 0 & sin\ \theta_z & cos\ \theta_z \end{bmatrix}. \tag{15}$$

If the distance between a subject and a camera is longer than the focal length, the camera rotation is the dominant factor in the warping matrix rather than camera translation [57–59]. Therefore, according to the above assumption, Equation (14) can be approximated as Equation (16).

$$H(\theta) = KR(\theta)K^{-1}. \tag{16}$$

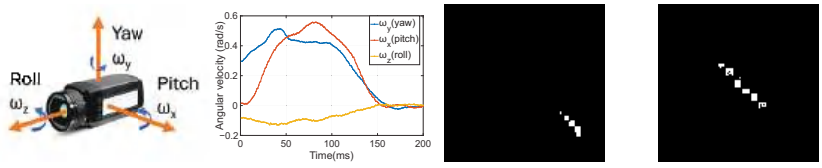
It is reported in several studies that the path of a light point source, which is called a light streak in blurry images, corresponds to the shape of a blur kernel [60]. Generally, the blur kernel is expressed as the cumulative sum of unit impulse functions during the exposure time T in a camera using the photon detector. Therefore, the relationship between a camera motion and a blur kernel is as the following Equation (17). $\delta[x, y]$ is the unit impulse function, f_g is the gyroscope frame rate, and N_p is the total number of gyroscope data during the exposure time.

$$\begin{aligned} k_p[x, y] &= \frac{1}{N_p}\sum_{i=1}^{N_p}\delta[x - x_i, y - y_i], \\ \text{where} \quad (x_i, y_i, 1) &= KR(\theta(t_i))K^{-1}(x_0, y_0, 1), \quad N_p = Tf_g. \end{aligned} \tag{17}$$

The warping matrix of a thermal detector is identical to that of a photon detector case, but their image generation models are different. The blur kernel rendering method in the thermal image domain is expressed in Equation (18) by combining Equations (11) and (16). Since the exponential decay term causes the signal attenuation effect in Equation (18), the result of blur kernel rendering resembles a comet tail shape. Figure 7 shows the camera axis and the blur kernel rendering results. Since the position of a point source transformed

through the warping matrix is not expressed as an integer, the bi-linear interpolation is conducted. $(1-(1-\alpha)^{N_t})$ is the normalization term to make the summation of the blur kernel be one. f_g and N_t are the gyroscope frame rate and the total number of gyroscope data during $m\tau$ in Equation (17), respectively.

$$k_t[x,y] = \frac{\alpha}{(1-(1-\alpha)^{N_t})} \sum_{i=1}^{N_t} (1-\alpha)^{N_t-i} \delta[x-x_i, y-y_i], \quad \text{where} \quad N_t = m\tau f_g. \tag{18}$$



(**a**) (**b**) (**c**) (**d**)

Figure 7. Illustration of camera rotation. (**a**) 3-axis rotation model, (**b**) Rotation motion measured by gyroscope sensor, (**c**) Blur kernel rendering result using the thermal detector model, (**d**) Blur kernel rendering result using the photon detector model.

The rotation matrix is required to implement the formula of blur kernel rendering. The angular information of each axis in the rotation matrix can be obtained through the gyroscope sensor. Since the gyroscope is a sensor that measures the angular velocity, the angle can be calculated by integrating the measured values over time. Next, we should decide the number of gyroscope data. In the case of a photon detector, the number of gyroscope data is easily determined by the exposure time, which induces the blur effect. In contrast, the blur effect of a thermal detector is caused by the thermal time constant in the microbolometer structure. Therefore, it is necessary to define the number of gyroscope data based on the thermal time constant τ. According to the modeling result in Equation (18), All gyroscope data stored during the entire duration are required for blur kernel rendering. However, the practical length of gyroscope data for rendering is limited due to the signal attenuation characteristics of the thermal detector. We confirmed that it is sufficient if the length of gyroscope data is at least five times the thermal time constant, or $m = 5$. For instance, if τ is 8 ms, obtaining gyroscope data for 40 ms is enough to synthesize the blur kernel.

3.2. Calibration and Blur Kernel Refinement

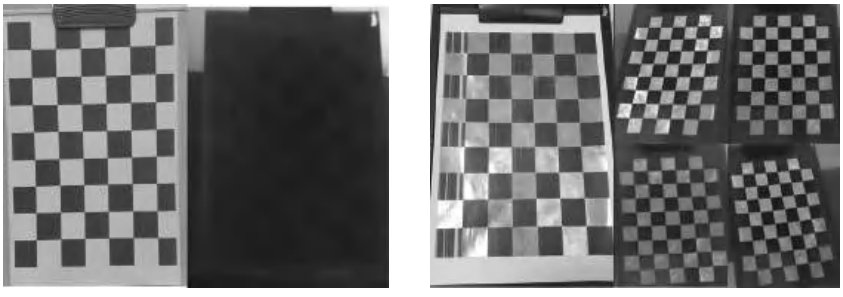
We calibrate a camera and a gyroscope using the open-source code for calibration [61]. Generally, the calibration process can be conducted by a standard checkerboard pattern in a visible image. On the other hand, the thermal camera cannot display a standard checkerboard pattern without temperature variations. To solve this problem, we use aluminum tapes whose emissivity is different from that of paper, as shown in Figure 8.

We conduct the refinement process for synthesizing the blur kernel as realistic as possible. The uniform blur effect appears even if there is no camera movement due to the optical Point Spread Function (PSF). The optical PSF is known to occur due to the diffraction and aberration of a camera lens system. Even for an ideal point source, a blur spot appears on the sensor plane by optical PSF [62]. Since diffraction increases as wavelength increases, the optical PSF is larger in an infrared band than in a visible band. Then, a refinement process considering the optical system is necessary to utilize the blur kernel rendering method in the infrared band. Precise optical measurement systems are required to synthesize an accurate optical PSF. However, these systems consume enormous time and cost. Instead, an efficient approximation formula is used in our method. As the primary cause of optical PSF, the diffraction blur spot size is expressed as an airy disk

function. The airy disk equation is approximated as Gaussian function, and its standard deviation is expressed by Equation (19) [63].

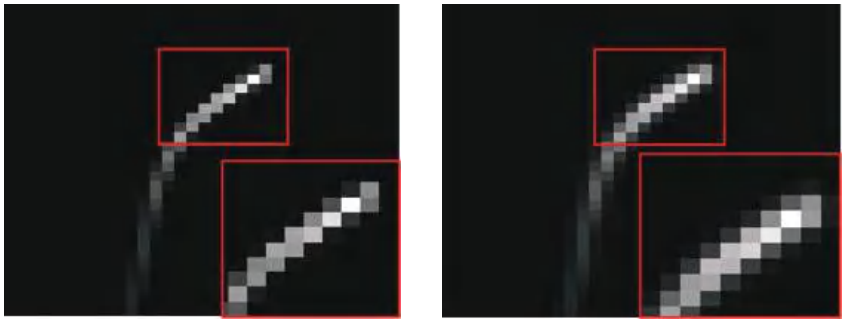
$$\sigma = 0.45 \cdot \frac{\lambda \cdot f/\sharp}{\beta} \quad . \tag{19}$$

where (19), λ is the infrared wavelength, $f/\sharp$ is the F-number, and β is the weighting factor to reflect the optical aberration effect. When β is 1, it directly means a diffraction-limited lens with no optical aberration effect. We determined the value of β with reference to the Strehl ratio to apply the optical aberration effect. Here, the Strehl ratio is defined as the peak intensity ratio of the center between a real PSF and an ideal PSF without aberrations [64]. Finally, the refined blur kernel can be calculated through the convolution between the blur kernel rendering result and the Gaussian function with the deviation value as σ shown in Equation (19). The blur kernel refinement results are presented in Figure 9.



(**a**) (**b**)

Figure 8. The calibration pattern for a thermal signal. (**a**) An ordinary checkerboard pattern (captured in visible-band and infrared band), (**b**) The checkerboard pattern improved by attaching aluminum material (captured in visible-band and infrared band).



(**a**) (**b**)

Figure 9. (**a**) Blur kernel before refinement, (**b**) blur kernel after refinement (given λ = 10 μm, $f/\sharp$ = 1.0, β = 0.6).

4. Experimental Setup

4.1. Construction of Synthetic Blurry Thermal Image Dataset

Most of the datasets for evaluating deblurring performance consist of visible band images, while thermal image datasets with ground truth images cannot be found. In this paper, we introduce the first Synthetic Blurry Thermal Image (SBTI) dataset with ground truth images in the thermal image domain. Firstly, we constructed the Sharp Thermal Image (STI) dataset using FLIR *A*655*sc* LWIR camera. The gyroscope sensor was mounted on the camera to measure the camera rotation speed. The LWIR camera was installed on a

tripod to synthesize the uniform blurry image by suppressing the roll movement. Table 1 shows the camera and gyroscope sensor parameters.

Table 1. The Parameters of Camera-Gyroscope integrated system.

Camera Parameters		**Gyroscope Parameters**	
Resolution (pixel)	640×480	Resolution (°/s)	0.0076
Frame rate (Hz)	50	Frame rate (Hz)	1000
FOV/IFOV (°)	$25 \times 19/0.0391$	Range (°/s)	± 200
Thermal time constant (ms)	8	Bias drift (°/s)	0.12
Focal length (mm)/$f/\sharp$	24.6/1.0	Total RMS noise (°/s)	0.05

As depicted in Figure 5, in order to synthesize a blurry thermal image without the stepping effect, adjacent images should be shifted by at most one pixel. Therefore, the maximum rotation angle of a camera between two adjacent images should be limited to the angle of IFOV. Since the IFOV of a FLIR camera is 0.0391°, and the frame rate is 50 Hz, the above condition can be satisfied if the camera rotation speed should be less than 1.955°/s. Since a gyroscope measures the angular velocity of a camera, the camera rotation speed is able to keep less than 1.955°/s during image acquisition. As shown in Table 2, the total number of images in each subset of the SBI dataset is between 1400 and 2000. The gyroscope data has been stored while synchronized with sharp images. Since the gyroscope frame rate is 1000 Hz, the camera rotation motion between adjacent images has been paired with 20 consecutive gyroscope data.

Table 2. Configuration of STI Dataset.

STI Dataset	**Subject**	**# of Images**	**# of Gyro.**	**Collection Environment**	**Bit Depth**
[1]	Test pattern	1400	28000	Indoor	16 bits
[2]	Vehicle, Road	1600	32000	Outdoor	16 bits
[3]	Person, Road	2000	40000	Outdoor	16 bits
[4]	Person, Vehicle	2000	40000	Outdoor	16 bits

The SBTI dataset is generated through Equation (10) based on the STI dataset. In Equation (10), the blur size is determined by α which consists of τ and h. Here, τ is thermal time constant, and h is interval time between two consecutive images (where h is 20 ms, having camera frame rate as 50 Hz). We adjust the blur size by changing the value of h. The real interval time of two sharp images is 20 ms, but we can control the blur size by replacing this interval time with a specific value. For example, assuming h is 1/1280, the frame rate between two sharp images becomes 1280 Hz. In other words, the time consumed to collect 1280 images is no longer 25.6 s but 1 s. The camera rotation speed also is converted from 1.955°/s to 50°/s. This range is about 25.6 times higher than a real camera rotation speed. Using this time compression method, we can generate blurry images corresponding to any camera rotation speed. Finally, the blurry images are sampled every 20 frames and converted to 8-bit images for comparison. Figure 10 and Table 3 show the configurations of STI and SBTI datasets. In the SBTI dataset, there are seven different blur sizes, and the maximum camera rotation speed intuitively expresses the blur size.

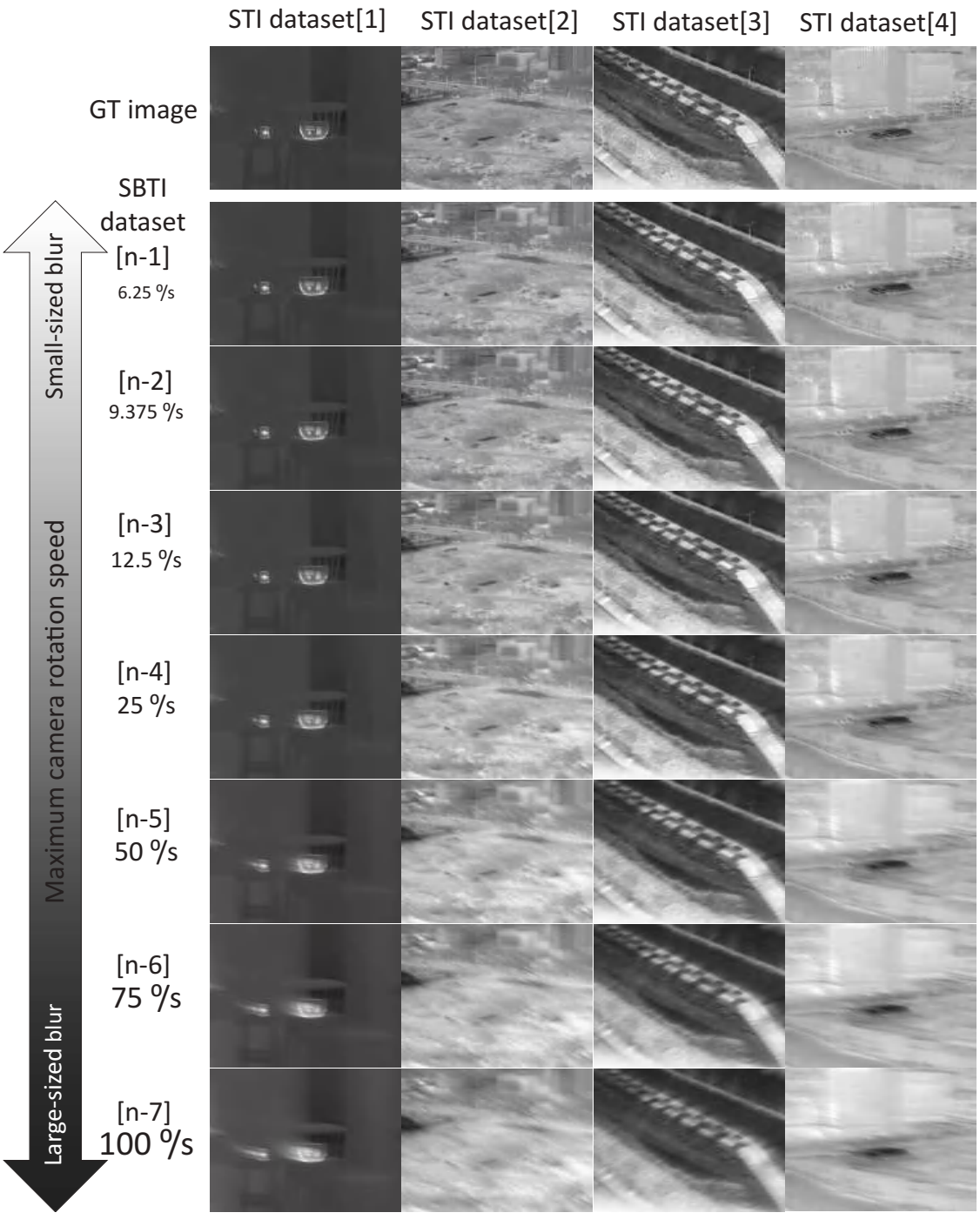


Figure 10. Overview of STI and SBTI datasets.

Table 3. Configuration of SBTI dataset.

STI Dataset	SBTI Dataset						
	Maximum Camera Rotation Speed (°/s)						
	6.25	**9.375**	**12.5**	**25**	**50**	**75**	**100**
[1]	[1-1]	[1-2]	[1-3]	[1-4]	[1-5]	[1-6]	[1-7]
[2]	[2-1]	[2-2]	[2-3]	[2-4]	[2-5]	[2-6]	[2-7]
[3]	[3-1]	[3-2]	[3-3]	[3-4]	[3-5]	[3-6]	[3-7]
[4]	[4-1]	[4-2]	[4-3]	[4-4]	[4-5]	[4-6]	[4-7]

4.2. Construction of Real Blurry Thermal Image Dataset

We collected an additional dataset containing real motion blur for evaluating our method in a real-world environment. The process for acquiring real blurry images is as same as the one for collecting sharp images as presented in Section 4, except that there is no limitation in camera rotation speed for the real effect of a blur. Another difference is that, since we use only one camera, we cannot acquire sharp images at the same time when collecting real blurry images. Specifically, the camera rotation speed varies from 30°/s to 100°/s. In addition, since infrared images are greatly affected by environmental temperature change, we collected daytime and nighttime images, respectively.

4.3. Our Deblurring Procedure

We evaluate the accuracy of our proposed blur kernel rendering result through the deblurring procedure. Therefore, we selected the deconvolution algorithm [65] which can be combined with blur kernel rendering result to construct a non-blind deblurring method. Actually, we used the public code version of [66] implementing [65]. In our experiment, we set parameters as follows: $\lambda = 0.001 \sim 0.003$, $\alpha = 1$.

4.4. Evaluation Environment

Blur kernel rendering and non-blind deblurring are implemented in MATLAB. NVIDIA GeForce GTX 1080 Ti GPU with 11 GB memory and Intel core i7-1065 G7@1.3G HZ with 16 GB memory have been adopted.

5. Experimental Results

Our experimental results are compared to the state-of-the-art deblurring methods, including the single image deblurring methods [33,35,36] and the deep learning-based video deblurring method [34]. We conducted both qualitative and quantitative comparisons on our SBTI dataset. Additionally, we used the real blurry thermal images to qualitatively evaluate the deblurring performance in actual situations.

5.1. Performance Evaluation on SBTI Dataset

The peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) [67] index were leveraged as the metrics of quantitative evaluation. The experimental results are summarized in Tables 4–7 as average values. Relatively higher PSNR and SSIM have been observed from [1-1] to [1-7] compared to the others in the SBTI dataset. As can be observed in the Tables 4–7, PSNR and SSIM tend to gradually decrease when the blur size increases. In most cases, our proposed method produces relatively higher PSNR and SSIM values compared to the state-of-the-art methods.

Table 4. Comparison of quantitative deblurring performance on the SBTI dataset [1-1]–[1-7].

SBTI Dataset	SRN [33]		SIUN [36]		DeblurGAN.v2 [35]		CDVD [34]		Ours	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
[1-1]	40.33	0.9881	41.03	0.9914	41.30	0.9910	39.62	0.9905	**41.57**	**0.9926**
[1-2]	37.96	0.9849	38.45	0.9889	38.37	0.9872	37.09	0.9874	**38.79**	**0.9906**
[1-3]	35.94	0.9815	36.35	0.9858	36.13	0.9835	35.05	0.9840	**36.42**	**0.9880**
[1-4]	30.97	0.9675	**31.11**	0.9714	30.91	0.9695	30.36	0.9699	31.06	**0.9756**
[1-5]	26.69	0.9419	**26.74**	0.9476	26.64	0.9456	26.32	0.9453	26.65	**0.9526**
[1-6]	24.59	0.9221	**24.67**	0.9298	24.57	0.9273	24.34	0.9271	24.52	**0.9337**
[1-7]	23.21	0.9049	**23.33**	0.9141	23.22	0.9118	23.07	0.9130	23.11	**0.9165**
Average	31.38	0.9558	31.67	0.9613	31.59	0.9594	30.84	0.9596	**31.73**	**0.9642**

Table 5. Comparison of quantitative deblurring performance on the SBTI dataset [2-1]–[2-7].

SBTI Dataset	SRN [33]		SIUN [36]		DeblurGAN.v2 [35]		CDVD [34]		Ours	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
[2-1]	28.66	0.8573	29.74	0.9026	32.25	0.9458	28.12	0.8358	**32.98**	**0.9600**
[2-2]	27.06	0.8247	27.97	0.8719	30.06	0.9221	26.54	0.8076	**30.93**	**0.9504**
[2-3]	26.02	0.8048	26.72	0.8455	28.69	0.9014	25.57	0.7891	**29.55**	**0.9396**
[2-4]	23.82	0.7603	24.32	0.7805	25.81	0.8405	24.04	0.7679	**26.38**	**0.9034**
[2-5]	21.78	0.7128	22.54	0.7421	23.36	0.7738	22.74	0.7674	**23.49**	**0.8492**
[2-6]	20.29	0.6743	21.01	0.7063	21.74	0.7262	21.53	0.7450	**21.86**	**0.8104**
[2-7]	19.11	0.6487	19.66	0.6776	20.28	0.6902	20.47	0.7204	**20.61**	**0.7757**
Average	23.82	0.7547	24.56	0.7895	26.03	0.8286	24.14	0.7762	**26.54**	**0.8841**

Table 6. Comparison of quantitative deblurring performance on the SBTI dataset [3-1]–[3-7].

SBTI Dataset	SRN [33]		SIUN [36]		DeblurGAN.v2 [35]		CDVD [34]		Ours	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
[3-1]	29.20	0.8606	29.64	0.8862	35.69	**0.9603**	34.034	0.9240	**36.556**	0.9600
[3-2]	27.93	0.8305	28.66	0.8597	33.79	0.9368	32.43	0.9081	**35.02**	**0.9525**
[3-3]	27.05	0.8053	27.92	0.8394	32.66	0.9201	31.45	0.8965	**33.95**	**0.9452**
[3-4]	25.34	0.7556	26.25	0.7961	30.10	0.8772	29.21	0.8657	**31.10**	**0.9177**
[3-5]	24.29	0.7348	24.90	0.7656	27.27	0.8237	26.72	0.8263	**28.00**	**0.8786**
[3-6]	23.38	0.7196	23.90	0.7435	25.52	0.7882	25.14	0.7982	**25.93**	**0.8427**
[3-7]	22.48	0.7034	22.94	0.7215	24.21	0.7605	23.82	0.7726	**24.53**	**0.8128**
Average	25.67	0.7728	26.32	0.8017	29.89	0.8667	28.97	0.8559	**30.73**	**0.9013**

Table 7. Comparison of quantitative deblurring performance on the SBTI dataset [4-1]–[4-7].

SBTI Dataset	SRN [33]		SIUN [36]		DeblurGAN.v2 [35]		CDVD [34]		Ours	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
[4-1]	30.37	0.8925	31.42	0.9271	33.63	0.9552	32.19	0.9258	**34.05**	**0.9640**
[4-2]	29.02	0.8742	29.78	0.9066	31.78	0.9373	30.77	0.9177	**32.34**	**0.9589**
[4-3]	28.14	0.8620	28.71	0.8900	30.67	0.9262	29.86	0.9110	**31.22**	**0.9532**
[4-4]	25.98	0.8294	26.40	0.8531	27.87	0.8923	27.44	0.8937	**28.20**	**0.9312**
[4-5]	23.88	0.7947	24.22	0.8137	**25.19**	0.8506	24.81	0.8636	25.02	**0.8956**
[4-6]	22.53	0.7731	22.82	0.7869	**23.53**	0.8216	23.22	0.8390	23.41	**0.8704**
[4-7]	21.52	0.7567	21.74	0.7662	**22.33**	0.8022	22.06	0.8175	22.30	**0.8460**
Average	25.92	0.8261	26.44	0.8491	27.86	0.8836	27.19	0.8812	**28.08**	**0.9170**

The qualitative comparing results are shown in Figures 11–14. Figure 11 shows the deblurring results on the 54th frame of the SBTI dataset [1-4]. The main subjects of the SBTI dataset [1-4] consist of a cross pattern and a 4-bar pattern. Unlike the other methods, which partially removed the blur effect, our proposed method dramatically recover the blur effect. The shape of the small spot at the edge of the cross-pattern reveals the signal attenuation characteristics of the blurry thermal image. This signal attenuation effect makes

the small subject disappear in the blurry image. As shown in other algorithm results, it is not easy to restore the blurry image with an extreme loss of signal. In this case, the size of the blur kernel rendered by our proposed method is 20 by 20. Figure 12 shows the deblurring results on the 49th frame of the SBTI dataset [2-5], and the main subject is a group of vehicles. In this blurry image, it is difficult to recognize either the number of vehicles or their shapes. The result of SRN shows that it is almost impossible to recognize a vehicle in the deblurred image. Further, the other methods still fail to restore the shapes of vehicles due to the signal attenuation effect. In this dataset, the signal attenuation effect makes the subject and the background indistinguishable. In contrast, our result shows high restoration performance enough to recognize the number of vehicles and distinguish their external shapes. In this case, the size of the blur kernel rendered by our proposed method is 54 by 54. Figure 13 shows the deblurring results on the 51th frame of the SBTI dataset [3-4]. The main subject is people. Our method most clearly restores the shape of human arms and legs than other competing methods. Further, SRN and CDVD methods show distorted restoration results regarding the tree's shape in the promenade center. In the case, the size of the blur kernel rendered by our proposed method is 24 by 24. Figure 14 shows the deblurring results on the 91th frame of the SBTI dataset. It is very difficult to recognize the number of subjects or their shapes without referring to the ground truth image. Our proposed method successfully restores the blurry image so that the details are sufficiently revealed, such as the number of people and the shapes of vehicles. Most people and vehicles' edges disappeared in this blurry image due to the signal attenuation effect. It is challenging to predict the blur kernel in an image where the subject and the background cannot be distinguished. It is also difficult to show good restoration results without learnable knowledge, even using a deep learning-based approach. In the case, the size of the blur kernel rendered by our proposed method is 107 by 107.

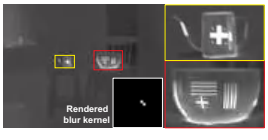
(**a**) Dataset[1-4]52th image

(**b**) PSNR/SSIM: 28.96/0.9703

(**c**) PSNR/SSIM: 28.90/0.9653

(**d**) PSNR/SSIM: **29.01**/0.9688

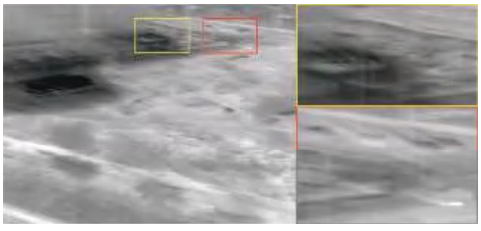
(**e**) PSNR/SSIM: 28.27/0.9662



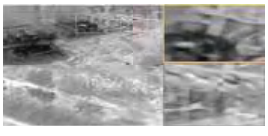
(**f**) PSNR/SSIM: 28.28/**0.9737**

(**g**) Ground truth image

Figure 11. Qualitative comparison of deblurring results on the SBTI dataset [1-4]54th. (**a**) Synthetic blurry thermal image, (**b**) SRN [33], (**c**) SIUN [36], (**d**) DeblurGAN.v2 [35], (**e**) CDVD [34], (**f**) Ours, (**g**) GT.

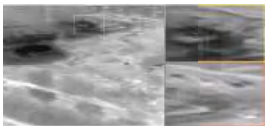


(**a**) Dataset[2-5]49th image

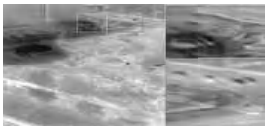


(**b**) PSNR/SSIM: 22.40/0.7791

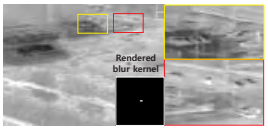
(**c**) PSNR/SSIM: 21.79/0.7401



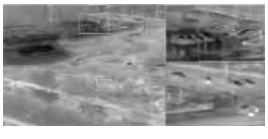
(**d**) PSNR/SSIM: **23.02**/0.8191



(**e**) PSNR/SSIM: 22.98/0.8509



(**f**) PSNR/SSIM: 22.98/**0.9015**



(**g**) Ground truth image

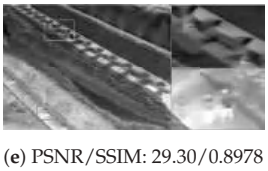
Figure 12. Qualitative comparison of deblurring results on the SBTI dataset [2-5]49th. (**a**) Synthetic blurry thermal image, (**b**) SRN [33], (**c**) SIUN [36], (**d**) DeblurGAN.v2 [35], (**e**) CDVD [34], (**f**) Ours, (**g**) GT.

(**a**) Dataset[3-4]51th image

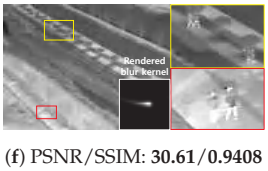
(**b**) PSNR/SSIM: 26.23/0.8058

(**c**) PSNR/SSIM: 24.72/0.7374

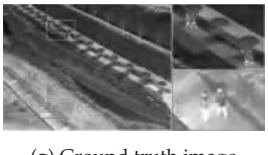
(**d**) PSNR/SSIM: 29.45/0.8884



(**e**) PSNR/SSIM: 29.30/0.8978

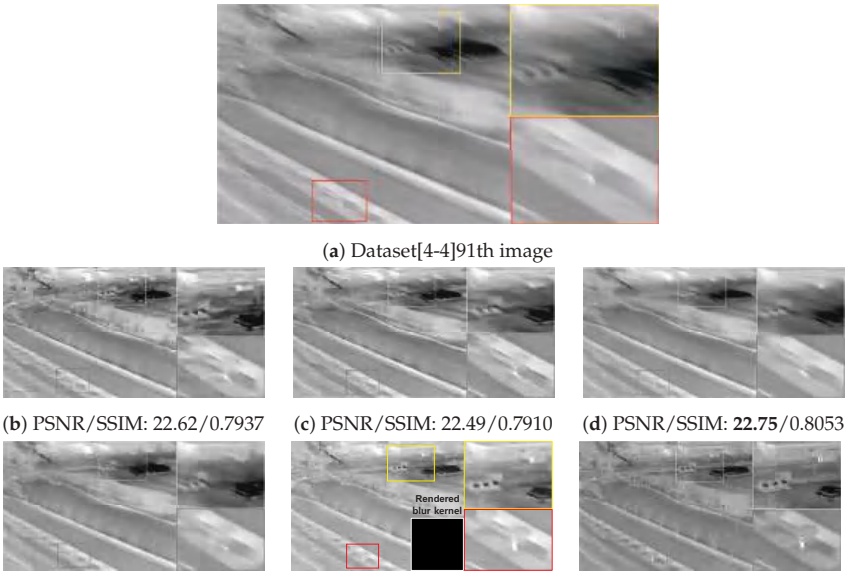


(**f**) PSNR/SSIM: **30.61**/**0.9408**



(**g**) Ground truth image

Figure 13. Qualitative comparison of deblurring results on the SBTI dataset [3-4]51th. (**a**) Synthetic blurry thermal images, (**b**) SRN [33] (**c**) SIUN [36], (**d**) DeblurGAN.v2 [35], (**e**) CDVD [34], (**f**) Ours, (**g**) GT.



(**a**) Dataset[4-4]91th image

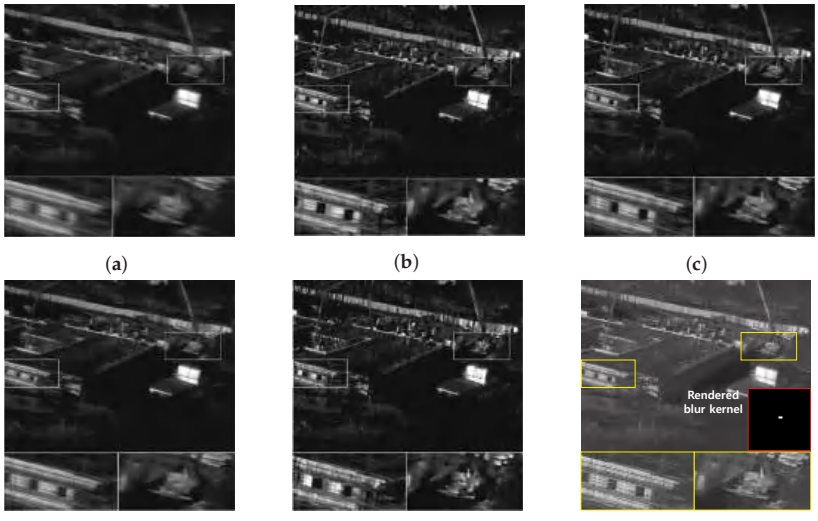
(**b**) PSNR/SSIM: 22.62/0.7937 (**c**) PSNR/SSIM: 22.49/0.7910 (**d**) PSNR/SSIM: **22.75**/0.8053

(**e**) PSNR/SSIM: 22.56/0.8357 (**f**) PSNR/SSIM: 22.60/**0.8723** (**g**) Ground truth image

Figure 14. Qualitative comparison of deblurring results on the SBTI dataset [4-4]91th. (**a**) Synthetic blurry thermal image, (**b**) SRN [33], (**c**) SIUN [36], (**d**) DeblurGAN.v2 [35], (**e**) CDVD [34], (**f**) Ours, (**g**) GT.

5.2. Performance Evaluation on Real Blurry Thermal Images

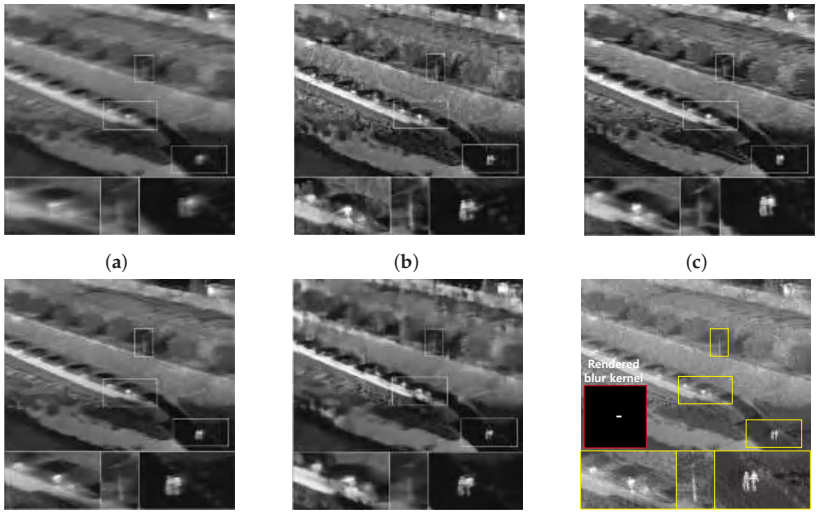
Furthermore, we conduct a qualitative comparison between our proposed method and other methods on real blurry images. Since the real blurry images cannot have the supplementary sharp images as ground truth, only qualitative comparisons are performed. Figures 15 and 16 show the blurry thermal images of building, construction equipment and people, collected when the camera rotation speed has been about 30°/s. Even though the blur effect is low in these images, the competing algorithm results show a residual blur effect in their restoration images. In contrast, our proposed method successfully recovers blurry images, so the shape of the subject is distinguished well. Figures 17 and 18 show the blurry thermal images of vehicles, buildings, and people, collected while the camera rotation speed has been about 40°/s. Because of the effect of a motion blur, we can barely know the shape of the subject in the real blurry images. As can be seen in Figures 17c and 18e, the shape of a person still has the blur effect in the restoration image. On the other hand, our proposed method shows the restoration result that has the fully recognizable shape of the person's arms and legs and contains the details of the vehicle's wheels. Figures 19 and 20 depict the results of images acquired when the camera rotation speed has been about 80°/s. Because of the large level of blur effect, it is impossible to recognize the shape or number of any subject. Although the competing methods reduced the blur effect, their restoration images are not enough to recognize the details of a subject. On the other hand, our proposed method recovers the details of subjects better than the competing methods. In Figure 21, the blurry image was obtained while the camera rotation speed has been about 100°/s. The blur effect had been so huge that the contour or presence of a subject is barely recognizable. However, our method remarkably restores the shape of a person, and all competing methods failed. Figure 22 is the image data collected at night, when the camera rotation speed has been 40°/s. Similar to the above results, our method restores the shape of a person, while the competing methods do not.



(**a**) (**b**) (**c**)

(**d**) (**e**) (**f**)

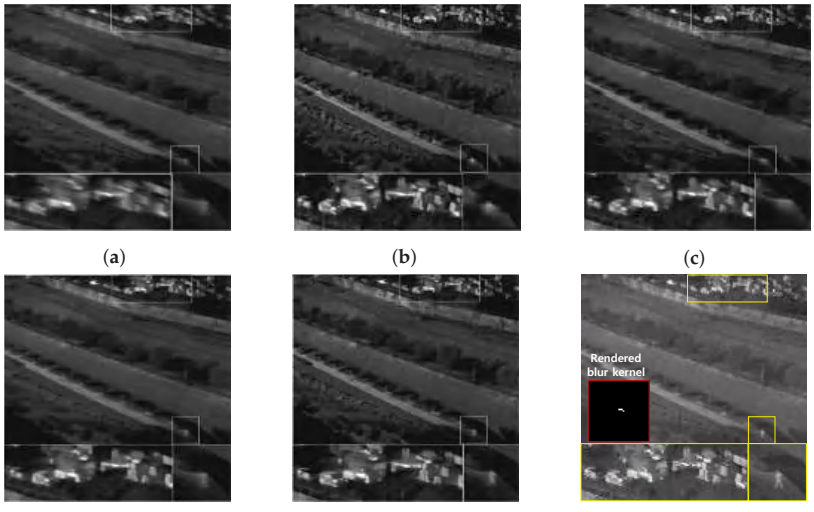
Figure 15. Qualitative comparison of motion deblurring results on the real blurry thermal image. (**a**) Real blurry thermal image acquired with a camera rotating at 31°/*s*, (**b**) SRN [33], (**c**) SIUN [36], (**d**) DeblurGAN.v2 [35], (**e**) CDVD [34], (**f**) Ours.



(**a**) (**b**) (**c**)

(**d**) (**e**) (**f**)

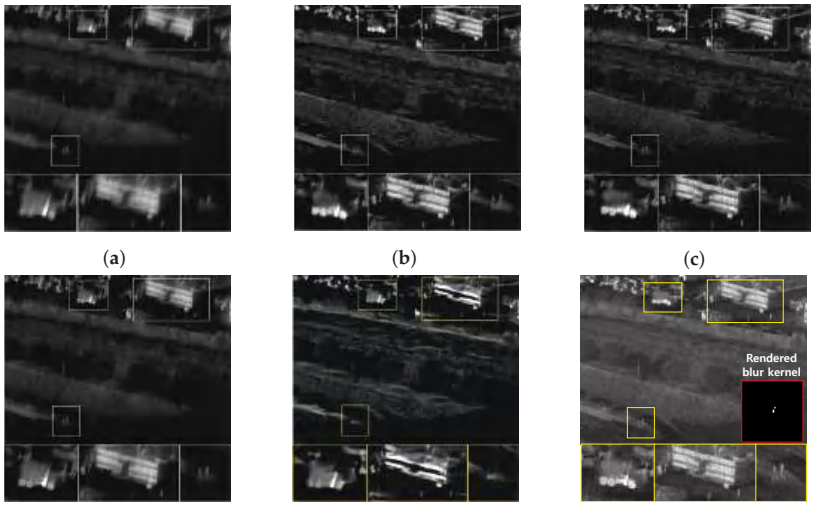
Figure 16. Qualitative comparison of motion deblurring results on the real blurry thermal image. (**a**) Real blurry thermal image acquired with a camera rotating at 39°/*s*, (**b**) SRN [33], (**c**) SIUN [36], (**d**) DeblurGAN.v2 [35], (**e**) CDVD [34], (**f**) Ours.



(**a**) (**b**) (**c**)

(**d**) (**e**) (**f**)

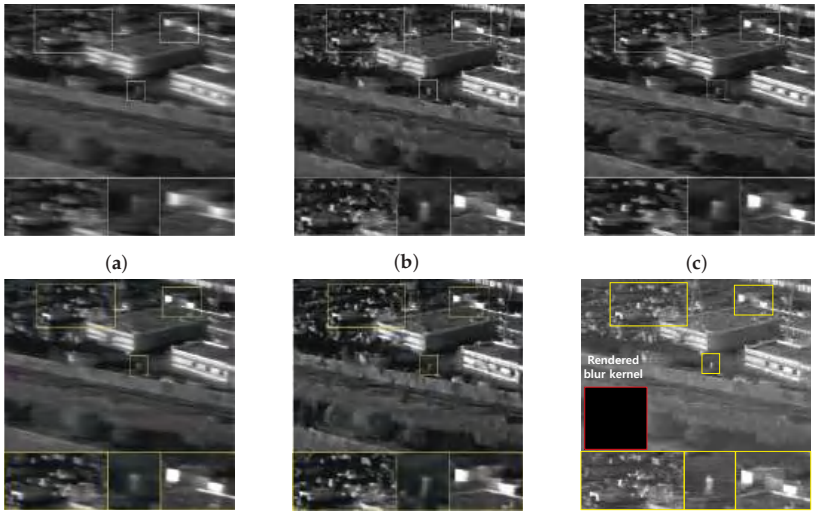
Figure 17. Qualitative comparison of motion deblurring results on the real blurry thermal image. (**a**) Real blurry thermal image acquired with a camera rotating at 43°/s, (**b**) SRN [33], (**c**) SIUN [36], (**d**) DeblurGAN.v2 [35], (**e**) CDVD [34], (**f**) Ours.



(**a**) (**b**) (**c**)

(**d**) (**e**) (**f**)

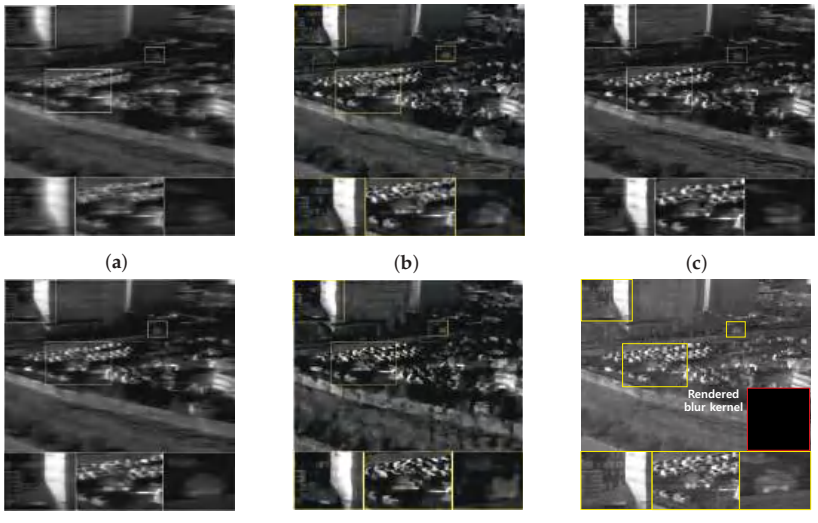
Figure 18. Qualitative comparison of motion deblurring results on the real blurry thermal image. (**a**) Real blurry thermal image acquired with a camera rotating at 44°/s, (**b**) SRN [33], (**c**) SIUN [36], (**d**) DeblurGAN.v2 [35], (**e**) CDVD [34], (**f**) Ours.



(**a**) (**b**) (**c**)

(**d**) (**e**) (**f**)

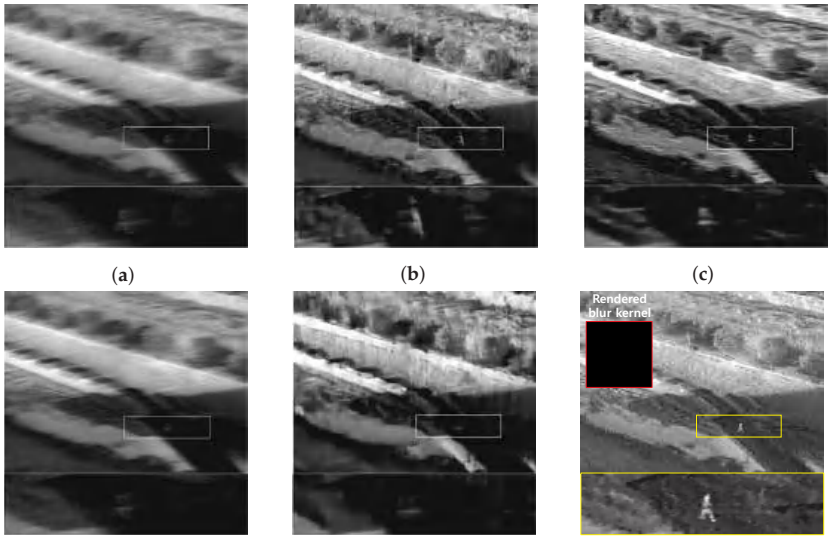
Figure 19. Qualitative comparison of motion deblurring results on the real blurry thermal image. (**a**) Real blurry thermal image acquired with a camera rotating at $84°/s$, (**b**) SRN [33], (**c**) SIUN [36], (**d**) DeblurGAN.v2 [35], (**e**) CDVD [34], (**f**) Ours.



(**a**) (**b**) (**c**)

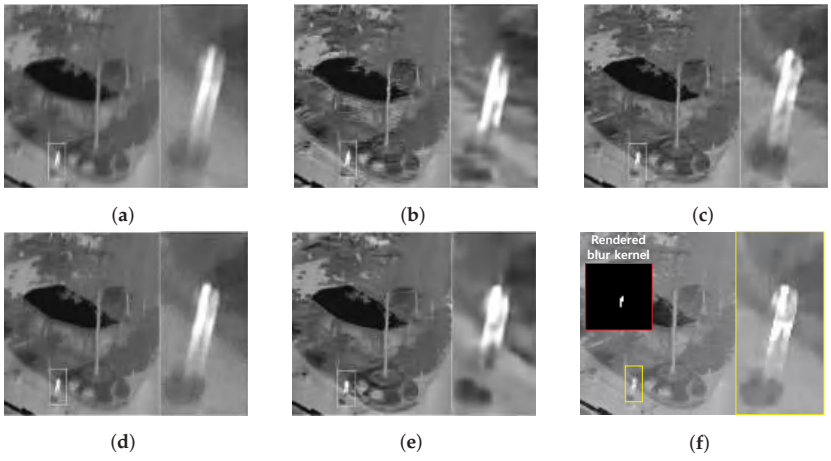
(**d**) (**e**) (**f**)

Figure 20. Qualitative comparison of motion deblurring results on the real blurry thermal image. (**a**) Real blurry thermal image acquired with a camera rotating at $85°/s$, (**b**) SRN [33], (**c**) SIUN [36], (**d**) DeblurGAN.v2 [35], (**e**) CDVD [34], (**f**) Ours.



(**a**) (**b**) (**c**) (**d**) (**e**) (**f**)

Figure 21. Qualitative comparison of motion deblurring results on the real blurry thermal image. (**a**) Real blurry thermal image acquired with a camera rotating at $100^\circ/s$, (**b**) SRN [33], (**c**) SIUN [36], (**d**) DeblurGAN.v2 [35], (**e**) CDVD [34], (**f**) Ours.



(**a**) (**b**) (**c**) (**d**) (**e**) (**f**)

Figure 22. Qualitative comparison of motion deblurring results on the real blurry thermal image. (**a**) Real blurry thermal image acquired with a camera rotating at $40^\circ/s$, (**b**) SRN [33], (**c**) SIUN [36], (**d**) DeblurGAN.v2 [35], (**e**) CDVD [34], (**f**) Ours.

Extensive experimental results show that our proposed method outperforms other methods. The reason is that our approach is able to estimate more accurate blur kernels using a physical model and inertial sensor. There are two explanations regarding how our method can render the exact blur kernel. Firstly, our method leverages the physical mechanism of a thermal detector for accurate blur kernel rendering. As shown in Figure 2, the pixel structure of a thermal detector loses its stored thermal energy over time which appears as the effect of attenuation of an image signal. This attenuation effect causes motion blur similar to a comet tail shape. As shown in Figures 14 and 17–21, when a small-sized subject has its temperature similar to the background, the subject is barely distinguished

from the background due to its attenuation effect of motion blur. It is extremely challenging to obtain a blur kernel from an intensely blurred image where the subject has almost disappeared. Further, even with a deep learning-based method, high performance is hardly achieved without learnable information. In contrast, our method shows high deblurring performance even for vanishing subjects with a large amount of motion blur. For this reason, our proposed method, which is designed considering the characteristics of the thermal detector, is able to show high feasibility compared to other methods in the thermal image domain. Secondly, accurate blur kernel rendering is possible since our proposed method is free from the synchronization problem between the gyroscope data length and the image sensor exposure time. In general, to combine photon detector and gyroscope data, the synchronization problem between photon detector exposure time and gyroscope sensor data length must be resolved. A photon detector adjusts the exposure time in real-time according to the amount of ambient light in a scene. The exposure time range is generally set from a few microseconds to several seconds. Due to the dynamic change in exposure time, the length of gyroscope data also needs to be changed simultaneously. In contrast, in a thermal detector, the concept corresponding to the exposure time of the photon detector is the thermal time constant. Since the thermal time constant is a fixed value determined when a thermal detector is fabricated, the length of gyroscope data used for blur kernel rendering is not changed. Therefore, a thermal detector combined with a gyroscope is more feasible to render the accurate blur kernel.

6. Conclusions

In this paper, we observed that a thermal detector and a photon detector have different inherent characteristics, which accordingly cause different motion blur effects. Based on this observation, we have analyzed the physical and theoretical differences between a thermal detector and a photon detector in order to precisely model a motion blur effect in the thermal image. We suggest a novel motion blur model for thermal images by interpreting the physical mechanism of a thermal detector. The proposed motion blur model is leveraged to enable blur kernel rendering to accurate using gyroscope sensor information. We constructed the first blurry thermal image dataset that contains both synthetic blurred images and sharp thermal images in the thermal image domain. Finally, extensive qualitative and quantitative experiments were conducted to show that our proposed method outperforms the state-of-the-art methods.

Author Contributions: Conceptualization, K.L.; methodology, K.L.; software, K.L. and Y.B.; validation, K.L. and Y.B.; data collection, K.L. and Y.B.; writing—original draft preparation, K.L. and Y.B.; writing—review and editing, C.K.; visualization, K.L. and Y.B.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external or third party funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Anyone who wants to use the dataset presented in this paper can receive the dataset by filling out a simple request form at the following link. Link: https://forms.gle/ZRK1R1imETkzCWkh8 (accessed on 20 January 2022).

Acknowledgments: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Huda, A.N.; Taib, S. Application of infrared thermography for predictive/preventive maintenance of thermal defect in electrical equipment. *Appl. Therm. Eng.* **2013**, *61*, 220–227. [CrossRef]
2. Mayer, S.; Lischke, L.; Woźniak, P.W. Drones for search and rescue. In Proceedings of the 1st International Workshop on Human-Drone Interaction, Glasgow, UK, 4–9 May 2019.

3. Apvrille, L.; Tanzi, T.; Dugelay, J.L. Autonomous drones for assisting rescue services within the context of natural disasters. In Proceedings of the 2014 XXXIth URSI General Assembly and Scientific Symposium (URSI GASS), Beijing, China, 16–23 August 2014; pp. 1–4.
4. Pinchon, N.; Cassignol, O.; Nicolas, A.; Bernardin, F.; Leduc, P.; Tarel, J.P.; Brémond, R.; Bercier, E.; Brunet, J. All-weather vision for automotive safety: Which spectral band? In *International Forum on Advanced Microsystems for Automotive Applications*; Springer: Berlin, Germany, 2018; pp. 3–15.
5. Wikipedia. Infrared—Wikipedia, The Free Encyclopedia. 2021. Available online: http://en.wikipedia.org/w/index.php?title=Infrared&oldid=1052704429 (accessed on 3 November 2021).
6. Kimata, M. Uncooled infrared focal plane arrays. *IEEJ Trans. Electr. Electron. Eng.* **2018**, *13*, 4–12. [CrossRef]
7. Buades, A.; Coll, B.; Morel, J.M. A non-local algorithm for image denoising. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), San Diego, CA, USA, 20–26 June 2005; Volume 2, pp. 60–65.
8. Zha, Z.; Wen, B.; Yuan, X.; Zhou, J.; Zhu, C. Image Restoration via Reconciliation of Group Sparsity and Low-Rank Models. *IEEE Trans. Image Process.* **2021**, *30*, 5223–5238. [CrossRef]
9. Buades, A.; Coll, B.; Morel, J.M. A review of image denoising algorithms, with a new one. *Multiscale Model. Simul.* **2005**, *4*, 490–530. [CrossRef]
10. Zha, Z.; Yuan, X.; Wen, B.; Zhou, J.; Zhang, J.; Zhu, C. From Rank Estimation to Rank Approximation: Rank Residual Constraint for Image Restoration. *IEEE Trans. Image Process.* **2020**, *29*, 3254–3269. [CrossRef]
11. Stark, J.A. Adaptive image contrast enhancement using generalizations of histogram equalization. *IEEE Trans. Image Process.* **2000**, *9*, 889–896. [CrossRef]
12. Jung, C.; Jiao, L.; Qi, H.; Sun, T. Image deblocking via sparse representation. *Signal Process. Image Commun.* **2012**, *27*, 663–677. [CrossRef]
13. Zha, Z.; Yuan, X.; Wen, B.; Zhang, J.; Zhou, J.; Zhu, C. Image Restoration Using Joint Patch-Group-Based Sparse Representation. *IEEE Trans. Image Process.* **2020**, *29*, 7735–7750. [CrossRef]
14. Bertalmio, M.; Sapiro, G.; Caselles, V.; Ballester, C. Image inpainting. In Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, New Orleans, LA, USA, 23–28 July 2000; pp. 417–424.
15. Zha, Z.; Yuan, X.; Wen, B.; Zhou, J.; Zhang, J.; Zhu, C. A Benchmark for Sparse Coding: When Group Sparsity Meets Rank Minimization. *IEEE Trans. Image Process.* **2020**, *29*, 5094–5109. [CrossRef]
16. Pan, J.; Sun, D.; Pfister, H.; Yang, M.H. Blind image deblurring using dark channel prior. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1628–1636.
17. Yan, Y.; Ren, W.; Guo, Y.; Wang, R.; Cao, X. Image deblurring via extreme channels prior. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 17–25 July 2017; pp. 4003–4011.
18. Zha, Z.; Wen, B.; Yuan, X.; Zhou, J.T.; Zhou, J.; Zhu, C. Triply Complementary Priors for Image Restoration. *IEEE Trans. Image Process.* **2021**, *30*, 5819–5834. [CrossRef]
19. Zha, Z.; Yuan, X.; Zhou, J.; Zhu, C.; Wen, B. Image Restoration via Simultaneous Nonlocal Self-Similarity Priors. *IEEE Trans. Image Process.* **2020**, *29*, 8561–8576. [CrossRef]
20. Zha, Z.; Yuan, X.; Wen, B.; Zhou, J.; Zhu, C. Group Sparsity Residual Constraint With Non-Local Priors for Image Restoration. *IEEE Trans. Image Process.* **2020**, *29*, 8960–8975. [CrossRef]
21. Zhang, J.; Ghanem, B. ISTA-Net: Interpretable optimization-inspired deep network for image compressive sensing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 1828–1837.
22. Han, J.; Lee, H.; Kang, M.G. Thermal Image Restoration Based on LWIR Sensor Statistics. *Sensors* **2021**, *21*, 5443. [CrossRef]
23. Morris, N.J.W.; Avidan, S.; Matusik, W.; Pfister, H. Statistics of Infrared Images. In Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition, Minneapolis, MN, USA, 18–23 June 2007; pp. 1–7. [CrossRef]
24. Huang, Y.; Bi, D.; Wu, D. Infrared and visible image fusion based on different constraints in the non-subsampled shearlet transform domain. *Sensors* **2018**, *18*, 1169. [CrossRef]
25. Ban, Y.; Lee, K. Multi-Scale Ensemble Learning for Thermal Image Enhancement. *Appl. Sci.* **2021**, *11*, 2810. [CrossRef]
26. Choi, Y.; Kim, N.; Hwang, S.; Kweon, I.S. Thermal image enhancement using convolutional neural network. In Proceedings of the 2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, Korea, 9–14 October 2016; pp. 223–230.
27. Lee, K.; Lee, J.; Lee, J.; Hwang, S.; Lee, S. Brightness-based convolutional neural network for thermal image enhancement. *IEEE Access* **2017**, *5*, 26867–26879. [CrossRef]
28. Oswald-Tranta, B. Temperature reconstruction of infrared images with motion deblurring. *J. Sens. Sens. Syst.* **2018**, *7*, 13–20. [CrossRef]
29. Nihei, R.; Tanaka, Y.; Iizuka, H.; Matsumiya, T. Simple correction model for blurred images of uncooled bolometer type infrared cameras. In Proceedings of the Infrared Imaging Systems: Design, Analysis, Modeling, and Testing XXX, International Society for Optics and Photonics, Baltimore, MD, USA, 14–18 April 2019; Volume 11001, p. 420–427.
30. Ramanagopal, M.S.; Zhang, Z.; Vasudevan, R.; Roberson, M.J. Pixel-Wise Motion Deblurring of Thermal Videos. In Proceedings of the Robotics: Science and Systems XVI, Cambridge, MA, USA, 12–16 July 2020; Volume 16.
31. Zhao, Y.; Fu, G.; Wang, H.; Zhang, S.; Yue, M. Infrared Image Deblurring Based on Generative Adversarial Networks. *Int. J. Opt.* **2021**, *2021*, 9946809. [CrossRef]

32. Batchuluun, G.; Lee, Y.W.; Nguyen, D.T.; Pham, T.D.; Park, K.R. Thermal image reconstruction using deep learning. *IEEE Access* **2020**, *8*, 126839–126858. [CrossRef]
33. Tao, X.; Gao, H.; Shen, X.; Wang, J.; Jia, J. Scale-recurrent network for deep image deblurring. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8174–8182.
34. Pan, J.; Bai, H.; Tang, J. Cascaded deep video deblurring using temporal sharpness prior. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–16 June 2020; pp. 3043–3051.
35. Kupyn, O.; Martyniuk, T.; Wu, J.; Wang, Z. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27–28 October 2019; pp. 8878–8887.
36. Ye, M.; Lyu, D.; Chen, G. Scale-iterative upscaling network for image deblurring. *IEEE Access* **2020**, *8*, 18316–18325. [CrossRef]
37. Wang, S.; Zhang, S.; Ning, M.; Zhou, B. Motion Blurred Star Image Restoration Based on MEMS Gyroscope Aid and Blur Kernel Correction. *Sensors* **2018**, *18*, 2662. [CrossRef]
38. Liu, D.; Chen, X.; Liu, X.; Shi, C. Star Image Prediction and Restoration under Dynamic Conditions. *Sensors* **2019**, *19*, 1890. [CrossRef]
39. Audi, A.; Pierrot-Deseilligny, M.; Meynard, C.; Thom, C. Implementation of an IMU Aided Image Stacking Algorithm in a Digital Camera for Unmanned Aerial Vehicles. *Sensors* **2017**, *17*, 1646. [CrossRef]
40. Bae, H.; Fowlkes, C.C.; Chou, P.H. Accurate motion deblurring using camera motion tracking and scene depth. In Proceedings of the 2013 IEEE Workshop on Applications of Computer Vision (WACV), Beach, FL, USA, 15–17 January 2013; pp. 148–153.
41. Zhang, Y.; Hirakawa, K. Combining inertial measurements with blind image deblurring using distance transform. *IEEE Trans. Comput. Imaging* **2016**, *2*, 281–293. [CrossRef]
42. Hu, Z.; Yuan, L.; Lin, S.; Yang, M.H. Image deblurring using smartphone inertial sensors. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1855–1864.
43. Hee Park, S.; Levoy, M. Gyro-based multi-image deconvolution for removing handshake blur. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 3366–3373.
44. Mustaniemi, J.; Kannala, J.; Särkkä, S.; Matas, J.; Heikkilä, J. Inertial-aided motion deblurring with deep networks. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 8–10 January 2019, pp. 1914-1922. [CrossRef]
45. Joshi, N.; Kang, S.B.; Zitnick, C.L.; Szeliski, R. Image deblurring using inertial measurement sensors. *ACM Trans. Graph. (TOG)* **2010**, *29*, 1–9.
46. Ji, S.; Hong, J. -P.; Lee, J.; Baek, S. -J.; Ko, S.-J. Robust Single Image Deblurring Using Gyroscope Sensor. *IEEE Access* **2021**, *9*, 80835–80846. [CrossRef]
47. Sindelar, O.; Sroubek, F. Image deblurring in smartphone devices using built-in inertial measurement sensors. *J. Electron. Imaging* **2013**, *22*, 011003. [CrossRef]
48. Nah, S.; Hyun Kim, T.; Mu Lee, K. Deep multi-scale convolutional neural network for dynamic scene deblurring. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3883–3891.
49. Zhang, K.; Luo, W.; Zhong, Y.; Ma, L.; Stenger, B.; Liu, W.; Li, H. Deblurring by realistic blurring. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 2737–2746.
50. Navarro, F.; Serón, F.J.; Gutierrez, D. Motion blur rendering: State of the art. In *Computer Graphics Forum*; Wiley: Hoboken, NJ, USA, 2011; Volume 30, pp. 3–26.
51. Lancelle, M.; Dogan, P.; Gross, M. Controlling motion blur in synthetic long time exposures. In *Computer Graphics Forum*; Wiley Online Library, Wiley: Hoboken, NJ, USA, 2019; Volume 38, pp. 393–403.
52. Kruse, P.W. Chapter 2 Principles of Uncooled Infrared Focal Plane Arrays. In *Uncooled Infrared Imaging Arrays and Systems*; Kruse, P.W.; Skatrud, D.D., Eds.; Elsevier: Amsterdam, The Netherlands, 1997; Volume 47, pp. 17–42. [CrossRef]
53. Oh, J.; Song, H.s.; Park, J.; Lee, J.K. Noise Improvement of a-Si Microbolometers by the Post-Metal Annealing Process. *Sensors* **2021**, *21*, 6722. [CrossRef]
54. Numerical Differential Equation Methods. In *Numerical Methods for Ordinary Differential Equations*; John Wiley & Sons, Ltd.: Hoboken, NJ, USA, 2008; Chapter 2, pp. 51–135. [CrossRef]
55. Pradham, P.; Younan, N.H.; King, R.L. 16—Concepts of image fusion in remote sensing applications. In *Image Fusion*; Stathaki, T., Ed.; Academic Press: Oxford, UK, 2008; pp. 393–428. [CrossRef]
56. Hartley, R.; Zisserman, A. Scene planes and homographies. In *Multiple View Geometry in Computer Vision*; Cambridge University Press: Cambridge, UK, 2004; pp. 325–343. [CrossRef]
57. Köhler, R.; Hirsch, M.; Mohler, B.; Schölkopf, B.; Harmeling, S. Recording and playback of camera shake: Benchmarking blind deconvolution with a real-world database. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 27–40.
58. Whyte, O.; Sivic, J.; Zisserman, A.; Ponce, J. Non-uniform deblurring for shaken images. *Int. J. Comput. Vis.* **2012**, *98*, 168–186. [CrossRef]
59. Bell, S.; Troccoli, A.; Pulli, K. A non-linear filter for gyroscope-based video stabilization. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 294–308.
60. Hu, Z.; Cho, S.; Wang, J.; Yang, M.H. Deblurring low-light images with light streaks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 3382–3389.

61. Bouguet, J.Y. Camera Calibration Toolbox for Matlab. 2004. Available online: http://www.vision.caltech.edu/bouguetj/calib_doc/index.html (accessed on 4 November 2021)
62. Kino, G.S.; Corle, T.R. *Confocal Scanning Optical Microscopy and Related Imaging Systems*; Academic Press: Cambridge, MA, USA, 1996.
63. Zhang, B.; Zerubia, J.; Olivo-Marin, J.C. Gaussian approximations of fluorescence microscope point-spread function models. *Appl. Opt.* **2007**, *46*, 1819–1829. [CrossRef]
64. Guenther, B.D.; Steel, D. *Encyclopedia of Modern Optics*; Academic Press: Cambridge, MA, USA, 2018.
65. Krishnan, D.; Fergus, R. Fast image deconvolution using hyper-Laplacian priors. *Adv. Neural Inf. Process. Syst.* **2009**, *22*, 1033–1041.
66. Pan, J.; Hu, Z.; Su, Z.; Yang, M.H. Deblurring text images via L0-regularized intensity and gradient prior. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 2901–2908.
67. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef]

sensors

MDPI

Article

Hyperspectral Image Labeling and Classification Using an Ensemble Semi-Supervised Machine Learning Approach

Vidya Manian [1,*], Estefanía Alfaro-Mejía [1] and Roger P. Tokars [2]

[1] Department of Electrical and Computer Engineering, University of Puerto Rico, Mayaguez, PR 00681, USA; estefania.alfaro@upr.edu
[2] NASA Glenn Research Center, 21000 Brookpark Rd, Cleveland, OH 44135, USA; roger.p.tokars@nasa.gov
* Correspondence: vidya.manian@upr.edu

Abstract: Hyperspectral remote sensing has tremendous potential for monitoring land cover and water bodies from the rich spatial and spectral information contained in the images. It is a time and resource consuming task to obtain groundtruth data for these images by field sampling. A semi-supervised method for labeling and classification of hyperspectral images is presented. The unsupervised stage consists of image enhancement by feature extraction, followed by clustering for labeling and generating the groundtruth image. The supervised stage for classification consists of a preprocessing stage involving normalization, computation of principal components, and feature extraction. An ensemble of machine learning models takes the extracted features and groundtruth data from the unsupervised stage as input and a decision block then combines the output of the machines to label the image based on majority voting. The ensemble of machine learning methods includes support vector machines, gradient boosting, Gaussian classifier, and linear perceptron. Overall, the gradient boosting method gives the best performance for supervised classification of hyperspectral images. The presented ensemble method is useful for generating labeled data for hyperspectral images that do not have groundtruth information. It gives an overall accuracy of 93.74% for the Jasper hyperspectral image, 100% accuracy for the HSI2 Lake Erie images, and 99.92% for the classification of cyanobacteria or harmful algal blooms and surface scum. The method distinguishes well between blue green algae and surface scum. The full pipeline ensemble method for classifying Lake Erie images in a cloud server runs 24 times faster than a workstation.

Keywords: hyperspectral images; semi-supervised learning; groundtruth; labeling; feature extraction; principal components analysis; normalization; image classification and reconstruction

Citation: Manian, V.; Alfaro-Mejía, E.; Tokars, R.P. Hyperspectral Image Labeling and Classification Using an Ensemble Semi-Supervised Machine Learning Approach. *Sensors* **2022**, *22*, 1623. https://doi.org/10.3390/s22041623

Academic Editor: Manuel José Cabral dos Santos Reis

Received: 15 December 2021
Accepted: 15 February 2022
Published: 18 February 2022

1. Introduction

Hyperspectral imaging (HSI) provides a high density of spectral information in the hundreds of bands of the imaged material. Most modern hyperspectral sensors also have a high spatial resolution enabling the images to have a range of applications in agriculture, ecosystem monitoring, astronomy, molecular biology, biomedical imaging, geosciences, physics, and surveillance. Hyperspectral unmixing is the method of identifying the percentage of material or endmember contributions in each pixel, hence useful for material identification or detection. There are linear and nonlinear methods for hyperspectral unmixing [1]. They can be used to gain preliminary knowledge on the site before embarking on a field campaign. These images are particularly useful for informed decision-making on a terrestrial or aquatic ecosystem.

Hyperspectral image classification requires preprocessing methods to reduce dimensionality and requires algorithms to solve the issues of few labeled samples, and low spatial resolution [2]. Traditionally, hyperspectral images have been classified using supervised, semi-supervised, and unsupervised Machine Learning (ML) methods. HSI classification is usually done after applying dimensionality reduction, feature extraction, and/or band

subset selection. A review of the ranking, clustering, searching, sparsity, embedding, and hybrid scheme-based methods for band selection are given in [3]. A review of non-negative matrix factorization techniques and benchmark datasets for unmixing are presented in [4]. Low spatial rank tensor factorization methods are popular for unmixing hyperspectral images with mixed pixels [5]. ML approaches such as Random Forest (RF) and XGBoost have been applied in precision agriculture for the estimation of biomass [6]. HSI are also used for change detection in the ocean. A spatial–spectral attention network with PCA-based features is used for change detection [7].

The challenging problems with HSI classification are the unlabeled pixels and the high dimensionality of hyperspectral images. It is also expensive to assign labels to the pixels from field sampling requiring human supervision. To address the problem of unlabeled samples, ML algorithms have been developed which are described below. A graph-based semisupervised learning or ensemble label propagation method using spectral–spatial similarity measurements from a graph representation is proposed in [8]. Recently, Deep Learning (DL) methods are being developed and used for HSI classification [9]. Autoencoders have been used for hyperspectral unmixing and extended to the classification of HSI [10–12]. DL networks require a large number of labeled samples, which is overcome by few shot learning from spectral–spatial features, and training and testing using a 3D CNN in a metric space [13]. One of the disadvantages of using DL techniques is the computational complexity and cost. ML techniques are promising, but require pre-processing and feature extraction stages before training and validation. A local and global modeling approach for pseudo labeling using Active Learning (AL) is proposed in [14] for HSI classification. Tree-based approaches have gained attention in semi-supervised HSI classification. An ensemble semi-supervised random forest method is used for adaptively labeling unlabeled data and adding them to the training dataset [15]. AL and semi-supervised learning are combined to improve the performance of random forest method for HSI classification in [16]. The current ensemble classifiers and semi-supervised methods do not consider all the samples without labeling. The novelty of our ensemble semi-supervised scheme takes into account all the unlabeled samples in the HSI. Moreover, considering the computational complexity of DL networks, we propose a scheme for improving the performance of ML approach for HSI classification by image preprocessing using spectral textural and statistical feature extraction for image enhancement and semi-supervised ensemble labeling and classification in the following way:

- Unlabeled samples are labeled without a pre-trained labeled model by extracting spectral textural and statistical features and incorporating them in the image enhancement stage.
- The textural energy and statistical features computed in the image enhancement stage are input to a k-means clustering stage.
- The novel workflow consists of assigning labels to the unlabeled samples using spectral textural and statistical information in the unsupervised stage, followed by the application of an ensemble of four ML classifiers in the supervised stage, and a decision block that selects the best classifier for the classification of the image.

We apply our ensemble semi-supervised ML scheme for labeling and classification of hyperspectral images acquired over water bodies with Harmful Algal Blooms (HABs). HABs occur in fresh, marine (salt), and brackish (mixture of salt and fresh) water bodies around the world. They are caused by noxious and toxic phytoplankton, cyanobacteria, benthic algae, and microalgae. They are also produced by the overabundance of nutrients such as nitrates, ammonia, urea, and phosphates in the water. These nutrients runoff into the water from agriculture, fertilizers, and urban activity. The HABs lower oxygen levels in the water causing harm to organisms, animals, the environment, and the economy. The bloom lifespan lasts as long as there are favorable conditions but typically ranges from a few days to many months. HABs have been increasing in size and frequency worldwide, and it is caused by possible global climate change. Hence, HAB monitoring is key to the management of the health and utility of waterbodies. NOAA has used

hyperspectral sensors to detect HABs in Lake Erie, one of the Great Lakes that border the U.S. and Canada. The hyperspectral camera collects information on the location, size, the concentration of the blooms, and types of algae [17]. The NASA Glenn Research Center (GRC) has developed an in-house hyperspectral camera, the airborne HSI2 that operates in the wavelength of 400 to 900 nm useful for HAB identification [18]. It can collect data at a high spatial resolution of 1 m, with the advantage of on-demand airborne flight paths not affected by cloud cover [19]. The HSI2 camera images have been used for assessing spatial and temporal variability of blue-green algae, chlorophyll, and temperature [20]. The airborne imagery serves as a complement to satellite-based measurements. HAB detection has been done using varimax-rotated principal components to isolate noise, extracting spectral components, and spatial patterns [21]. Satellite imagery from Sentinel-2A has been used for retrieval of chlorophyll-a concentration using empirical algorithms applied to the image bands, and an ensemble method. An ensemble is a set of base estimators that can be combined to make new predictions [22]. Moreover, Sentinel-2A images have been used for the estimation of chlorophyll-a concentration from regionally and locally adaptive models. Several empirical models were evaluated and found that the single global model constructed by the top-performing empirical algorithm performed best in estimating Chlorophyll-a concentration from both the multispectral and hyperspectral airborne images [23].

In this paper, we present a semi-supervised approach for labeling and classification of HSI that combines the best classifiers to provide optimal classification results. The rest of the paper is organized as follows. Section 2 presents an overview of the methodology and the algorithms used for preprocessing, feature extraction, clustering, and classification. It presents the ensemble ML models: (1) for labeling HSI in the absence of groundtruth data, which requires a preliminary clustering procedure, and (2) labeling and classification of HSI with groundtruth data. Section 3 presents results, while Section 4 discusses the results and compares them with those of state-of-the-art methods. The limitations and future work are presented here. The conclusions are provided in Section 5.

2. Materials and Methods

This section describes the images, and the methods used for preprocessing, labeling, and classification of the hyperspectral images. Two types of HSI are used: ones without groundtruth data and another with groundtruth data. The ones without groundtruth data are from airborne HSI sensors flown by NASA Glenn Research Center (GRC).

The processing of hyperspectral images involves calibration of the images in the laboratory and georeferencing of the data in flight. The calibration in the laboratory utilizes a known National Institute of Standards and Technology (NIST) calibrated radiance source to convert image intensity counts to radiance units. The calibration also utilizes a HgAr light source to convert the spatial pixel axis into known wavelength units. Additionally, in-flight O_2 absorption lines fine-tune these wavelength calibrations to negate the effects of temperature and pressure differences. In-flight measurements of latitude, longitude, and attitude allow for georeferencing of the images. Figure 1 shows the HSI2 sensor installed on the NASA Twin Otter aircraft.

We have used two HSI2 camera images from the Ohio Supercomputer Center (OSC) and one image developed by GRC for HAB monitoring in near real-time in Lake Erie [19]. The two HSI2 images are one-meter resolution with 51 bands from 400 to 900 nanometers of size 5000 rows and 495 columns, and the CyanoHAB hyperspectral image has 170 bands from 400 to 900 nm with a spectral resolution of 2.5 nm, also with a spatial resolution of 1 m. Figure 2 shows the block diagram for the proposed semi-supervised classification scheme.

Figure 1. HSI2 installed on the NASA Twin Otter aircraft.

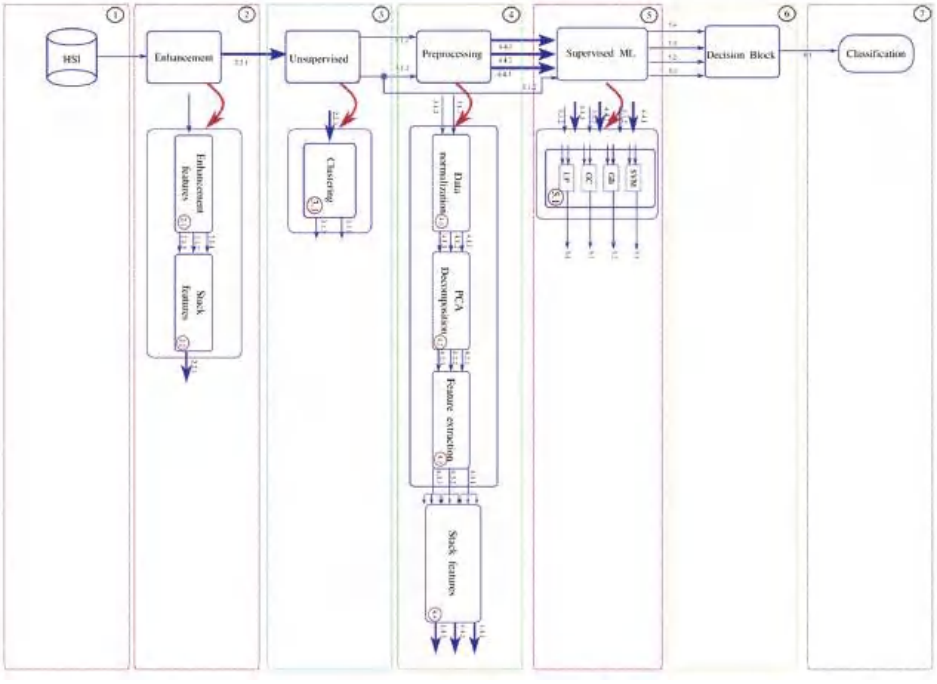


Figure 2. Block diagram for the semi-supervised hyperspectral image classification framework. The numbers on the arrows from top to bottom indicate the 3 types of scaling. The blue arrows represent the input and output for each stage, and the stages are represented by the blocks. The red arrow is a zoom into the corresponding stage showing what happens in that stage in detail.

The proposed semi-supervised HSI classification workflow is illustrated in Figure 2. The workflow has seven stages. The HSI2 images have NaN entries for some pixel data points. The sunlight reflects off the water causing imager saturation by glare or speckle. Hence, NaN is inserted at these data locations. The first stage corresponds to the input and is the hyperspectral image, the image is read and is processed using a data frame structure and the NaN values are replaced by the mean of the five neighborhood pixels. After this filtering process, the enhancement stage has two sub-processes 2.1 and 2.2 (shown in Figure 2). In Section 2.1 relevant features are extracted using the stacked 51 bands of

the image using the first-two statistical moments (mean, standard deviation) and texture information as an energy feature. These features are described below:

$$\xi = \sum_{n=1}^{N_1 N_2} |x[n]|^2 \tag{1}$$

is the energy. N_1 and N_2 are the batch size [24]. The mean is computed as:

$$\mu = \frac{1}{N_1 N_2} \sum_{j=1}^{N_2} \sum_{i=1}^{N_1} x[i][j] \tag{2}$$

and the standard deviation feature is computed as:

$$\sigma = \sqrt{\frac{1}{N_1 N_2} \sum_{j=1}^{N_2} \sum_{i=1}^{N_1} (x[i][j] - \mu)^2} \tag{3}$$

In the 2.2 sub-process, the enhancement vectors are stacked. The 3 arrows indicate the 3 features which are then stacked into one data frame. In the 2.2 sub-process, the stacked vectors are then input to the unsupervised stage. In the 3.1 sub-process, the label assignment is made. The data preparation for this stage requires a 1-D tensor representation of the image. The experiment consists of various trials of cluster numbers, k = 2 to 5, to result in an output image label representation from the original image after the enhancement stage. Once the best label assignment for the Lake Erie image is determined, we have the data and the corresponding labels. Since the images do not have specific groundtruth data, the unsupervised stage produces a label representation of the original image for the best number of clusters. The next is stage 4 processing which includes 4 sub-processes. Sub-process 4.1 is a data normalization process using three different kinds of normalization: normalization scaling (ns), maximum scaling (ms), and scaling (sc). After the data normalization process, sub-process 4.2 is PCA decomposition and selection of 3, 5, or 7 bands. In sub-process 4.3, the feature vectors **ft** from the enhancement stage are computed. In sub-process 4.4, the resulting vectors are stacked into an array **Y** and are concatenated with the labels provided by the unsupervised stage. Stacked vectors and the labels are then input to the process 5, supervised Machine Learning (ML) stage.

Stacked vectors **Y** and the labels go through a batch selection process before being input to the supervised ML stage. The Supervised ML stage has four machine learning techniques in an ensemble configuration: Support Vector Machines (SVM), Gradient Boost Classifier (GB), Gaussian Classifier, and a Linear Perceptron (LP) [25]. SVMs represent the training samples as points in p-dimensional space, mapped so that the samples of the data classes are separated by a (p-1) dimensional hyperplane. The hyperplane is chosen such that it maximizes the margin on either side of the hyperplane between two classes. Hence, SVM performs binary classification but can be extended to multi-class problems. Gradient boost classifiers combine many weak learning models to create a strong predictive model. It minimizes a loss function by iteratively choosing a function that points towards the negative gradient. A Gaussian classifier is a naïve Bayes classifier. It is a generative approach that models the class posterior and input-class conditional distribution. The LP is a linear feedforward network with an input and an output layer.

The stage 6 process is the decision block that decides the best classifier based on the classification accuracy results obtained from testing the trained models. The final classification stage 7 receives the decision block results and labels the HSI pixels to fixed class labels. The classification stage results are also evaluated using three metrics. They are the classification accuracy, F1-score, and the Structural Similarity Index Metric (SSIM). The SSIM compares the reconstructed image with the labeled image and rates how good the

reconstructed image from the classification is compared to groundtruth labeled image. The SSIM is given by:

$$\mathrm{SSIM}(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{\left(\mu_x^2 + \mu_y^2 + C_1\right)\left(\sigma_x^2 + \sigma_y^2 + C_2\right)} \tag{4}$$

where x and y are two non-negative image signals, μ_x and μ_y are their means, and σ_x and σ_y are their standard deviations, σ_{xy} the correlation, and C_1 and C_2 are constants [26,27]. The SSIM is widely used for the assessment of image quality and it satisfies the conditions of symmetry, boundedness, and unique maximum.

We used Amazon Web Services (AWS) [28] to run the training models written in Python for classification of the two HSI2 Lake Erie images. AWS hardware resource is an EC2 instance of type R5 extra large which has six Virtual CPUs (VCPUs), and 32 GB RAM. This particular instance provides optimized memory computing.

2.1. Workflow for Supervised Classification of Jasper Image

We have used the Jasper HSI, because it is similar to the Lake Erie image as it has land cover and an inland water body. Figure 3 shows the Jasper image along with the groundtruth. The Jasper image has 100 rows, 100 columns, and has 224 bands. Figure 4 shows the four endmember abundances for the materials present in the Jasper image. The endmembers are road, soil, water, and tree. We did not consider the road class because of an insufficient number of pixels for training. The available groundtruth has endmember abundances for each of the pixels. In [29] random labeling of the HSI pixels is used for creating labels. Here, we conduct two classification experiments by generating labels based on groundtruth endmember abundances. For the first experiment, to perform a fixed classification of each pixel to a particular class, we created three labels for each pixel from the endmember abundances as strongly belong, weakly belong, and does not belong to one of the three original groundtruth classes. If the fractional abundance is greater than 0.8, then the pixel is labeled as strongly belonging to the class. If the fractional abundance is less than 0.8, then the pixel weakly belongs to the class, and if the abundance is 0 the pixel does not belong to the class. All the four machines are trained with training batches for the three groundtruth classes and for the three labels for each of the three groundtruth classes resulting in training of nine classes. We also conduct a second classification experiment with two labels for pixels. The pixel is labeled as strongly belonging to the class if the abundance is less than the groundtruth maximum value for the class and greater than 0.4. If the pixel value is greater than the minimum groundtruth value and less than 0.4, it is labeled as not belonging to the class, resulting in the training of 6 classes. For both the experiments, 10 fold cross-validation is done which results in the training of a total of 90, and 60 models for both experiments, respectively. We have effectively converted an unmixing problem into a classification problem by assigning fixed labels to pixels with fractional abundances by thresholding. The procedure for preprocessing and extraction of batch sizes for training and testing are explained below.

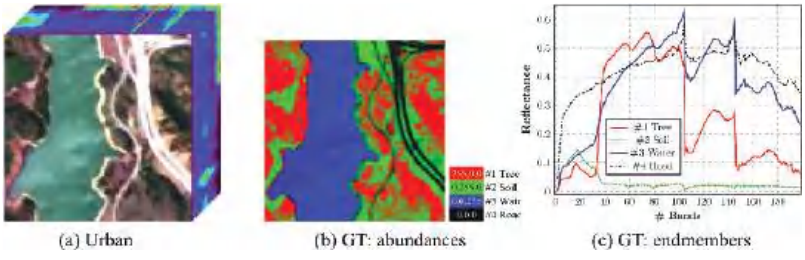


Figure 3. Jasper HSI (**a**) original image, (**b**) Groundtruth abundances, (**c**) Groundtruth endmembers.

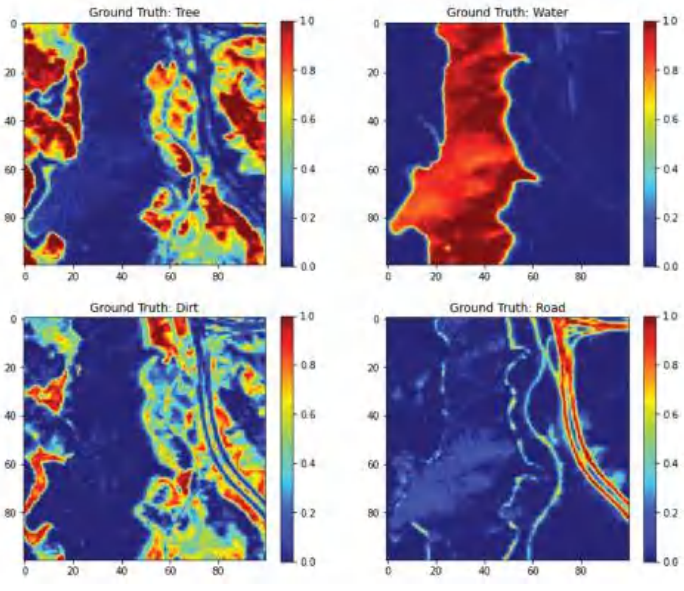


Figure 4. Endmember abundances for the four endmembers for Jasper image.

Firstly, PCA is applied to the Jasper image, and three, five, and seven dominant PCA bands are selected. The batch selection process consists of the random extraction of parts of the image by class. The batches are divided into groups for training, testing, and left-over data. Batch sizes for the training data are 820, 1000, and 1500 pixels. The data is split into training data, testing data corresponding to the same selected batch size as training data, and the remaining data not used for training or testing is used only for the image reconstruction. This data is around 400 pixels. The training is done with less than 2% of the pixels of Jasper HSI for each class. Figure 5 explains the batch size extraction process for three PCA bands with min–max scaling. The experiment is repeated with max-scaling and normalization. Finally, the batches are stacked for training the models. For two labels (strongly belong, and does not belong), the training batch sizes are (6×820 pixels), where 6 corresponds to two batches for each of the three PCA bands. The six batches per class are stacked together for training the models for all three classes. The testing batch sizes are (9×820 pixels) where 9 corresponds to three batches for each of the three PCA bands which are stacked together for testing for the three classes. There is a remaining 980 pixels of left-over data which is used for image reconstruction. The experiment is repeated for batch sizes of 1000 and 1500 pixels. For three labels (strongly belong, weakly belong, and does not belong), the batch sizes are smaller: 300, 500, and 600 pixels. The training is done on the features extracted from the batches.

The features are energy, mean, and standard deviation which are calculated on the batches of pixels. The ML models are trained with the computed features. The ensemble model for the training process for the three classes, trees, water, and soil, is shown in Figure 6. The labeled testing pixels are then used to reconstruct the classified Jasper image with color code for each class.

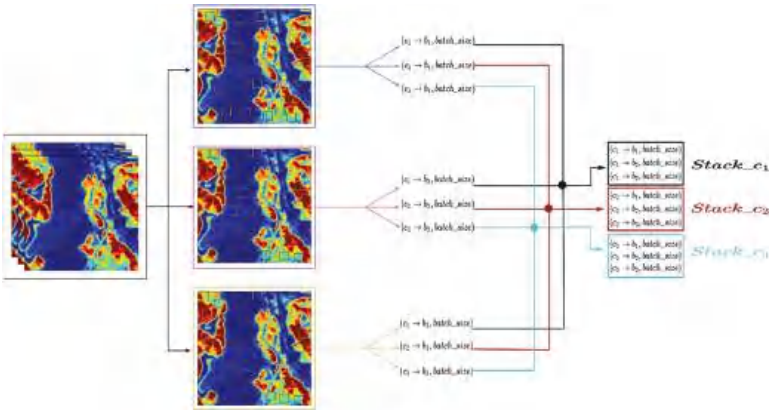


Figure 5. Batch size selection process for the three PCA bands from the Jasper HSI.

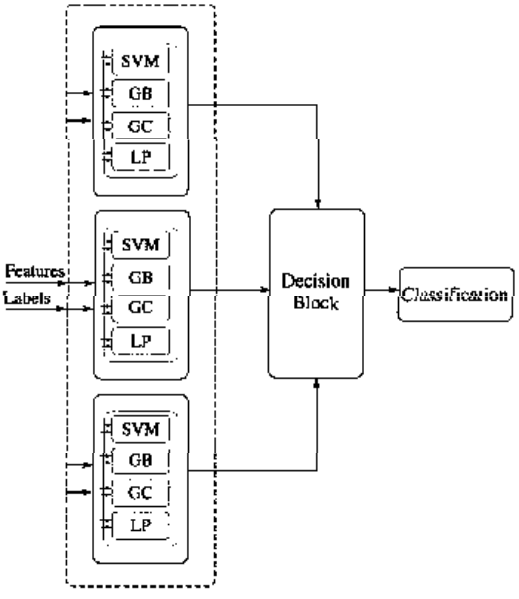


Figure 6. Energy, mean, and variance features are calculated from the Jasper HSI training samples and are input to the ML algorithms. The decision block selects the best machine for each class and uses the selected machines to label the testing samples.

Pseudocode description of the algorithms for the image enhancement features block, supervised ML block, and decision block are given below.

A. Pseudo code feature enhancement block

Input: Hyperspectral Image
Output: Stacked vector Enhancement
Begin:
compute the energy feature using Equation (1)
compute the mean using Equation (2)
compute the standard deviation using Equation (3)
concatenate the energy, mean and standard deviation in to a data frame
Return Stacked enhancement vector

The enhancement features block is applied to obtain the spectral features representation. The input is the 1-D reshaped hyperspectral image vector placed as columns for each of the bands, then the energy, mean, and standard deviation feature are extracted. Finally, the data is stacked in to a data frame.

B. Pseudo code supervised machine learning block

Input: dataset train (data), label for dataset train (label), tolerance, kernel, depth, estimators
Output: Models
Begin: Initialize variables for accuracy, F1 score, confusion matrix for the models (metrics)
For 10-fold cross validation of the data
    compute **SVM** Model using data, label, and tolerance
    compute **GB** Model using data, label, estimators, and depth
    compute **LP** Model using data, label, and tolerance
    compute **GC** Model using data, label, and kernel
    compute accuracy score for the four models
    compute F1 score for the four models
    compute confusion matrix score for the four models
    save (SVM Model, GB Model, LP Model, GC Model)
    append accuracy, F1-score, confusion matrix
Return Models, metrics

The unsupervised machine learning block proposed is composed of four machine learning methods: SVM, GB, GC, LP. The models are trained using a 10-fold cross-validation methodology. Then, the input of the machine learning blocks is the selected training data, the respective labels, and the tuning parameters. The tuning parameters are configured for each machine learning technique as follow:

SVM is set using a linear kernel, and hinge as a loss function and tolerance values of 1×10^{-3}. The Gradient Boosting parameter is the depth of the individual regression estimator which is set to 10, the number of boosting stages is 100, and the learning rate for each tree is 1.0. The LP classifier is set to tolerance or stopping criteria of 1×10^{-5}. The Gaussian classifier is set with the RBF kernel using L-BFGS quasi-Newton methods as an optimization function.

C. Pseudo code decision block

Input: data_test (batch_size, features), label (batch_size), models
Output: Best classifiers
Begin: Initialize dictionary metrics variable (accuracy, F1 score, confusion matrix, training data, predicted labels), maximum accuracy variable
For each folderModels
    For each Model
        load model
        compute accuracy
        compute F1 score
        compute confusion matrix
        append accuracy, F1 score, confusion matrix, model, and variables in dictionary metrics
    concatenate dictionary metrics in a pandas data frame
obtain the best model classifier using the accuracy criteria
Return best classifier

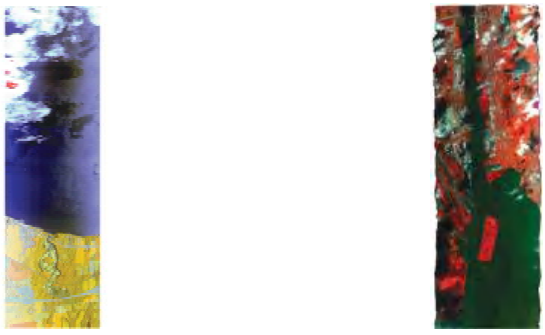
The above pseudocode procedure is for the principal blocks of the workflow in Figures 2 and 6. The rest of the blocks that include preprocessing methods for scaling, and dimensionality reduction using PCA are straightforward to compute.

3. Results

This section presents and discusses the results of applying the ensemble method for the labeling, classification, and reconstruction of the HSI2 images and Jasper hyperspectral images.

3.1. Classification and Reconstruction of HSI2 Images

The semi-supervised classification pipeline is applied to two HSI2 images over Lake Erie. The HSI2 images (Image 1 and Image 2) are shown in Figure 7. Image 1 is of size 3270 × 960, where 3270 is the number of lines, and 960 is the number of samples per line. Image 2 in Figure 7b is of size 4444 × 960, where 4444 is the number of lines, and 960 is the number of samples per line. The semi-supervised classification scheme shown in Figure 2 is applied to the images shown in Figure 7a,b.



(**a**) (**b**)

Figure 7. HSI2 Hyperspectral images over Lake Erie (**a**) Image 1 (white—clouds, blue—water, yellow—land), and (**b**) Image 2 (white—clouds, green—water, red—land).

The unsupervised stage for segmenting the image into clusters is applied for a choice of 2, 3, 4, and 5 clusters. This stage performs k-means clustering after image enhancement using the standard deviation and energy features. This combination and 3 numbers of clusters give the best results for labeling and obtaining the groundtruth image. Following unsupervised classification which identifies the best number of clusters image preprocessing is performed. PCA is used for selecting the best number of a subset of bands. There are 51 bands in the HSI2. The covariance matrix of the image and its Eigenvalues are computed. Figure 8 shows the percentage of contribution of the first ten bands to the Eigenvalues of the covariance matrix of the image. It can be seen that all the energy is compacted in the first three bands of the image.

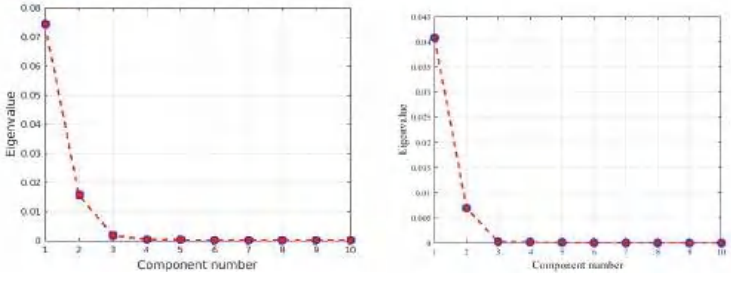


Figure 8. Scree plot of the contribution of each principal component bands for the two hyperspectral images in Figure 7 (Left Image 1 and Right Image 2).

Following preprocessing, extraction of the three mean, standard deviation, and energy features is performed. For supervised classification, all the three features are used. The features are stacked for training the four ML methods. The Ensemble of the four machines is applied to HSI2 images 1 and 2 in Figure 7. The decision block loads all the models of the 10 fold cross-validation process, and classifies the images with all the models, and choose the best model using the classification accuracy as the selection metric.

Tables 1 and 2 shows the accuracy and F1 score obtained from 10-fold cross validation for three clusters with 3, 5, and 7 PCA bands, using the three scaling methods of standardization normalization (ns), max scaling (ms), and min–max scaling(sc) for the HSI2 Image 1 in Figure 7a, and HSI2 Image 2 in Figure 7b, respectively. The three clusters are land, water, and clouds. Tables 3 and 4 show the accuracy and F1 score obtained from 10-fold cross validation for three clusters with 3, 5, and 7 PCA bands, using the three scaling methods of normalization scaling, min–max scaling, and max scaling for the HSI2 Image 2 in Figure 7b.

Table 1. The classification accuracy for HSI2 image 1 using PCA 3, 5, and 7 bands and the different scaling and normalization methods using the four machines for a 1500 pixels batch size.

	Accuracy			
	SVM	**LP**	**GB**	**GC**
PCA-3 ns	68.46	56.88	100.00	87.77
PCA-3 ms	72.67	56.88	100.00	99.94
PCA-3 sc	72.67	56.88	100.00	99.90
PCA-5 ns	63.80	56.88	99.65	81.63
PCA-5 ms	58.55	56.88	99.97	98.87
PCA-5 sc	58.55	56.88	99.93	97.77
PCA-7 ns	58.75	56.88	98.95	79.13
PCA-7 ms	58.07	56.88	99.92	97.26
PCA-7 sc	58.07	56.88	99.97	97.26

Table 2. The F1 score for classification of HSI2 image 1 using PCA 3, 5, and 7 bands and the different scaling and normalization methods using the four machines.

	F1-Score			
	SVM	**LP**	**GB**	**GC**
PCA-3 ns	61.80	41.22	100.00	89.75
PCA-3 ms	70.43	41.22	100.00	100.00
PCA-3 sc	70.43	41.22	100.00	99.92
PCA-5 ns	54.97	41.22	99.79	83.41
PCA-5 ms	44.46	41.22	99.90	90.06
PCA-5 sc	44.46	41.22	99.95	85.85
PCA-7 ns	43.64	41.22	99.21	79.53
PCA-7 ms	43.64	41.22	99.95	88.56
PCA-7 sc	43.64	41.22	99.88	87.96

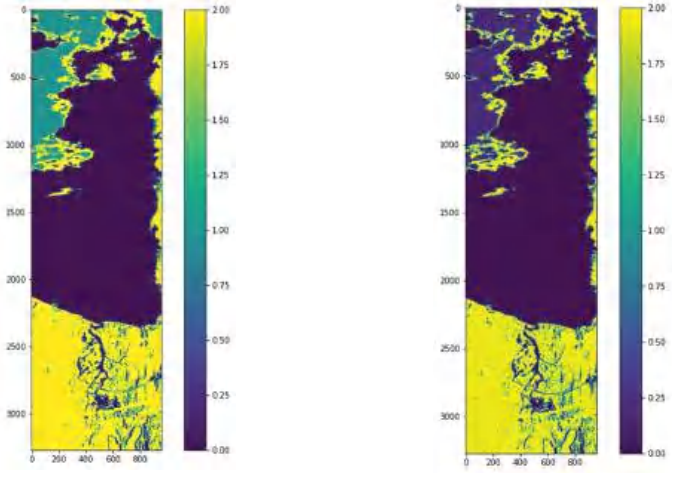
Table 3. The classification accuracy for HSI2 image 2 using PCA 3, 5, and 7 bands and the different scaling and normalization methods using the four machines for a 1500 pixels batch size.

	Accuracy			
	SVM	**LP**	**GB**	**GC**
PCA-3 ns	83.97	52.88	100	98.46
PCA-3 ms	53.54	53.54	100	94.94
PCA-3 sc	53.54	53.54	100	93.76
PCA-5 ns	64.81	52.88	99.81	94.74
PCA-5 ms	52.88	52.88	98.19	88.75
PCA-5 sc	52.88	52.88	98.64	88.16
PCA-7 ns	56.58	52.88	86.47	79.98
PCA-7 ms	52.88	52.88	86.88	79.7
PCA-7 sc	52.88	52.88	87.08	79.95

Table 4. The F1 score for classification of HSI2 image 2 using PCA 3, 5, and 7 bands and the different scaling and normalization methods using the four machines.

	F1-Score			
	SVM	**LP**	**GB**	**GC**
PCA-3 ns	79.12	36.58	100	98.46
PCA-3 ms	48.25	48.25	100	94.91
PCA-3 sc	48.25	48.25	100	93.78
PCA-5 ns	61.17	36.58	99.81	94.73
PCA-5 ms	36.58	36.58	98.18	88.78
PCA-5 sc	36.58	36.58	98.63	88.19
PCA-7 ns	52.09	36.58	86.22	79.8
PCA-7 ms	36.58	36.58	86.33	80.23
PCA-7 sc	36.58	36.58	86.57	80.53

The models are trained with the extracted features for a batch size of 1500 for HSI2 image 1 and image 2 for three classes. For both images, the best batch size is found to be 1500 pixels compared to 1000 pixels batch size. The trained models are then used to classify the images into three classes. The classified images are reconstructed. The best accuracy is obtained with the GB model and 3 PCA bands for both images. The reconstructed labeled image and reconstructed classified image for HSI2 image 1 are shown in Figure 9a,b, respectively. The SSIM between the labeled image and the reconstructed image is 0.6743.



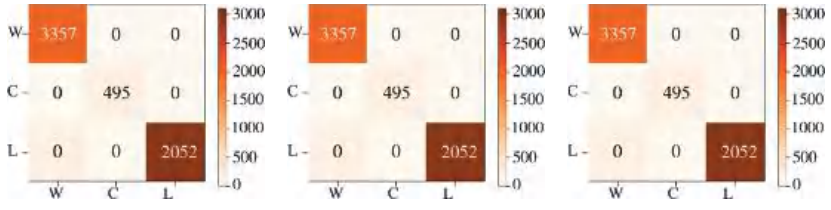
(**a**) (**b**)

Figure 9. (**a**) HSI2 Image 1 with labels for 3 classes using the unsupervised stage (k-means clustering method), (**b**) Reconstructed image from classified samples using the supervised stage.

Figure 10 shows the confusion matrices for the reconstruction of HSI2 Image 1 using 3 PCA bands with the three types of scaling methods. All of the scaling methods give 100% accuracy using GB classifier, and second best classifier is GC.

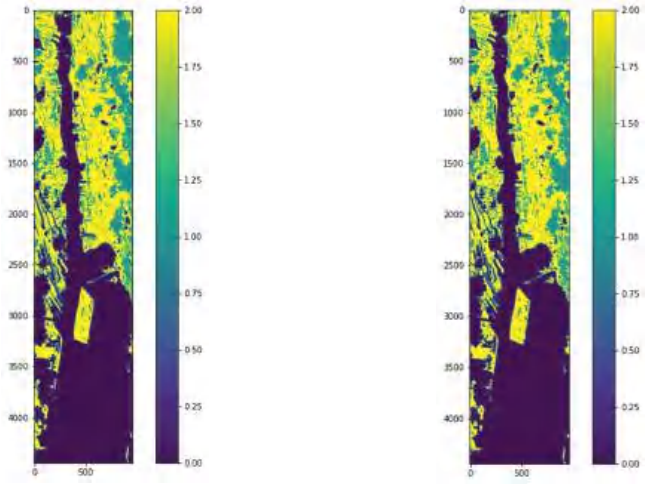
The labeled image and classified reconstructed image for HSI2 image 2 are shown in Figure 11a,b, respectively. The classified image of HSI2 image 2 using 3 PCA bands and maximum scaling (ms) gives the best similarity with the labeled image with the SSIM being 1.0. Figure 12 shows the confusion matrices for the reconstruction of HSI2 Image 2 using 3 PCA bands with the three types of scaling methods. Overall, the scaling and maximum scaling methods give 100% accuracy, and the normalization scaling gives 99.95% accuracy. For HSI image 1, the highest accuracies are obtained for using 3 PCA bands and maximum

scaling, and for HSI image 2, the highest accuracies are obtained for using 3 PCA bands and normalization scaling. The confusion matrices in Figures 10 and 12 have different number of testing samples, as the HSI2 images 1 and 2 are of different size.



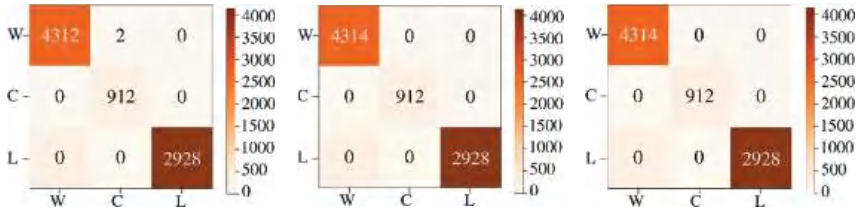
(**a**) (**b**) (**c**)

Figure 10. Confusion matrices for classification of HSI2 Image 1 into 3 classes for a batch size of 1500 using 3 PCA bands with the three types of scaling methods (**a**) normalization scaling (ns), (**b**) maximum scaling (ms), (**c**) scaling (sc). The three classes are indicated as W—water, C—clouds, L—land.



(**a**) (**b**)

Figure 11. (**a**) HIS2 Image 2 with labels for 3 classes using the unsupervised stage (k-means clustering method), (**b**) Reconstructed image from classified samples using the supervised stage.

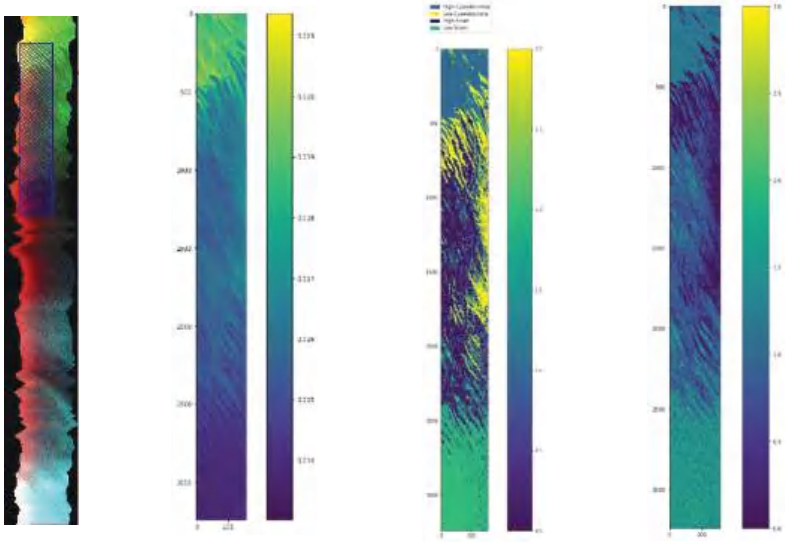


(**a**) (**b**) (**c**)

Figure 12. Confusion matrices for classification of HSI2 Image 2 into 3 classes for a batch size of 1500 using 3 PCA bands with the three types of scaling methods (**a**) normalization scaling (ns), (**b**) maximum scaling (ms), (**c**) scaling (sc).

3.2. Classification of Cyanohab from Lake Erie Image

We use another HSI image shown in Figure 13 to show that the ensemble semi-supervised scheme can identify blue green algae or cyanoHAB (cyanobacteria) from other materials in the lake. This image was acquired using a different sensor than HSI2, the data has a different format. The image is of size 5000 lines, with 496 samples per line. We used ENVI to obtain the ROIs for the cyanobacteria and surface scum. This image shows higher concentrations of cyanobacteria and also surface scum. A Region of Interest (ROI) with a high concentration of cyanobacteria in the East side of the lake, highlighted by a blue rectangle in Figure 13a is extracted. The ROI image is of size 3240 lines, with 311 samples per line. The ROI image is shown enlarged in Figure 13b. The image is stored in 'tif' format in 170 bands and the proposed workflow shown in Figure 2 is applied, similar to the classification of HSI2 images. The enhancement stage performs feature extraction of the textural energy and statistical mean and standard deviation features. Then, the vectors are stacked using a Pandas data frame structure. The next stage is the unsupervised stage for label assignment in the image using a k-means clustering that takes as input the stacked features vectors. The output of this block are the labels and data. The labeled image with four clusters is shown in Figure 13c. A preprocessing stage is performed using data normalization followed by feature extraction. The previous outputs are the inputs for the supervised machine learning ensemble trained with a batch size of 1000 by 3 features similar to the previous experiment on Lake Erie HSI2 images. After training, the decision block decides the best of the four machines using majority voting, using which the final classification and reconstructed image are obtained. The classified reconstructed image is shown in Figure 13d.



(**a**) (**b**) (**c**) (**d**)

Figure 13. (**a**) Hyperspectral image of Lake Erie with extracted ROI shown as blue rectangle, (**b**) Zoomed ROI subimage, (**c**) Output image from unsupervised stage with four clusters, (**d**) Classified reconstructed image using 3 PCA bands and normalization scaling (Legends are the same as image in (**c**)).

The classification accuracies using the supervised stage of the ensemble method for the 3 scaling methods, and 3, 5, and 7 PCA bands are given in Table 5. As can be seen, the three PCA bands result in higher accuracies using the Gradient Boosting classifier. The F1 shores are given in Table 6.

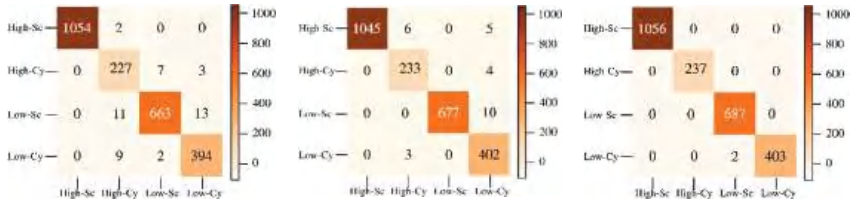
Table 5. The classification accuracy for CyanoHAB HSI using PCA 3, 5, and 7 bands, with the three types of scaling using the four machines for a 1000 pixels batch size.

	Accuracy				
	SVM	**LP**	**GB**	**GC**	
PCA-3 ns	63.48	53.88	99.92	91.53	PCA-3
PCA-3 ms	53.88	53.88	99.33	90.15	
PCA-3 sc	53.71	53.88	98.53	83.73	
PCA-5 ns	61.03	50.04	96.33	76.93	PCA-5
PCA-5 ms	50.04	50.04	97.36	64.18	
PCA-5 sc	49.89	50.04	97.08	74.11	
PCA-7 ns	50.13	48.39	90.71	67.78	PCA-7
PCA-7 ms	48.39	48.39	95.11	56.17	
PCA-7 sc	48.81	48.39	91.25	62.66	

Table 6. The F1 score for classification of CyanoHAB HSI using PCA 3, 5, and 7 bands with the three types of scaling using the four machines.

	F1-Score				
	SVM	**LP**	**GB**	**GC**	
PCA-3 ns	53.76	43.52	99.92	91.32	PCA-3
PCA-3 ms	43.52	43.52	99.33	89.50	
PCA-3 sc	43.38	43.52	98.54	83.25	
PCA-5 ns	51.30	37.91	96.35	77.70	PCA-5
PCA-5 ms	37.91	37.91	97.38	64.13	
PCA-5 sc	37.80	37.91	97.10	73.01	
PCA-7 ns	38.49	35.18	90.88	64.37	PCA-7
PCA-7 ms	35.18	35.18	95.14	57.10	
PCA-7 sc	40.19	35.18	91.60	63.39	

The confusion matrices for the 4 classes are given in Figure 14. The classified reconstructed image shown in Figure 13d has the same color legends as Figure 13c.



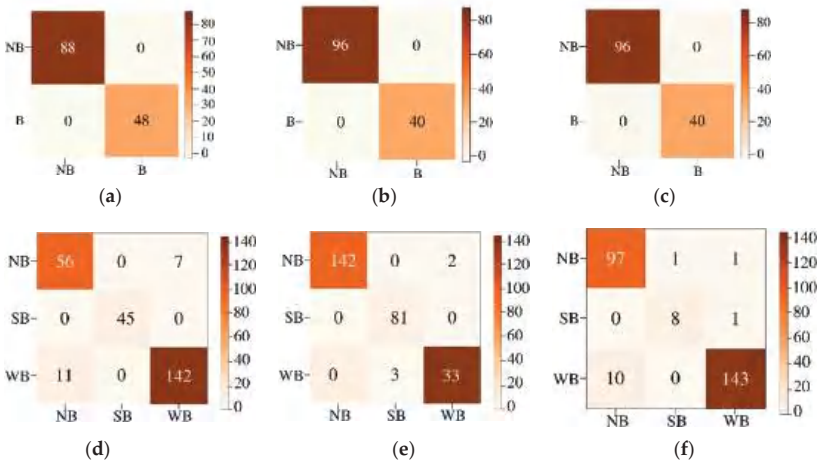
(**a**) (**b**) (**c**)

Figure 14. The confusion matrices for 4 labels High Scum, High Cynanobacteria, Low Scum, and Low Cyanobacteria pixels of the CyanoHAB image for three types of scaling- (**a**) normalization scaling (ns), (**b**) maximum scaling (ms), and (**c**) scaling (sc).

3.3. Classification and Reconstruction of Jasper Image

Jasper HSI has 4 different materials with mixing in each pixel. We did not consider the road class because of insufficient data for training. The three considered classes are trees, water, and soil. The Jasper image pixels have been classified into subcategories: Belong (B) and Not Belong (NB) and to three subcategories: Strong Belong (SB), Weak Belong (WB), and Not Belong (NB) to give fixed labels to the groundtruth pixels with fractional abundances. For two labels within the three classes of trees, water, and soil, 2×2 confusion matrices are obtained for each of the three classes, and for three labels within the three

classes, 3 × 3 confusion matrices are obtained (shown in Figure 15) for each of the three classes. For the two subcategories experiment, we have a total of 136 testing samples, and for the three subcategories experiment, we have a total of 261 testing samples.



(**a**) (**b**) (**c**)

(**d**) (**e**) (**f**)

Figure 15. The confusion matrices for 2 labels Not Belong (NB) and Belong (B) given to the (**a**) trees, (**b**) water, and (**c**) soil pixels of the Jasper HSI in the top row. Bottom row shows the confusion matrices for 3 labels Strong Belong (SB), Weak Belong (WB), and Not Belong (NB) given to the (**d**) trees, (**e**) water and (**f**) soil pixels of the Jasper HSI.

We divided the data into batches of training and testing sizes and computed the classification accuracies using 10-fold cross validation. Figure 16 shows the classified images and the original groundtruth endmembers for each of the classes for classifying with three labels per class.

Tables 7 and 8 show the accuracy and F1 score obtained from 10-fold cross validation for three clusters with 3, 5, and 7 PCA bands, using the three scaling methods of standardization normalization, min–max scaling, and maximum scaling for the Jasper HSI. The structural similarity index measure (SSIM) between the original and reconstructed image pixels is 1.0 for the tree, water, and soil classes. Best results are obtained with three PCA bands and maximum scaling. The batch sizes for training and classification for three labels per class are 300, 500, and 600. The best batch size is found to be 300 pixels.

Table 7. The classification accuracy for Jasper HSI using PCA 3, 5, and 7 bands, with maximum scaling using the four machines for a 300 pixels batch size.

	Accuracy				
	SVM	**LP**	**GB**	**GC**	
Tree	63.60	60.54	95.02	82.76	
Water	69.73	55.17	96.93	77.78	PCA-3
Soil	62.07	66.28	89.27	72.80	
Tree	58.62	58.62	77.47	72.41	
Water	61.61	55.17	93.10	69.20	PCA-5
Soil	64.14	63.45	85.29	71.49	
Tree	58.62	59.44	68.80	68.97	
Water	55.83	55.17	76.52	64.20	PCA-7
Soil	60.26	62.73	80.13	69.13	

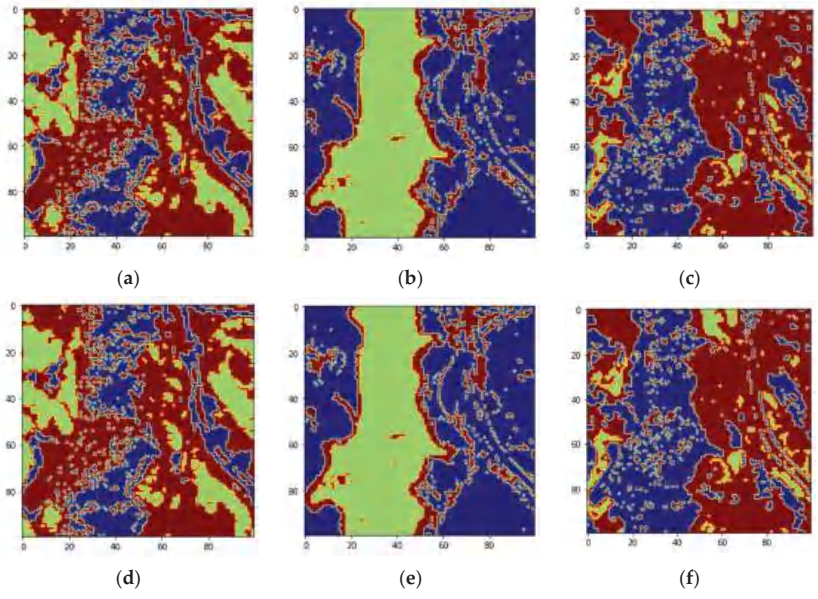


Figure 16. (**a**) Tree class reconstructed image with label 0 (blue) for pixels Not Belonging to tree class, label 1(green) for Strong Belonging pixels to tree class, and label 2 (brown) for pixels Weakly Belonging to tree class. (**b**) Water class reconstructed image with label 3 (blue) for pixels Not Belonging to water class, label 4 (green) for Strong Belonging to water class, and label 5 (brown) for pixels Weakly Belonging to water class, (**c**) Soil class reconstructed image with label 6 (blue) for pixels Not Belonging to soil class, label 7 (green) for pixels Strongly Belonging to soil class, and label 8 (brown) for pixels Weakly Belonging to Soil class, (**d**) Original groundtruth for tree class, (**e**) Original groundtruth for water class, and (**f**) Original groundtruth image for soil class.

Table 8. The F1 score for classification of Jasper HSI using PCA 3, 5, and 7 bands with maximum scaling using the four machines.

	F1-Score				
	SVM	**LP**	**GB**	**GC**	
Tree	54.67	47.31	95.03	82.93	
Water	64.27	39.23	96.90	78.86	PCA-3
Soil	51.64	58.98	89.35	71.67	
Tree	43.33	43.33	77.98	69.97	
Water	57.04	39.23	93.16	70.95	PCA-5
Soil	61.99	53.56	86.38	71.84	
Tree	43.33	45.12	69.64	64.98	
Water	52.00	39.23	76.87	65.12	PCA-7
Soil	59.43	51.93	82.53	66.53	

The Jasper image is also classified into two labels per pixel versus Not Belong and Strong Belong by thresholding the fractional abundances as discussed in Section 2.1. The batch sizes for training and classification for two labels per class are 820, 1000, and 1500. The results of the reconstructed images for each endmember are shown in Figure 17. The SSIM for these reconstructions is also 1.0 giving the highest similarity between original and reconstructed images.

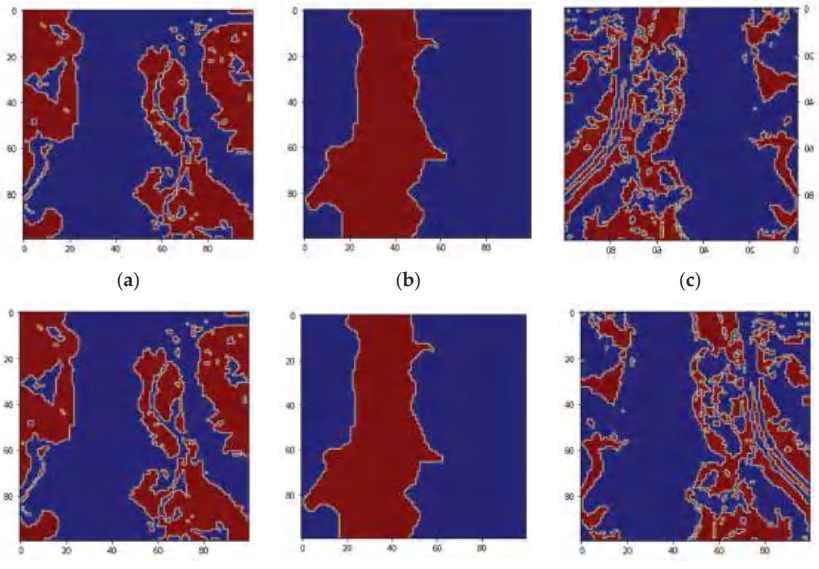


Figure 17. (**a**) Tree class reconstructed image with label 0 (blue) for pixels Not Belonging to tree class, label 1(brown) for pixels Belonging to tree class. (**b**) Water class reconstructed image with label 3 (blue) for pixels Not Belonging to water class, and label 4 (brown) for pixels Belonging to water class, (**c**) Soil class reconstructed image with label 6 (blue) for pixels Not Belonging to soil class, and label 7 (brown) for pixels Belonging to soil class, (**d**) Original groundtruth for tree class, (**e**) Original groundtruth for water class, and (**f**) Original groundtruth image for soil class.

4. Discussion

4.1. Discussion of Ensemble Model Results for HSI2 Images of Lake Erie

The semi-supervised ensemble method pipeline is larger for Lake Erie images because we do not have the labeled groundtruth data. The labeled data has to be created using the unsupervised stage of the pipeline. Moreover, the image enhancement stage makes use of all the 51 bands of the original image to compute the features that are input to the unsupervised stage. The enhancement stage is important as it improves the labeling of the original HSI dataset. Both images are labeled by the unsupervised stage into 3 classes: clouds, land, and water. The supervised stage implements four ML models and the output classified images are obtained after 10-fold cross validation. The best batch size is 1500 pixels stacked for the 3 features. The SSIM is 0.6743 for Lake Erie Image 1 while it is 1.0 for image 2 which has more land cover than image 1. This is because of the higher cloud cover in image 1. Optical remote sensing imagery has the problem of cloud cover and thresholding methods are applied for their removal from hyperspectral imagery [30]. Onboard spectral–spatial method is proposed in [31] for cloud detection. A deep learning neural network method is proposed for cloud detection in [32]. Our method can be used for masking and filtering cloud cover pixels before classification of the image. The advantage of our ensemble method is that the identification of cloud pixels is part of the labeling process in the pipeline, which is followed by supervised classification using the ensemble ML technique.

4.2. Discussion of Ensemble Model Results for CyanoHAB Image of Lake Erie

The spectral signature of pixels in the four clusters from the ROI image in Figure 13c is shown in Figure 18. The bands 679 nm, 664 nm, and 709 nm, and the bands 667 nm and 858 nm are used to calculate the Cyanobacteria Index (CI) and Surface Scum Index

(SSI), respectively in [19]. As can be seen from the output of the unsupervised stage the regions of High CyanoHAB, Low CyanoHAB, High Scum, and Low Scum are identified correctly compared to the images obtained from the CI and SSI in [19]. The classification accuracies for the supervised classification of CyanoHAB is 99.92%. Our classification of High CyanoHAB, High Scum, and Low Scum are good, but the accuracy is low for low cyanobacteria concentration. The classification of the low cyanobacteria class can be improved by spectral feature extraction. Our semi-supervised ensemble scheme can be used for the identification of cyanobacteria from hyperspectral images in an automatic manner without human intervention and the need for labeled samples. Moreover, the CI and SSI give a fractional index of the materials with one image per material. While our classification pipeline gives fixed labels for each pixel which will be more useful for water management as they know definitely which areas pertain to harmful cyanobacteria, and which are safer for recreation and other activities.

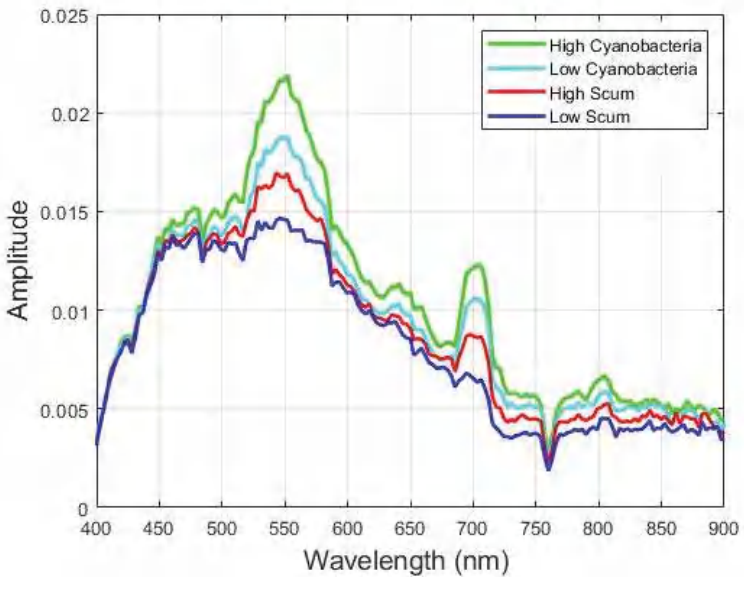


Figure 18. Spectral signature profiles of the CyanoHAB found in the Lake Erie Hyperspectral image.

4.3. Discussion of Ensemble Model Results for Jasper HSI

For two subclasses by label, the classification results were compared for using three, five, and seven PCA bands. The best classification performance was obtained with the GB classifier, with three PCA bands giving an accuracy of 91.57%, and 100%, 100% for the tree, water, and soil classes, respectively. For the tree and soil classes a better accuracy is obtained with five PCA bands. On the other hand, for water classification, the performance is better with seven PCA bands. For larger batch sizes, e.g., 1000 and 1500, lower number of PCA bands such as three, gives an accuracy of 100% for the three classes with GB classifier. For the three labels per class configuration of SB, WB, and NB, three PCA bands for a batch size of 300 gives the best classification accuracy of 95.02%, 96.23%, and 89.27% for the tree, water, and soil classes, respectively, using the GB classifier. For the water class, we obtained better accuracy of 99.07% using a batch size of 600. The three labels per class configuration also gives a SSIM of 1.0 for the reconstructed image compared to the original groundtruth image. The best results for the Jasper image can be summarized as the use of three PCA bands with min–max scaling and GB classifier. The GB is found to be the best classifier as it is based on decision trees and it combines many weak learners to create a strong predictive model.

Jasper dataset has four endmembers' contributions corresponding to tree, soil, water, and road. A graph-based architecture is proposed in [33] to classify the endmember contributions and, the authors compare the classification performance with ML techniques such as SVM, KNN, LDAKNN, PCAKNN, KPCAKNN, LDASVM, PCASVM KPCASVM, Convolution Neural Network (CNN), the abbreviation Linear Discriminant Analysis (LDA), Kernel PCA (KPCA) in the machine learning methods means the preprocessing step before the classification techniques. The accuracy measurements for Jasper image for classification into the endmembers contribution by class is as follows: Soil Class obtained 100% accuracy results using the PCAKNN method and SVM 99.905%, for water, and TLM-2 classifier obtained 98.959%, Finally, for tree class, TLM-2 obtained 97.622%. From our experiments, in the separate analysis for three classes using a labeled subset of non-belong (NB), strong-belong (SB), and weak-belong (WB), and using the feature extracted from three PCA bands, we obtained the following accuracies for three classes: trees 95.02%, water class 96.93%, and soil class 89.27% for the GB classifier. On the other hand, for two classes using as a label sub-set of non-belong (NB) and belong (B), we obtained 91.67%, 100.00%, and 100.00% for the three classes of trees, water, and soil, respectively. We improve the results compared to the method proposed in [33] for our two sub-labels approach. The best scaling method for Jasper dataset was the min–max-scaling. Our ensemble method improves the water and soil classification accuracies using 24.6% of the dataset for training and the remaining data for testing. In [29], the authors propose a Kernelized Extreme Learning Machine (K-ELM) using 2000 samples for training and the obtained accuracy score for a groundtruth labeling in the re-testing procedure for road, soil, water, and trees as: 84.7%, 98.06%, 69.4%, and 71.1% by class which are lower than the accuracies obtained by our ensemble method, and also requires a larger number of training samples.

The ensemble model can handle unlabeled samples. However, it needs sufficient unlabeled samples for training the machines. Since there are four machines involved the model is time-consuming. In the cloud server, the model takes 6 h 47 min for classifying the Lake Erie images which are still faster than a DELL desktop computer which takes about a week to classify one image. Currently, the model classifies pixels as belonging to particular classes, the future work will involve developing the model to determine fractional abundances of each pixel. Moreover, the future work will involve optimizing the training to work with fewer unlabeled samples using other machines such as DL networks.

5. Conclusions

A semi-supervised ensemble method is presented for labeling pixels in an HSI and classifying the image. The method performs well for airborne HSI over Lake Erie and the Jasper benchmark HSI. In the absence of groundtruth, this method can be used as a preprocessing step for labeling pixels and creating groundtruth data. Moreover, the unsupervised stage effectively detects cloud pixels in the HSI and can be used for cloud removal. The method is able to identify cyanobacteria and other water pollutants from HSI. As with any ML method, sufficient training samples are necessary for adequate training of the machines. The best normalization scheme is found to be maximum scaling, and the number of PCA bands depends on the spectral bands and characteristics of the HSI. For the Lake Erie images and Jasper image dataset, the best number of PCA bands is found to be three. The best ML classifier is found to be the GB classifier for both the Lake Erie and Jasper HSIs. A lower number of PCA bands implies a lesser running time of the models. In the AWS cloud server, the models run in about 6 h and 47 min compared to a regular PC which takes a week for training the models and classification.

Author Contributions: Conceptualization, V.M., E.A.-M.; methodology, E.A.-M., V.M.; software, E.A.-M.; validation, E.A.-M., V.M., R.P.T.; formal analysis, E.A.-M., V.M.; investigation, V.M., E.A.-M., R.P.T.; data curation, E.A.-M., R.P.T.; writing—original draft preparation, V.M., E.A.-M., R.P.T.; writing—review and editing, V.M., E.A.-M., R.P.T.; visualization, V.M., E.A.-M.; supervision, V.M.; project administration, V.M.; funding acquisition, V.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by NASA EPSCoR, grant number 80NSSC21M0155. The APC is funded by 80NSSC21M0155. Opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NASA.

Institutional Review Board Statement: Not Applicable.

Informed Consent Statement: Not Applicable.

Data Availability Statement: The Lake Erie hyperspectral image datasets are available at: https://oceandata.sci.gsfc.nasa.gov/directaccess/HSI-HABS-RAW/ (accessed on 14 December 2021). The Jasper hyperspectral image dataset is available at: http://lesun.weebly.com/hyperspectral-data-set.html (accessed on 14 December 2021).

Acknowledgments: The material contained in this document is based upon work supported by National Aeronautics and Space Administration (NASA) grant 80NSSC19M0167. Opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of NASA.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Dobigeon, N.; Altmann, Y.; Brun, N.; Moussaoui, S. *Linear and Nonlinear Unmixing in Hyperspectral Imaging*, 1st ed.; Elsevier B.V.: Amsterdam, The Netherlands, 2016; pp. 185–224.
2. Sabale, S.P.; Jadhav, C.R. Supervised, Unsupervised, and Semisupervised Classification Methods for Hyperspectral Image Classification—A Review. *Int. J. Sci. Res.* **2014**, *3*, 2319–7064.
3. Sun, W.; Du, Q. Hyperspectral Band Selection: A Review. *IEEE Geosci. Remote Sens. Mag.* **2019**, *7*, 118–139. [CrossRef]
4. Zhu, F. Hyperspectral Unmixing: Groundtruth Labeling, Datasets, Benchmark Performances and Survey. *arXiv* **2017**, arXiv:1708.05125.
5. Navas-Auger, W.; Manian, V. Spatial Low-Rank Tensor Factorization and Unmixing of Hyperspectral Images. *Computers* **2021**, *10*, 78. [CrossRef]
6. Zhang, Y.; Xia, C.; Zhang, X.; Cheng, X.; Feng, G.; Wang, Y.; Gao, Q. Estimating the maize biomass by crop height and narrowband vegetation indices derived from UAV-based hyperspectral images. *Ecol. Indic.* **2021**, *129*, 107985. [CrossRef]
7. Wang, Z.; Jiang, F.; Liu, T.; Xie, F.; Li, P. Attention-Based Spatial and Spectral Network with PCA-Guided Self-Supervised Feature Extraction for Change Detection in Hyperspectral Images. *Remote Sens.* **2021**, *13*, 4927. [CrossRef]
8. Zhang, Y.; Cao, G.; Shafique, A.; Fu, P. Label Propagation Ensemble for Hyperspectral Image Classification. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2019**, *12*, 3623–3636. [CrossRef]
9. Ahmad, M.; Shabbir, S.; Roy, S.K.; Hong, D.; Wu, X.; Yao, J.; Khan, A.M.; Mazzara, M.; Distefano, S.; Chanussot, J. Hyperspectral Image Classification—Traditional to Deep Models: A Survey for Future Prospects. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2021**, *15*, 968–999. [CrossRef]
10. Li, H.; Borsoi, R.A.; Imbiriba, T.; Closas, P.; Bermudez, J.C.M.; Erdogmus, D. Model-Based Deep Autoencoder Networks for Nonlinear Hyperspectral Unmixing. *IEEE Geosci. Remote Sens. Lett.* **2021**, *19*, 1–5. [CrossRef]
11. Jia, P.; Zhang, M.; Shen, Y. Deep spectral unmixing framework via 3D denoising convolutional autoencoder. *IET Image Process.* **2021**, *15*, 1399–1409. [CrossRef]
12. Guo, A.J.; Zhu, F. Improving deep hyperspectral image classification performance with spectral unmixing. *Signal Process.* **2021**, *183*, 107949. [CrossRef]
13. Liu, B.; Yu, X.; Yu, A.; Zhang, P.; Wan, G.; Wang, R. Deep Few-Shot Learning for Hyperspectral Image Classification. *IEEE Trans. Geosci. Remote Sens.* **2018**, *57*, 2290–2304. [CrossRef]
14. Liu, C.; Li, J.; Paoletti, M.E.; Haut, J.M.; Plaza, A.; Shi, Q. Accessibility-Free Active Learning for Hyperspectral Image Classification. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 409–412. [CrossRef]
15. Galilée, I. Ensemble margin based semi-supervised random forest for the classification of hyperspectral image with limited training data. In Proceedings of the IGARSS 2019—2019 IEEE International Geoscience and Remote Sensing Symposium, Yokohama, Japan, 28 July–2 August 2019; pp. 971–974.
16. Zhang, Y.; Cao, G.; Li, X.; Wang, B.; Fu, P. Active Semi-Supervised Random Forest for Hyperspectral Image Classification. *Remote Sens.* **2019**, *11*, 2974. [CrossRef]
17. NOAA and Partners Use Hyperspectral Imagery to Detect HABs in Lake Erie. Available online: https://www.regions.noaa.gov/great-lakes/index.php/highlights/noaa-and-partners-use-hyperspectral-imagery-to-detect-harmful-algal-blooms-habs-in-lake-erie/ (accessed on 15 September 2021).
18. Lekki, J.D. *Airborne Hyperspectral Sensing of Harmful Algal Blooms in the Great Lakes Region: System Calibration and Validation from Photons to Algae Information: The Processes In-Between*; National Aeronautics and Space Administration, Glenn Research Center: Cleveland, OH, USA, 2017; p. 78.

19. Sawtell, R.W.; Anderson, R.; Tokars, R.; Lekki, J.D.; Shuchman, R.A.; Bosse, K.R.; Sayers, M.J. Real time HABs mapping using NASA Glenn hyperspectral imager. *J. Great Lakes Res.* **2019**, *45*, 596–608. [CrossRef]
20. Vander Woude, A.; Ruberg, S.; Johengen, T.; Miller, R.; Stuart, D. Spatial and temporal scales of variability of cyanobacteria harmful algal blooms from NOAA GLERL airborne hyperspectral imagery. *J. Great Lakes Res.* **2019**, *45*, 536–546. [CrossRef]
21. Ortiz, J.D.; Avouris, D.M.; Schiller, S.J.; Luvall, J.C.; Lekki, J.D.; Tokars, R.P.; Anderson, R.C.; Shuchman, R.; Sayers, M.; Becker, R. Evaluating visible derivative spectroscopy by varimax-rotated, principal component analysis of aerial hyperspectral images from the western basin of Lake Erie. *J. Great Lakes Res.* **2019**, *45*, 522–535. [CrossRef]
22. Xu, M.; Liu, H.; Beck, R.; Lekki, J.; Yang, B.; Shu, S.; Kang, E.; Anderson, R.; Johansen, R.; Emery, E.; et al. A spectral space partition guided ensemble method for retrieving chlorophyll-a concentration in inland waters from Sentinel-2A satellite imagery. *J. Great Lakes Res.* **2018**, *45*, 454–465. [CrossRef]
23. Xu, M.; Liu, H.; Beck, R.; Lekki, J.; Yang, B.; Shu, S.; Liu, Y.; Benko, T.; Anderson, R.; Tokars, R.; et al. Regionally and Locally Adaptive Models for Retrieving Chlorophyll-a Concentration in Inland Waters From Remotely Sensed Multispectral and Hyperspectral Imagery. *IEEE Trans. Geosci. Remote Sens.* **2019**, *57*, 4758–4774. [CrossRef]
24. Chapparo, L.F.; Akan, A. Discrete-Time Signals and Systems. In *Signals and Systems Using MATLAB*, 3rd ed.; Academic Press: Cambridge, MA, USA, 2019.
25. Hart, P.E.; Stork, D.G.; Duda, R.O. *Pattern Classification*, 2nd ed.; Wiley: Hoboken, NJ, USA, 2001.
26. Ponomarenko, M.; Egiazarian, K.; Lukin, V.; Abramova, V. Structural Similarity Index with Predictability of Image Blocks. In Proceedings of the International Conference on Mathemathical Methods in Electromagnetic Theory, Kyiv, Ukraine, 2–5 July 2018; pp. 115–118. [CrossRef]
27. Bovik, A.; Wang, Z.; Sheikh, H. *Structural Similarity Based Image Quality Assessment, Digital Video, Image Quality, and Perceptual Coding*, 1st ed.; CRC Press: Boca Raton, FL, USA, 2006; pp. 225–241. [CrossRef]
28. Amazon Web Services. Available online: http://www.aws.com (accessed on 20 October 2021).
29. Hmad, M. Ground truth labeling and samples selection for Hyperspectral Image Classification. *Optik* **2021**, *230*, 166267. [CrossRef]
30. Stournara, P.; Tsakiri-Strati, M.; Patias Candidate, P. Detection and removal of cloud and cloud shadow contamination from hyperspectral images of Hyperion sensor. *South-East. Eur. J. Earth Obs. Geomat. Issue* **2013**, *2*, 33–45.
31. Li, H.; Zheng, H.; Han, C.; Wang, H.; Miao, M. Onboard Spectral and Spatial Cloud Detection for Hyperspectral Remote Sensing Images. *Remote Sens.* **2018**, *10*, 152. [CrossRef]
32. Sun, L.; Yang, X.; Jia, S.; Jia, C.; Wang, Q.; Liu, X.; Wei, J.; Zhou, X. Satellite data cloud detection using deep learning supported by hyperspectral data. *Int. J. Remote Sens.* **2019**, *41*, 1349–1371. [CrossRef]
33. Huang, B.; Ge, L.; Chen, G.; Radenkovic, M.; Wang, X.; Duan, J.; Pan, Z. Nonlocal graph theory based transductive learning for hyperspectral image classification. *Pattern Recognit.* **2021**, *116*, 107967. [CrossRef]

sensors

MDPI

Article

A Convolutional Neural Network-Based Method for Discriminating Shadowed Targets in Frequency-Modulated Continuous-Wave Radar Systems

Ammar Mohanna *, Christian Gianoglio, Ali Rizik and Maurizio Valle *

Department of Electrical, Electronic and Telecommunication Engineering and Naval Architecture (DITEN), University of Genoa, Via Opera Pia 11, 16145 Genoa, Italy; christian.gianoglio@edu.unige.it (C.G.); ali.rizik@edu.unige.it (A.R.)

* Correspondence: ammar.mohanna@unige.it (A.M.); maurizio.valle@unige.it (M.V.)

Abstract: The radar shadow effect prevents reliable target discrimination when a target lies in the shadow region of another target. In this paper, we address this issue in the case of Frequency-Modulated Continuous-Wave (FMCW) radars, which are low-cost and small-sized devices with an increasing number of applications. We propose a novel method based on Convolutional Neural Networks that take as input the spectrograms obtained after a Short-Time Fourier Transform (STFT) analysis of the radar-received signal. The method discerns whether a target is or is not in the shadow region of another target. The proposed method achieves test accuracy of 92% with a standard deviation of 2.86%.

Keywords: radar; shadow effect; machine learning; CNN; transfer learning

Citation: Mohanna, A.; Gianoglio, C.; Rizik, A.; Valle, M. A Convolutional Neural Network-Based Method for Discriminating Shadowed Targets in Frequency-Modulated Continuous-Wave Radar Systems. *Sensors* **2022**, *22*, 1048. https://doi.org/10.3390/s22031048

Academic Editor: Manuel José Cabral dos Santos Reis

Received: 23 December 2021
Accepted: 27 January 2022
Published: 28 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

In recent years, the application of radars for target detection at short and medium ranges has become ubiquitous [1]. The use of short-range Ultra-Wide-Band (UWB) and Continuous-Wave (CW) radars is becoming an attractive solution for localization purposes. Some radar systems applications include through-wall and through-fire detection [2,3], the tracking of moving targets during security operations [4], the detection of trapped people after an avalanche or earthquake [5], and the detection, tracking, and classification of multiple targets passing through a security gate [6].

Up to now, the bi-static radars (with at least one transmitting antenna and at least one receiving antenna) have resolved the detection and localization of a single stationary target, yet the problem of multi-stationary target detection has been less addressed. The bi-static radars are able to accurately detect targets that are closer to the radar antennas, whereas the greater the distance of the targets from the radar, the lower the accuracy of the detection [7]. This is mainly attributed to two factors. Firstly, as the transmission distance increases, the energy of the electromagnetic wave is attenuated; hence, the energy of electromagnetic waves reaching farther targets is inevitably smaller than that reaching the closest target. Secondly, some targets, named recessive targets, can lie in the shadowed region of a dominant target (i.e., the closest to the radar). Thus, because the highest energy of the electromagnetic waves is reflected from the dominant target to the radar, the electromagnetic illumination of the recessive targets could decrease to the point where they are not detected [8]. Therefore, radar systems suffer from what is called the shadowing effect. This effect occurs when two targets stand in front of the antenna, one in the shadowing region of the other. The radar is usually not reliably capable of detecting the target that is standing in the shadow region [7]. This problem is common for most radar technologies, particularly, Ultra-Wide-Band (UWB) radar [2] and Frequency-Modulated Continuous-Wave (FMCW) radar [9]. Unlike pulse and Ultra-Wide-Band (UWB) radars, FMCW systems require lower sampling rates and

lower peak-to-average power ratios to detect the distance and speed of multiple moving targets [10,11]. Accordingly, the FMCW radar is a good solution for detection and localization purposes but performs poorly whenever the shadow effect occurs. The shadow effect is more relevant in low-cost radars. This is due to their lower resolution compared to the high-end radars (higher range and velocity resolution) [12].

In the literature, the shadow effect has been targeted only by a few papers [2,7,8,13–19]. The authors proposed non-scalable solutions, thus requiring expert intervention for applying their methods in different environments. In this paper, we propose a novel solution for solving the issue of target identification in the shadow region and we adopt Deep Learning (DL) techniques. This method is quantitatively analyzed and results are presented in Section 6. It benefits from the promising achievements presented in the literature of applying AI techniques on post-processed radar data. These techniques can help to dynamically learn suitable filters. This proposed solution is also scalable and does not need expert intervention.

In general, DL methods have proven to be very efficient in real-world image classification [20]. Moreover, DL techniques that use radar input are adopted for a wide range of applications, such as target classification [21], object tracking [22], and gesture recognition applications [23]. Among DL techniques, Convolutional Neural Networks (CNNs) are particularly suited for addressing image processing problems [24,25]. Our proposed method uses a lightweight CNN model based on ImageNet (i.e., convolutional filters have been pre-learned based on ImageNet data [26]) to target the discrimination of shadowed targets, fine-tuning only the weights of the last layer (i.e., dense layer). The convolutional layers perform the feature extraction without any prior knowledge of the user. To validate the proposed solution, we address a two-class classification problem: one target vs. two targets. In the latter, one target is in the shadowing region of the other. Four models have been tested using the collected dataset. The best model in terms of accuracy is the MobileNet_V3 Large version; it achieves a generalization performance on the test set of 92.2%. The results encourage us to extend the adoption of CNNs in applications such as identifying and tracking more shadowed targets.

The rest of this paper is organized as follows. In Section 2, the state of the art of the targeted research domain is extensively presented. In the following Section 3, the problem statement is explained. Section 4 presents and discusses the methodology adopted to identify and solve the shadowing effect. The experimental setup is considered in Section 5, explaining the data acquisition process, time frequency analysis, and training process. The experimental results and discussion are presented in Section 6. Finally, the conclusions and some proposals for future work are provided in Section 7.

2. State of the Art

In [13], the shadow effect and its removal using PCL radar is investigated. A study on PCL radar performance under the shadowing effect is presented, when a distant, weak target echo is shadowed by strong echoes. In [7,8], the authors outlined the origin of the shadow effect as the impact of the mutual shadowing of targets in a multiple-person tracking scenario. This explanation is confirmed by the experimental measurements. Other researchers investigated the shadowing effect for the purposes of person detection and tracking with UWB radars [14]. The results confirm the existence of additional attenuation within the shadow zones. In [15], a technique based on wavelet entropy is proposed because of the significant difference in frequency ratio components between the echo signal of the tested target and that of the masked target generated by dynamic clutter. Wavelet entropy can accurately detect multiple human targets in the presence of dynamic clutter, even if the distant human targets are in the shadow area of the closer target, as compared to the reference techniques of adaptive line enhancement and energy accumulation. In [2], a significant difference in frequency was detected between the echo signal of the human target and that of noise in the shadowing region. The authors concluded that the target detection using the power spectrum is not effective. Therefore, an auto-correlation algorithm is

applied to the pre-processed signals in order to compute the wavelet entropy. Results show that the proposed approach is capable of detecting a shadowed target. Other applications have been addressed in the literature [16–19]. In general, none of the previous works have presented a scalable solution for solving the shadowing effect. In fact, these solutions require expert intervention for applying them in different environments.

Several works involving the use of FMCW radar have been reported in the literature. In [23], the authors introduced a novel system for dynamic continuous hand gesture recognition based on a Frequency-Modulated Continuous-Wave radar sensor. They employed a recurrent 3D CNN to perform the classification of dynamic hand gestures and achieved a recognition rate of 96%. In [27], the authors proposed a prototype of an FMCW radar system for the classification of multiple targets passing through a road gate. The classification covered four classes: pedestrians, motorcycles, cars, and trucks. It achieved accuracy of 88.4%. Many other applications have been tackled in the literature [9,28–35]. Most of the systems presented in the aforementioned works suffer from the shadow effect. However, none of them have proposed a solution.

Deep learning classification techniques for radar target classification have also been adopted in the literature. The practical classification of a moving target system, based on automotive radar and deep neural networks, is presented in [36]. The study presents results for the classification of different classes of targets using automotive radar data and different neural networks. In addition, a human–robot classification system based on 25 GHz FMCW radar using micro-Doppler features was introduced in [37]. The raw Range-Doppler images were directly fed into a CNN, resulting in performance with 99% accuracy for distinguishing humans from robots. Many other applications that use neural networks for radar problems have been tackled in the literature [38–42].

3. Problem Statement

3.1. FMCW Radar Device

The multi-chirp FMCW algorithm is considered the standard for detecting and measuring the range and speed of multiple targets [43]. The concept of multi-chirp is to send a frame containing a number of chirps (Nc) in saw-tooth modulation and in a short period, with the chirp time (Tc) being very small (in μs), where Tf is the time of the data frame (Tf is in ms). In the current scenario, the "Position2Go" [44] cheap radar is used. It is an FMCW radar board developed by Infineon technologies. This development kit allows the user to implement and test several sensing applications at the 24 GHz ISM band, such as tracking and collision avoidance. This is possible by using fast chirp FMCW and two receiving antennas to obtain the angle, distance, speed, and direction of motion. The radar is equipped with a pair of arrays of microstrip patch antennas (one for transmitting and two for receiving) characterized by a 12 dBi gain and 19 × 76 degree beam-widths, defining the Field of View (FoV). The kit consists of the BGT24MTR12 transceiver MMIC and a XMC4700 32-bit ARM® Cortex®-M4 for signal processing and communication via USB. The radar is connected via USB to a PC that is running MATLAB. A MATLAB script sends the order to the radar to initiate the data acquisition procedure through the USB port. Table 1 shows the radar sensor parameters.

3.2. Shadow Effect

Figure 1 shows different cases of target detection using a radar. In particular, Figure 1a illustrates the case of a single target standing in the range of the radar, Figure 1b depicts two targets both detectable by the radar, while Figure 1c represents the shadowing effect where target B is masked by target A and thus target B is not visible to the radar.

Table 1. Position2Go radar specifications.

Parameters	Value
Sweep Bandwidth	200 MHz
Center Frequency	24 GHz
Up-Chirp Time	300 µs
Number of Samples/Chirp (Ns)	128
Number of Chirps/Frame (Nc)	32
Maximum Range	50 m
Maximum Velocity	5.4 km/h
Range Resolution	0.75 m
Velocity Resolution	0.4 km/h
Sampling Rate	42 KHz

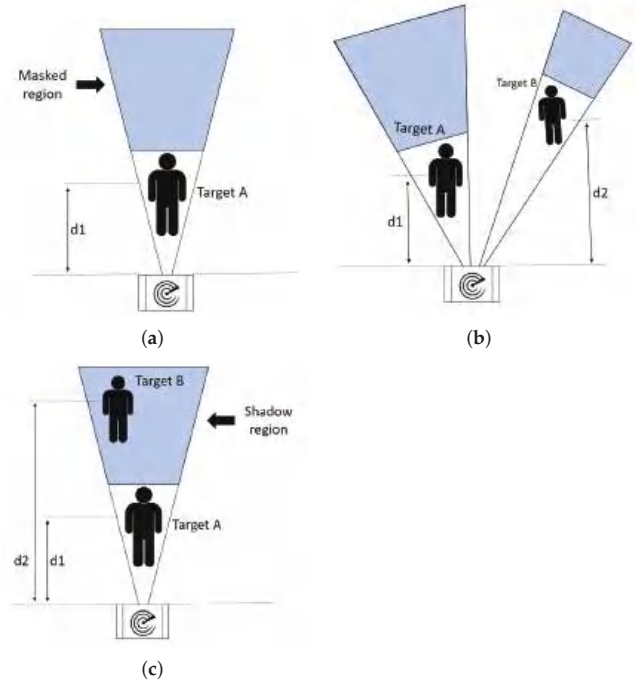


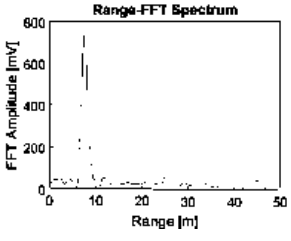
Figure 1. Illustration of the data collection setup. (**a**) One target in range of the radar. (**b**) Two targets in range of radar, both visible to the radar. (**c**) Two targets in range of radar, only one visible to the radar.

The shadowing effect creates a region behind the target (Target A in Figure 1c) where the electromagnetic waves emitted by the radar transmitting antenna or reflected by another object are not able to propagate. In fact, computing the power spectrum on the multi-chirp data acquired by the FMCW radar, it is possible to detect the masked target, but the detection is accompanied by a lot of variability in the measurements. The reason for such variability is that a few radar waves penetrate and slip through the shadowing target to the masked one, reflecting to the radar with a very low intensity. These waves in particular cause huge variability that can affect the detection parameters of both targets in the Field of View (FoV) of the radar. Figure 2 shows three examples of the range representation obtained after the fast-time FFT (range-FFT) on multi-chirp signals [43], positioning the radar 1.5 m from the floor. Each target on the spectrum is represented by a peak. A fixed target detection threshold (red horizontal line) is used to determine the valid target

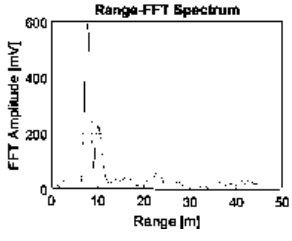
identifications, i.e., each target passing in the FoV of the radar with a peak higher than the fixed threshold is considered a valid detection by the radar. The threshold is a user-defined parameter that affects radar performance directly by causing a trade-off between detection accuracy and false alarm probability. If it is chosen to be too high, the algorithm will fail to identify some targets. If it is too low, the algorithm will detect many artifacts as targets. Figure 2a shows only one peak at a distance of 7 m; this situation is illustrated in Figure 1a, where only one target is in front of the radar ($d1 = 7$ m). In Figure 2b, two peaks appear at distances of 7 m and 10 m, respectively; this spectrum is the result of a trial where two people were standing in different positions (i.e., $d1 = 7$ m and $d2 = 10$ m) with no shadow effect on each other, as illustrated in Figure 1b. The magnitude of the peak at a 10 m is smaller than that at 7 m because of the attenuation of the electromagnetic wave of the radar as the distance increases. In Figure 2c, the maximum peak appears at a distance of 7 m, which corresponds to the location of target A ($d1 = 7$ m). However, target B ($d2 = 10$ m) cannot be detected, since he stands in the shadowing region. An example is illustrated in Figure 1c.

Therefore, the traditional spectrum method is not reliable for detecting multiple targets where the shadow effect occurs. The shadowed targets are hardly detected. This fact is dependent on the chosen power threshold, Radar Cross-Section (RCS) of the shadowing target [45], and the environmental clutter.

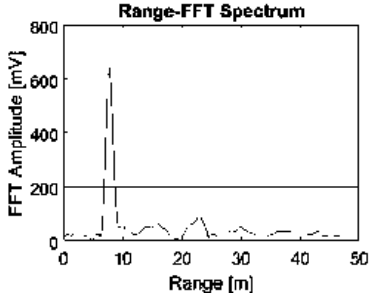
In the case when target B is not fully shadowed by target A, target B is expected to be detected with a weak signal, based on how much it is shadowed by target A. However, this detection is also relative to the chosen detection threshold. The radar is capable of discontinuously detecting target B when it is not completely aligned with target A [7]. However, the detection of target B is not reliable, and the partial shadow effect was excluded from our testing campaign because it represents a simplified version of the full shadow effect illustrated in Figure 1c.



(**a**) One target



(**b**) Two unmasked targets



(**c**) Two targets, one masked

Figure 2. Range-FFT power spectrum. The horizontal red line is the target detection threshold. Radar is positioned 1.5 m from the floor. (**a**) Only target A ($d1 = 7$ m), (**b**) both targets A ($d1 = 7$ m) and B ($d2 = 10$ m) without shadowing effect, (**c**) target B ($d2 = 10$ m) shadowed by target A ($d1 = 7$ m).

4. Methodology

To address the shadow effect, a novel approach is proposed. The idea is that a small portion of the waves slip through or around the shadowing target (target A in Figure 1c) towards the shadowed target (target B in Figure 1c). The masked target is receiving and reflecting these slithered electromagnetic waves, thus causing a slight but noticeable variation in the waves received by the radar. These reflections are used to identify whether there is a masked target or not. This goal could be achieved using time frequency analysis to construct images (i.e., spectrograms) that feed CNNs, addressing a two-class image classification problem (one target vs. two targets).

4.1. Time Frequency Analysis

Spectrograms are a popular signal processing tool used to reveal the instantaneous spectral contents of the time-domain signal. They also show the variations in the spectral content over time. A spectrogram is obtained by applying the squared magnitude of the STFT computed over a discrete input signal. The STFT can be formalized as:

$$STFT\{x[n]\} = X(m,k) = \sum_{n=-\infty}^{\infty} x[n]w[n-m]e^{-j2\pi kn/N} \quad (1)$$

where $x[n]$ is the discrete signal, $w[n]$ is the discrete window function, which is non-zero in $[0 \ldots N]$ and zero elsewhere, N is the number of samples in the window, and k is the discrete frequency. The window's location is indicated by the index m. The spectrogram can be generated by continuously computing the STFT with increasing m by a step size Δm. The step size Δm can be used to achieve an overlap between two consecutive analysis windows, resulting in a smoother time dimension output. Eventually, to use the computationally quicker Fast Fourier Transform (FFT), a power of 2 must be selected for N, or N can be zero padded to a power of 2. As a rule of thumb, a large window size indicates a high resolution in the time domain and low resolution in the frequency and vice versa.

4.2. Deep Neural Network Models

To address the shadowing effect problem in its most simplified form, only two classes were considered in this study (one target and two targets). To address the two-class classification problem, we employed CNNs trained over the spectrogram images. The CNNs proved to be very efficient in image classification. In particular, MobileNet models are suitable for deployment on embedded systems since they achieve similar accuracies in the object classification problem, while requiring less parameters than ResNets and VGGs. The peculiarity of the MobileNet models is the adoption of the depth-wise separable convolution [46], i.e., the standard convolution operator is replaced by two separate layers: the first layer involves one convolutional filter per input channel, while the second is a point-wise convolution. For an input of size $H \times W \times D$, and a 2D convolutional layer presenting N_k kernels of size $K \times K$, the computational cost C_{SC} of the standard convolution is:

$$C_{SC} = H \times W \times D \times N_k \times K \times K \quad (2)$$

while, using the depth-wise separable convolution, the cost C_{DSC} is:

$$C_{DSC} = H \times W \times D \times (K^2 + N_k) \quad (3)$$

which is significantly smaller than (2).

Table 2 shows a comparison of some state-of-the-art MobileNet models (i.e., V2 and Small V3) with ResNet50 and VGG19 networks [47], all trained on the Imagenet dataset. The first column reports the models, the second represents the number of parameters, the third shows the generalization accuracy on the Imagenet dataset, the fourth displays the model sizes in megabytes, while the last column presents the average inference time of the models running on GPU Tesla A100. The table demonstrates that the MobileNet models

can achieve similar generalization performance, employing few parameters with respect to more complex models.

Table 2. Sample of the available models.

Model	Num of Params (Million)	Top Accuracy (%)	Size (MB)	Inference Time (ms) on GPU
ResNet50	25.6	74.9	98	4.55
VGG19	143.6	71.3	549	4.38
MobileNet_V2	3.53	71.3	14	3.83
Small MobileNet_V3	2.0	73.8	12	3.57

Following the results of Table 2, four different MobileNet-based architectures were compared. The four models were pre-trained on the Imagenet dataset; thus, the weights and biases were statically loaded, before eventually fine-tuning the last trainable dense layer using the collected dataset. Hence, the convolutional layers of the MobileNet models provided the filters, learned on the Imagenet dataset, to process the input image. Eventually, the features extracted by the convolutional layers were fed to the dense layer, which classified the data among the two possible classes (one target vs. two targets). The data collection procedure is presented in Section 5.1.

5. Experimental Setup

Four persons were involved in a series of experiments with the aim of collecting data to validate the proposed solution. In the following section, the data retrieval process is described. In addition, the spectrogram hyperparameter selection is explained. Finally, the CNN training phase is described. A block diagram of the proposed system is illustrated in Figure 3.

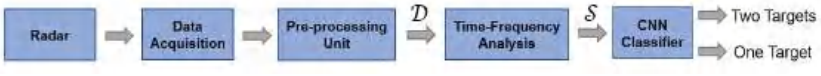


Figure 3. Block diagram of the proposed system.

5.1. Data Acquisition

In order to overcome the possible problem of the multi-path effect, a clutter removal technique proposed in [27] was used to remove the environmental clutter (i.e., the potential ghost effect) from the source.

Two sets of experiments were carried out for this study. Measurements took place in a thirty-meter-long and three-meter-wide corridor. The corridor environment was chosen because it maximized the clutter, thus making it harder for the radar to detect the shadowed target. The goal of the experiments was to detect all human targets standing in range of the radar. The radar was placed one and a half meters from the ground. Figure 4 schematizes the experimental environment.

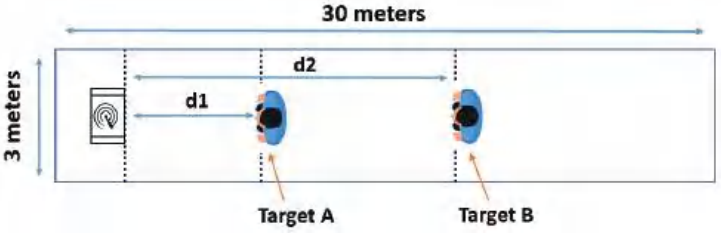


Figure 4. Illustration of the corridor data collection environment.

As illustrated in Figure 3, data are collected and pre-processed to reach a spectrogram format (i.e., image). These spectrograms are then fed to the CNN classifier. Figure 5 shows the data processing pipeline from the raw radar outcome towards the spectrogram format. The data corresponding to chirps are stored as the rows of a matrix of dimension $Nc \times Ns$ (i.e., Nc is the number of chirps and Ns is the number of samples of each chirp). To convert the data type, an Analog to Digital Converter (ADC) was used. Range FFT is then applied on each row, which results in a range representation. Multiple slices ($Slices = 50$ in this study) of this matrix are then collected to form a tensor ($Nc \times Range \times Slices$). The slices are collected consecutively: as soon as the n-th slice is collected, the radar immediately starts to collect the slice n-th + 1. Finally, STFT is applied on this 3D tensor to obtain the spectrogram.

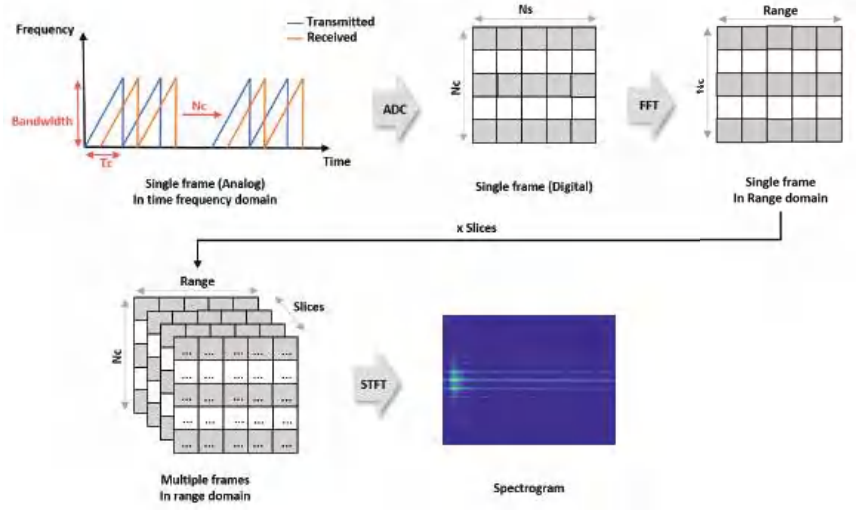


Figure 5. Data processing pipeline.

For the first experiment, illustrated in Figure 1a, a single target (target A) stood in range of the radar. The target was standing in different positions gradually through all reference positions $d1$ along an eleven-meter range. Four different human targets were involved in this experiment to increase the data diversity. One hundred and fifty measurements were collected for each target. Therefore, for the first experiment, six hundred measurements were collected. This dataset is called the One Target (OT) class.

For the second experiment, illustrated in Figure 1c, two targets, A and B, stood in range of the radar. Target A, who was closer to the radar antennas, was standing gradually through all reference positions $d1$ along the eleven-meter range in front of the antennas, and target B, who was farther away from the antennas, stood in the shadowing region of target A (setup is illustrated in Figure 1c), two meters behind him. Three persons were involved in this experiment, exchanging their mutual positions. Six hundred measurements were collected during this experiment; this dataset is called the Two Targets (TT) class. To sum up, the complete collected dataset consists of one thousand two hundred samples, divided in half among the two classes. The extracted dataset is formalized in:

$$\mathcal{D} = \{(\mathcal{X}, y)_i;\ \mathcal{X}_i \in \mathbb{R}^{Nc \times Ns \times Slices},\ y_i \in \{OT, TT\}; i = 1, \ldots, Z = 1200\} \tag{4}$$

where $Nc = 32$, $Ns = 128$, and $Slices = 50$.

Table 3 summarizes the data collection setup. The first column represents the class (One Target vs. Two Targets). The second and the third columns present the distances from each target to the radar (i.e., target A and target B, respectively). The last column displays the number of measurements acquired for each combination of the targets involved in

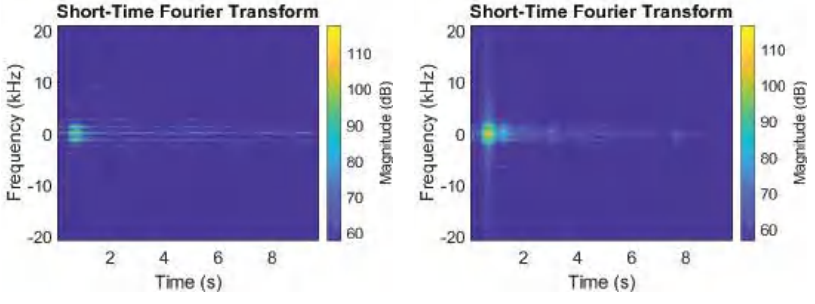
the experiments. In the case of the One Target class, 4 persons were involved (i.e., four combinations for each $d1$ distance); thus, there were 30 acquisitions per combination. For the Two Targets class, measurements were obtained on 3 persons exchanging their mutual position (i.e., 6 possible combinations for each pair $[d1, d2]$), hence leading to 20 acquisitions per combination.

Table 3. Data collection setup.

Class	Distance of Target A ($d1$) [m]	Distance of Target B ($d2$) [m]	Num of Meas. per Comb.
One Target	3	-	30
	5	-	30
	7	-	30
	9	-	30
	11	-	30
Two Targets	3	5	20
	5	7	20
	7	9	20
	9	11	20
	11	13	20

5.2. Spectrogram

According to Section 4.1, a time-frequency analysis was carried out on $\mathcal{D}$ to extract spectrograms in order to feed CNNs. To obtain a continuous spectrogram, a large window size was chosen with N = 2048, with a 50% overlap (Δm = 1024). Two samples of the obtained results are illustrated in Figure 6.



(**a**) (**b**)

Figure 6. Spectrogram examples. (**a**) One target. (**b**) Two targets.

The spectrograms of the 1200 collected samples were generated. The original dimensions of each spectrogram were (875, 656, 3); each was down-sampled and zero padded to fit the input size of our CNN, with the dimensions (224, 224, 3). The dataset containing spectrograms can be formalized as:

$$\mathcal{S} = \{(\hat{\mathcal{X}}, y)_i; \hat{\mathcal{X}}_i \in \mathbb{N}^{224\times224\times3},\ y_i \in \{OT, TT\}, i = 1, \ldots, Z = 1200\}. \quad (5)$$

Figure 6a shows an example of the generated spectrogram for the One Target class as illustrated in Figure 1a. In this example, the target was standing five meters away from the radar. Figure 6b shows an example for the Two Targets class as illustrated in Figure 1c. In this example, target A was five meters away from the radar while target B was two meters behind target A, so seven meters away from the radar. If inspected carefully, a difference is visible in Figure 6a,b; this difference represents the passive electromagnetic waves reflected from target B and received by the radar antenna.

5.3. Training

The authors adopted the four most common implementations of the MobileNet architectures. The number of neurons in the trainable dense layer was set to 128 for each network, using the ReLU as a non-linear activation function. Moreover, the Stratified K-Fold technique was adopted to ensure fair results. Stratified sampling consists of splitting the data of the original labeled dataset (i.e., the population) into subsets, having the same proportion of data as in the population. The subgroups are called 'strata'. Thus, adopting the stratified method in cross-validation guarantees that the training and test sets contain the same proportion of labeled dataset in each fold, leading to a close approximation of the generalization error on the test set. In each of the 'K' iterations of the K-fold cross-validation technique, where the data have been split into 'K' groups, one portion is used as the test set, while the remaining portions are used for training. In the current situation, 'K' was chosen to be equal to five. Therefore, five folds were generated, and results will be presented in the next section as the average of the five results from each of the folds. In this way, 80% of the data have been used for training and 20% for testing in each iteration run. Actually, the test data were split into two parts (validation and test sets) having the same number of samples. An early stopping criterion was implemented during training over the validation set, fixing the patience parameter to 10. All the results have been averaged over the 5 folds. The Adam optimizer function was used with a learning rate of 1/2000. Regarding the loss function, categorical cross entropy was used. Models were trained for one hundred epochs for each split.

6. Experimental Results and Discussion

The results achieved using the proposed approach are presented in Table 4. The first column provides the four adopted model architectures, the second shows the number of parameters, the third column reports the average accuracy computed on the test set of the five folds and the standard deviation for each model, the fourth column depicts the average inference time of the model running on a RTX-2080Ti GPU, while the last column presents the saved model sizes.

Table 4. Results over the four different models.

Model	Num of Params (Million)	Average Test Acc (%) $\pm$ STD	Inference Time (ms) on GPU	Size (MB)
MobileNet_V2	2.3	81.5 ± 4.36	2.35	7.2
MobileNet_V3 Large	3.2	92.2 ± 2.86	2.23	18.2
MobileNet_V3 Small	1.6	90.9 ± 1.4	1.91	6.8
MobileNet_V3 Small Minimalistic	1.06	88.7 ± 2.39	1.64	5.0

The table shows that all the MobileNet_V3-based networks generally perform better than MobileNet_V2. This could be explained by the introduction of the hard-swish activation function and the implementation of a squeeze-and-excitation module [46]. Among the three MobileNet_V3 versions (Large, Small, and Small Minimalistic), the testing accuracy results are directly proportional to the number of parameters used in the architecture: the higher the number of parameters, the higher testing accuracy is achieved.

A compromise should be taken when choosing the model. This compromise would be highly dependent on the application scenario. If the application scenario is critical and the accuracy is the main interest, Large MobileNet_V3 would be chosen. If the goal is to deploy on the edge, then memory and inference time would be the main goal, and Small Minimalistic MobileNet_V3 would be chosen. Usually, the main interest in using a low-cost radar is the possibility of edge deployment, and the main constraint of edge deployment is the

number of parameters, i.e., the model size. Therefore, the Small Minimalistic MobileNet_V3 best suits the proposed use-case.

Under the proposed circumstances, where either one or two targets are in the detection range of the radar, the model choice (i.e., number of parameters, architecture, etc.) affects the performance of the proposed algorithm. In addition, the radar parameters and hardware specifications (i.e., number of chirps, memory capacity, etc.) influence the performance of the algorithm; these parameters were chosen according to [44].

As introduced in Section 2, the authors in [2] proposed an algorithm based on the wavelet entropy for shadow effect removal for human targets using UWB radars. This method proved to be effective in detecting two stationary human targets despite one person being in the shadowing region of the other. Hence, static filters (i.e., wavelet) were used. For each new possible deployment environment, an on-site adjustment is required: the number of filter levels and the wavelet function need to be tuned to accurately fit the application scenario. Therefore, the solution is not easily scalable because it needs expert intervention whenever a new context occurs. On the other hand, our proposal uses filters (i.e., weights of the convolutional layers) learned on a massive dataset (i.e., Imagenet dataset). This guarantees a high level of scalability and ease of deployment for different environments. In addition, it is not necessary to retrain the filters for new problems: only one dense layer needs to be fine-tuned for the incoming dataset, preserving the same structure of the pre-trained architecture, without requiring any expert intervention.

7. Conclusions and Future Works

In the case of multi-target detection using an FMCW radar, the target closest to the radar antennas partially reflects the energy of the electromagnetic wave, and the person farther from the radar antennas is not detected continuously, especially when in the shadowing region of the closest person. In this paper, a novel solution for the radar shadowing effect has been proposed. The solution is based on a CNN model that classifies the spectrogram images, obtained after a time-frequency analysis of the radar data, among one of two classes: One Target vs. Two Targets. The model is based on MobileNet and is loaded with the Imagenet weights. The best solution in terms of testing accuracy achieved 92.2% with a standard deviation of 2.86%, while the lightest (i.e., 1.06 million parameters) model attained 88.7% with a standard deviation of 2.39% over five splits of input data. The latter model uses 1.06 million parameters only and has a size of 5 MB. The inference time using a GPU is 1.64 ms. In future research, the authors plan to deploy the proposed solution on a Raspberry Pi and test the model in a real scenario. In addition, the distance between the visible target and the masked target should be assessed using a regression model. The proposed solution could be extended to different types of targets (e.g., cars, robots, pedestrians, etc.). This novel solution uses a supervised learning method; in other words, it already knows all the possible classes (One Target or Two Targets). If the situation of multiple shadowed targets needs to be addressed, it is theoretically possible by collecting enough data for every possible class. This method might not be practical because the number of classes could not be predicted beforehand. Therefore, the recommended procedure would be to shift the problem into an unsupervised problem. We are also considering an extension of this proposed approach; the goal is to detect and track two or more moving targets, with different inner distances, in a cluttered environment.

Author Contributions: Conceptualization, A.M., C.G. and A.R.; methodology, A.M. and C.G.; software, A.M. and A.R.; validation, A.M., C.G. and A.R.; formal analysis, A.M., C.G. and A.R.; investigation, A.M.; resources, M.V.; data curation, A.M.; writing—original draft preparation, A.M.; writing—review and editing, A.M., C.G., A.R. and M.V.; visualization, A.M. and A.R.; supervision, C.G. and M.V.; project administration, M.V.; funding acquisition, M.V. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: All data and code are available at: https://github.com/AmmarMohanna/ShadowingEffect (accessed on 23 December 2021).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Thi Phuoc Van, N.; Tang, L.; Demir, V.; Hasan, S.F.; Duc Minh, N.; Mukhopadhyay, S. Microwave radar sensing systems for search and rescue purposes. *Sensors* **2019**, *19*, 2879. [CrossRef] [PubMed]
2. Xue, H.; Liu, M.; Zhang, Y.; Liang, F.; Qi, F.; Chen, F.; Lv, H.; Wang, J. An Algorithm based wavelet entropy for shadowing effect of human detection using ultra-wideband bio-radar. *Sensors* **2017**, *17*, 2255. [CrossRef] [PubMed]
3. Huang, K.; Zhong, J.; Zhu, J.; Zhang, X.; Zhao, F.; Xie, H.; Gu, F.; Zhou, B.; Wu, M. The method of forest fires recognition by using Doppler weather radar. In Proceedings of the 8Symposium on Fire and Forest Meteorology, Kalispell, MT, USA, 13–15 October 2007; pp. 1–7.
4. Capria, A.; Giusti, E.; Moscardini, C.; Conti, M.; Petri, D.; Martorella, M.; Berizzi, F. Multifunction imaging passive radar for harbour protection and navigation safety. *IEEE Aerosp. Electron. Syst. Mag.* **2017**, *32*, 30–38. [CrossRef]
5. Lemaitre, F.; Poussieres, J.C. Method and System for Sensing and Locating a Person, eg under an Avalanche. US Patent 6,031,482, 29 February 2000.
6. Rizik, A.; Randazzo, A.; Vio, R.; Delucchi, A.; Chible, H.; Caviglia, D.D. Low-Cost FMCW Radar Human-Vehicle Classification Based on Transfer Learning. In Proceedings of the 2020 32nd International Conference on Microelectronics (ICM), Aqaba, Jordan, 14–17 December 2020; pp. 1–4.
7. Kocur, D.; Rovňáková, J.; Urdzík, D. Experimental analyses of mutual shadowing effect for multiple target tracking by UWB radar. In Proceedings of the 2011 IEEE 7th International Symposium on Intelligent Signal Processing, Floriana, Malta, 19–21 September 2011; pp. 1–4.
8. Kocur, D.; Rovňáková, J.; Urdzík, D. Mutual shadowing effect of people tracked by the short-range UWB radar. In Proceedings of the 2011 34th International Conference on Telecommunications and Signal Processing (TSP), Budapest, Hungary, 18–20 August 2011; pp. 302–306.
9. Maaref, N.; Millot, P.; Pichot, C.; Picon, O. FMCW ultra-wideband radar for through-the-wall detection of human beings. In Proceedings of the 2009 International Radar Conference" Surveillance for a Safer World"(RADAR 2009), Bordeaux, France, 12–16 October 2009; pp. 1–5.
10. Mitomo, T.; Ono, N.; Hoshino, H.; Yoshihara, Y.; Watanabe, O.; Seto, I. A 77 GHz 90 nm CMOS transceiver for FMCW radar applications. *IEEE J. Solid-State Circuits* **2010**, *45*, 928–937. [CrossRef]
11. Lin Jr, J.; Li, Y.P.; Hsu, W.C.; Lee, T.S. Design of an FMCW radar baseband signal processing system for automotive application. *SpringerPlus* **2016**, *5*, 1–16. [CrossRef]
12. Zhou, H.; Wen, B.; Ma, Z.; Wu, S. Range/Doppler ambiguity elimination in high-frequency chirp radars. *IEE Proc.-Radar Sonar Navig.* **2006**, *153*, 467–472. [CrossRef]
13. Kulpa, K.; Czekała, Z. Masking effect and its removal in PCL radar. *IEE Proc.-Radar Sonar Navig.* **2005**, *152*, 174–178. [CrossRef]
14. Urdzík, D.; Zetík, R.; Kocur, D.; Rovnáková, J. Shadowing effect investigation for the purposes of person detection and tracking by UWB radars. In Proceedings of the 2012 The 7th German Microwave Conference, Ilmenau, Germany, 12–14 March 2012; pp. 1–4.
15. Xue, H.; Liu, M.; Lv, H.; Jiao, T.; Li, Z.; Qi, F.; Wang, P.; Wang, J.; Zhang, Y. A dynamic clutter interference suppression method for multiple static human targets detection using ultra-wideband radar. *Microw. Opt. Technol. Lett.* **2019**, *61*, 2854–2865. [CrossRef]
16. Claudepierre, L.; Douvenot, R.; Chabory, A.; Morlaas, C. Assessment of the Shadowing Effect between Windturbines at VOR and Radar frequencies. *Forum Electromagn. Res. Methods Appl. Technol. (FERMAT)* **2016**, *13*, 1464–1476.
17. Perez Fontan, F.; Espiñeira, P. *Shadowing Effects*; John Wiley & Sons: Hoboken, NJ, USA, 2008; pp. 29–60. [CrossRef]
18. Zetik, R.; Jovanoska, S.; Thomä, R. Simple Method for Localisation of Multiple Tag-Free Targets Using UWB Sensor Network. In Proceedings of the 2011 IEEE International Conference on Ultra-Wideband (ICUWB), Bologna, Italy, 14–16 September 2011; pp. 268–272. [CrossRef]
19. Radar Shadow. In *Dictionary Geotechnical Engineering/Wörterbuch GeoTechnik: English-German/Englisch-Deutsch*; Springer: Berlin/Heidelberg, Germany, 2014; p. 1069. [CrossRef]
20. O'Mahony, N.; Campbell, S.; Carvalho, A.; Harapanahalli, S.; Hernandez, G.V.; Krpalkova, L.; Riordan, D.; Walsh, J. Deep learning vs. traditional computer vision. In Proceedings of the Science and Information Conference, Las Vegas, NV, USA, 25–26 April 2019; pp. 128–144.
21. Heuel, S.; Rohling, H. Pedestrian classification in automotive radar systems. In Proceedings of the 2012 13th International RADAR Symposium, Warsaw, Poland, 23–25 May 2012; pp. 39–44.
22. Mukhtar, A.; Xia, L.; Tang, T.B. Vehicle detection techniques for collision avoidance systems: A review. *IEEE Trans. Intell. Transp. Syst.* **2015**, *16*, 2318–2338. [CrossRef]
23. Zhang, Z.; Tian, Z.; Zhou, M. Latern: Dynamic continuous hand gesture recognition using FMCW radar sensor. *IEEE Sens. J.* **2018**, *18*, 3278–3289. [CrossRef]

24. Hussain, M.; Bird, J.J.; Faria, D.R. A study on cnn transfer learning for image classification. In Proceedings of the UK Workshop on computational Intelligence, Nottingham, UK, 5–7 September 2018; pp. 191–202.
25. Lee, H.; Kwon, H. Going deeper with contextual CNN for hyperspectral image classification. *IEEE Trans. Image Process.* **2017**, *26*, 4843–4855. [CrossRef] [PubMed]
26. Huh, M.; Agrawal, P.; Efros, A.A. What makes ImageNet good for transfer learning? *arXiv* **2016**, arXiv:1608.08614.
27. Rizik, A.; Tavanti, E.; Chible, H.; Caviglia, D.D.; Randazzo, A. Cost-Efficient FMCW Radar for Multi-Target Classification in Security Gate Monitoring. *IEEE Sens. J.* **2021**, *21*, 20447–20461. [CrossRef]
28. Sacco, G.; Piuzzi, E.; Pittella, E.; Pisa, S. An FMCW radar for localization and vital signs measurement for different chest orientations. *Sensors* **2020**, *20*, 3489. [CrossRef] [PubMed]
29. Peng, Z.; Ran, L.; Li, C. A *K*-Band Portable FMCW Radar With Beamforming Array for Short-Range Localization and Vital-Doppler Targets Discrimination. *IEEE Trans. Microw. Theory Tech.* **2017**, *65*, 3443–3452. [CrossRef]
30. Han, K.; Hong, S. Vocal Signal Detection and Speaking-Human Localization With MIMO FMCW Radar. *IEEE Trans. Microw. Theory Tech.* **2021**, *69*, 4791–4802. [CrossRef]
31. Cong, J.; Wang, X.; Lan, X.; Huang, M.; Wan, L. Fast Target Localization Method for FMCW MIMO Radar via VDSR Neural Network. *Remote Sens.* **2021**, *13*, 1956. [CrossRef]
32. Stephan, M.; Hazra, S.; Santra, A.; Weigel, R.; Fischer, G. People Counting Solution Using an FMCW Radar with Knowledge Distillation From Camera Data. In Proceedings of the 2021 IEEE Sensors, Sydney, Australia, 31 October–3 November 2021.
33. Will, C.; Vaishnav, P.; Chakraborty, A.; Santra, A. Human Target Detection, Tracking, and Classification Using 24-GHz FMCW Radar. *IEEE Sens. J.* **2019**, *19*, 7283–7299. [CrossRef]
34. Vaishnav, P.; Santra, A. Continuous Human Activity Classification With Unscented Kalman Filter Tracking Using FMCW Radar. *IEEE Sens. Lett.* **2020**, *4*, 1–4. [CrossRef]
35. Wang, G.; Gu, C.; Inoue, T.; Li, C. A hybrid FMCW-interferometry radar for indoor precise positioning and versatile life activity monitoring. *IEEE Trans. Microw. Theory Tech.* **2014**, *62*, 2812–2822. [CrossRef]
36. Angelov, A.; Robertson, A.; Murray-Smith, R.; Fioranelli, F. Practical classification of different moving targets using automotive radar and deep neural networks. *IET Radar Sonar Navig.* **2018**, *12*, 1082–1089. [CrossRef]
37. Abdulatif, S.; Wei, Q.; Aziz, F.; Kleiner, B.; Schneider, U. Micro-doppler based human-robot classification using ensemble and deep learning approaches. In Proceedings of the 2018 IEEE Radar Conference (RadarConf18), Oklahoma City, OK, USA, 23–27 April 2018; pp. 1043–1048.
38. Khanna, R.; Oh, D.; Kim, Y. Through-wall remote human voice recognition using doppler radar with transfer learning. *IEEE Sens. J.* **2019**, *19*, 4571–4576. [CrossRef]
39. Bhattacharya, A.; Vaughan, R. Deep learning radar design for breathing and fall detection. *IEEE Sens. J.* **2020**, *20*, 5072–5085. [CrossRef]
40. Huang, X.; Ding, J.; Liang, D.; Wen, L. Multi-person recognition using separated micro-Doppler signatures. *IEEE Sens. J.* **2020**, *20*, 6605–6611. [CrossRef]
41. Kim, S.; Lee, K.; Doo, S.; Shim, B. Automotive radar signal classification using bypass recurrent convolutional networks. In Proceedings of the 2019 IEEE/CIC International Conference on Communications in China (ICCC), Changchun, China, 11–13 August 2019; pp. 798–803.
42. Kim, Y.; Alnujaim, I.; You, S.; Jeong, B.J. Human detection based on time-varying signature on range-Doppler diagram using deep neural networks. *IEEE Geosci. Remote Sens. Lett.* **2020**, *18*, 426–430. [CrossRef]
43. Richards, M.A. *Fundamentals of Radar Signal Processing*, 2nd ed.; McGraw-Hill: New York, NY, USA, 2014.
44. Infenion POSITION2GO Board. Available online: https://www.infineon.com/cms/en/product/evaluation-boards/demo-position2go/ (accessed on 17 December 2021).
45. Nicolaescu, L.; Oroian, T. Radar cross section. In Proceedings of the 5th International Conference on Telecommunications in Modern Satellite, Cable and Broadcasting Service. TELSIKS 2001. Proceedings of Papers (Cat. No. 01EX517), Nis, Yugoslavia, 19–21 September 2001; Volume 1, pp. 65–68.
46. Howard, A.; Sandler, M.; Chu, G.; Chen, L.C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; et al. Searching for mobilenetv3. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27–28 October 2019; pp. 1314–1324.
47. Deep Neural Networks. Available online: https:/keras.io/api/applications/ (accessed on 17 December 2021).

sensors

MDPI

Article

Towards Semantic Photogrammetry: Generating Semantically Rich Point Clouds from Architectural Close-Range Photogrammetry

Arnadi Murtiyoso [1,*], Eugenio Pellis [1,2], Pierre Grussenmeyer [1], Tania Landes [1] and Andrea Masiero [2]

[1] Université de Strasbourg, INSA Strasbourg, CNRS, ICube Laboratory UMR 7357, 67084 Strasbourg, France; eugenio.pellis@insa-strasbourg.fr (E.P.); pierre.grussenmeyer@insa-strasbourg.fr (P.G.); tania.landes@insa-strasbourg.fr (T.L.)
[2] Department of Civil and Environmental Engineering, University of Florence, 50121 Florence, Italy; andrea.masiero@unifi.it
* Correspondence: arnadi.murtiyoso@insa-strasbourg.fr

Abstract: Developments in the field of artificial intelligence have made great strides in the field of automatic semantic segmentation, both in the 2D (image) and 3D spaces. Within the context of 3D recording technology it has also seen application in several areas, most notably in creating semantically rich point clouds which is usually performed manually. In this paper, we propose the introduction of deep learning-based semantic image segmentation into the photogrammetric 3D reconstruction and classification workflow. The main objective is to be able to introduce semantic classification at the beginning of the classical photogrammetric workflow in order to automatically create classified dense point clouds by the end of the said workflow. In this regard, automatic image masking depending on pre-determined classes were performed using a previously trained neural network. The image masks were then employed during dense image matching in order to constraint the process into the respective classes, thus automatically creating semantically classified point clouds as the final output. Results show that the developed method is promising, with automation of the whole process feasible from input (images) to output (labelled point clouds). Quantitative assessment gave good results for specific classes e.g., building facades and windows, with IoU scores of 0.79 and 0.77 respectively.

Keywords: photogrammetry; semantic segmentation; deep learning; automation; dense matching; point cloud; classification

Citation: Murtiyoso, A.; Pellis, E.; Grussenmeyer, P.; Landes, T.; Masiero, A. Towards Semantic Photogrammetry: Generating Semantically Rich Point Clouds from Architectural Close-Range Photogrammetry. *Sensors* **2022**, *22*, 966. https://doi.org/10.3390/s22030966

Academic Editor: Manuel José Cabral dos Santos Reis

Received: 22 December 2021
Accepted: 24 January 2022
Published: 26 January 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

The use of artificial intelligence has seen an exponential increase in recent decades, aided by developments in computing power. Within the field of 3D surveying, such methods have been used to perform tasks such as semantic segmentation [1]. This process of automatically attributing semantic information into the otherwise geometric information stored in spatial 3D data (e.g., point clouds) is a major step in accelerating the surveying process. Semantic annotation also enables easier modelling and predictions using the available spatial data. Since spatial data annotation is traditionally performed manually, the use of artificial intelligence such as the deep learning approach has the potential to reduce both the significant time and resources required. However, current research mostly focuses on the application of deep learning on the 3D space. In this paper, we propose a method to introduce deep learning semantic segmentation into the classical photogrammetric workflow in order to benefit from some of photogrammetry's rigorous advantages, e.g., block bundle adjustment.

Photogrammetry as a discipline has a long history of use in the field of surveying. Starting with primarily small scale aerial use [2,3], the use of terrestrial images has also been

effectively applied for many applications in larger scale architectural survey [4]. In the past few decades, photogrammetry has also seen significant developments in both its theoretical and technical aspects. Major strides were made in subjects such as analytical photogrammetry [5,6], automatic image matching [7] and bundle adjustment [8]. Furthermore, the parallel development of the computer vision domain such as structure-from-motion [9,10], automatic feature extraction [11] and dense image matching [12,13] have helped in solving some traditional photogrammetric bottlenecks.

Recently, photogrammetry has seen a major democratisation by the advent of low-cost sensors [14,15], more powerful computing capacity [16,17] as well as availability of drones in the public market [18,19]. The latter greatly facilitated close-range photogrammetry as it enabled an aerial point of view which was previously a major constraint in data acquisition. Indeed, 3D reconstruction in general is slowly becoming the standard in mapping, replacing traditional 2D methods and products.

As 3D data both in the form of point clouds and meshes became more and more common as a geospatial product, a new research question rose [20–22]. All the major developments in 3D technology, including in photogrammetry, have focused on the geometrical reconstruction of existing objects. This is as far as surveyors are concerned, the main objective of the mapping activity. However, in order to be truly useful, specific tangible meanings must be attached to these geometric elements, i.e., annotating them with relevant semantic information [23,24]. Semantic information or attributes will give these 3D data richness and opens the possibility for various spatial analysis and modelling. Within the traditional 2D mapping, one of the most known framework for this mixture of geometry and semantic attribute is the Geographical Information System (GIS) [25,26]. Its extension into the 3D space can be seen in, for example, 3D GIS for smaller scale scenes [27,28] or Building Information Models (BIM) in larger scenes [29].

The problem of attribute annotation into geometric data was mostly addressed manually [30]. Indeed, historically, GIS layers were physical maps which were digitised and vectorised. When required, semantic annotation was performed at the same time by the operator. This method continued on with the arrival of BIM, where most users would create parametric 3D models from point clouds and attach attributes to them manually [31]. Attempts at automation can be seen in the current literature [32–34], and remains a major research question today. In practice, this process of data labelling translates into 3D data classification in geomatics or semantic segmentation in AI parlance [35].

Various methods for semantic segmentation have been proposed in the literature, with some review papers highlighting this fact [1,36]. Techniques based on heuristic information (e.g., geometric rules or tendencies for certain object classes) present a fast and generally precise results [37,38]. These algorithmic approaches are however often non-flexible and problems may occur when encountering complex cases, e.g., historical buildings or traditional architecture. More recent research into artificial intelligence, coupled with more powerful computing power, has opened the possibility to the application of machine learning (ML) to fulfil this purpose. Deep learning (DL), a subset of machine learning, has also seen major strides in performing semantic segmentation on 2D images [39]. Promising results can also be observed in 3D semantic segmentation, both indirectly [40] and directly [41]. It is worth noting that in DL-based solutions, a major bottleneck is the availability of labelled data for training. While over the past few years immense amounts of labelled images have become available in ML circles, 3D training data remain scarce [23] due to the higher complexity in manually labelling them.

While these solutions show promise, most are concerned with the segmentation of the point clouds or 3D meshes which are the product of 3D reconstruction techniques. In this regard, the process of creating these 3D inputs for classification matters a little as they may come from either photogrammetry, lidar, a combination of both, or other 3D sensors. Few studies (e.g., [42]) had addressed the potential of involving this semantic segmentation process directly within the photogrammetric workflow. Hypothetically speaking, such integration may benefit from several advantages. For instance, the application of semantic

segmentation on the input images may take advantage from the vastly more available 2D training datasets for DL. Furthermore, when applied on the dense image matching step, other mathematical conditions such as the epipolar constraint and error minimisation via bundle adjustment may help improve the results.

The idea of using AI to support photogrammetry has been previously explored by several studies. In [43], the authors reported several applications including aiding feature detection for the orientation of images with significant differences in scale and viewing angle. A study by [44] presented a similar approach to the one presented in this paper, i.e., using AI to generate classified point clouds from 2D images via photogrammetry although the authors did not prove numerical assessments of their method. The same authors also briefly reported their experiments in using masks to automatically clean dense point clouds, as well as to transfer 2D labels to 3D point cloud [42]. A similar approach using masks was reported in [45] for agricultural applications. Furthermore, the authors in [46] attempted to implement masks during dense matching in order to clean point cloud noise.

The aim of the study is therefore to propose a method which may benefit from both the abundance of 2D training data for DL purposes and the rigour of photogrammetric computations to create a faster and more precise approach to 3D point cloud semantic segmentation. To this end, in this paper we propose a practical and fully automatic workflow from images to classified 3D point clouds. As a proof of concept, the workflow was implemented within the context of close-range photogrammetry for architectural surveying purposes, e.g., building facade modelling. The input of the said workflow is 2D images acquired according to photogrammetric principles and oriented using a classical image matching and bundle adjustment. A DL-based neural network trained on a database of rectified building facade images was then used to perform semantic segmentation on the oriented images. The segmented images were thereafter used in dense image matching to generate semantically rich and classified 3D point cloud. The workflow was implemented using the open source photogrammetric suite Apero-Micmac [47], with additional coding in Matlab. Additional comparison was also implemented in the commercial software Agisoft Metashape. As the readers shall observe in this paper, the proposed method may be adapted into other photogrammetric situations, e.g., aerial mapping or heritage documentation by simply adjusting the applied neural network. As far as the paper's structure is concerned, the next section will explore some work related to the main idea presented in this paper. Section 3 will thereafter contain the main description of the proposed method, with experimental results and discussions presented in Sections 4 and 5, respectively. Finally, Section 6 shall put forward arguments to the potential of the proposed method, its drawbacks, and some ideas for improvements.

2. Literature Review

In the following exposition, a summary of existing literature on the subject of photogrammetry, AI semantic segmentation and their interaction shall be addressed. First, an overview of the photogrammetric workflow will be described. Arguments will also be put forward on the choice of software solution used in this study. Subsequently, a short description of deep learning methods for 3D semantic segmentation will be given. Finally, several existing solutions to the problem of projecting 2D image labels into the 3D space will be described.

2.1. Notions on Photogrammetry

Photogrammetry as a mapping technique attempts to convert 2D images into 3D coordinates using stereo vision principles [48]. While such concepts were first implemented in an empiric manner, as is the case with analogue photogrammetry, mathematical relations were soon developed to enable an analytic approach to the problem of 3D reconstruction. Notably, the collinearity and coplanarity conditions played an important role in establishing a relation between the 2D and 3D space [6]. For most of its history and even today, photogrammetry remains very focused on the problem of precision. This is in line with the

original objective of photogrammetry as a remote sensing mapping tool. However, almost in parallel developments in the computer vision domain saw significant leaps as evidenced by the popularity of Structure-from-Motion as a solution to image pose estimation [10]. This progress, in addition to other developments in both imaging sensor and computing technologies, enabled the unprecedented automation of the traditional photogrammetric workflow albeit sometimes at the expense of rigorous quality control [49]. Image matching algorithms further reduce the necessity of manual measurements, e.g., those involving the traditional six Von Grüber points [11,50].

Evolving from previous solutions for the problem of aerotriangulation, i.e., densification of ground controls [50] in analytical photogrammetry, the concept of bundle adjustment refers to the simultaneous computation of image exterior orientation parameters (also referred to as extrinsic parameters in computer vision [10,35]) and point coordinates in the 3D space. It typically involves a non-linear optimisation calculation based on either collinearity or coplanarity equations. This simultaneous "block" adjustment of the whole system provides a rigorous solution for the exterior orientation problem [49]. The bundle adjustment may also include the resolution of camera internal parameters in a process called self or auto-calibration [51]. Furthermore, modern bundle adjustment solutions may include damping techniques (e.g., Levenberg–Marquardt algorithm) to help the classical Gauss–Newton least-squares method in reaching final convergence [8]. This is the case, for example, in the software Apero-Micmac used in this study [47].

Another major breakthrough in the field of photogrammetry was the development of dense image matching. Work on Patch-based Multi View Stereo (PMVS) [9] and Semi-Global Matching [52] may be considered some of the most important developments. Dense image matching is a crucial development for photogrammetry which enables it to generate dense point clouds not unlike those created by lidar. This provides photogrammetry with the tool to compete with lidar systems [53], although in practice they are often complementary, especially in large-scale applications [54].

Various photogrammetric solutions exist in the market today, both of commercial and open source nature. A classical photogrammetric workflow starts with the acquisition of images. Certain rules must be respected in order to guarantee good results from photogrammetry, e.g., enough overlapping between images [55], configuration of image network [49,51,56] but also photographic quality [57]. From a surveying perspective, pre-acquisition steps such as determination of the required Ground Sampling Distance (GSD) [21] and distribution of Ground Control Points (GCP) or scale lines are equally important [18,58]. Image orientation with bundle adjustment is then usually performed before continuing with dense image matching in order to create dense point clouds.

However, these point clouds for the most part represent only the geometric aspect of the object in question. Semantic information is usually imbued by performing point cloud classification [59,60] as a post-processing of the point cloud generation process. Indeed, most studies including those with application of DL involve semantic segmentation on the point cloud [1]. In this paper, DL-based methods are introduced during the photogrammetric process with the final goal of creating a truly semantic photogrammetry workflow.

2.2. DEEP Learning for Semantic Segmentation

Semantic segmentation refers to the process of grouping parts of a data into several subsets that share similar feature characteristics. It can be considered as a fundamental step in the machine automatic comprehension and it is a key topic in a lot of computer vision problems such as scene understanding, autonomous driving, remote sensing, robotic perception, and many others. The continuously increasing number of applications concerning semantic segmentation makes it a very active research field, and different methods and approaches are proposed every year. Image segmentation, or 2D semantic segmentation, involves a pixel-level classification, in which each pixel is associated with a category or a class. Point cloud semantic segmentation is the extension of this task in the 3D space, in which irregular distributed points are used instead of regular distributed pixels in

a 2D image. Point cloud semantic segmentation is usually realised by supervised and unsupervised learning methods, including regular learning and deep learning [61]. In the last five years DL on point clouds has been attracting extensive attention, due to the remarkable results obtained on two-dimensional image processing, in particular after the introduction of Convolutional Neural Networks (CNN). Compared with two-dimensional data, working with 3D point clouds provides an opportunity for a better understanding of spatial and geometrical information, and a better comprehension and characterisation of complex scenarios. However, the use of deep learning on 3D point clouds still faces several significant challenges due to: (i) the unstructured and unordered nature of point clouds, which prevents the use of 2D network architectures, (ii) the large data size, which implies long computing time and (iii) the unavailability of large dedicated dataset for the networks training process. Studies exist which aim to remedy this problem [62].

Despite these challenges, more and more methods are proposed to work with point clouds. In the current literature, semantic segmentation techniques for 3D point cloud can be divided into two groups: (i) projection-based methods and (ii) point-based methods [63].

2.2.1. Projection-Based Methods

The main issue to solve in the problem of point cloud segmentation using standard neural network is its unstructured nature. To address this issue, projection-based techniques first apply a transformation to convert 3D point cloud into data with a regular structure, before subsequently performing semantic segmentation by exploiting the standards models and finally re-projecting the extracted features back to the initial point cloud. Although intermediate representation involves inevitably a spatial and geometrical information loss, the advantage of these methods is the ability to leverage well-established 2D network architectures. According to the type of representation, it is possible to distinguish four categories among these methods:

1. Multiview representation: These methods project firstly the 3D shape or point cloud into multiple images or views, then apply existing models to extract feature from the 2D data. The results obtained on the image representation are compared and analysed, and then re-projected on the 3D shape to obtain the segmentation of the 3D scene. Two of the most popular works are MVCNN [64] which proposed the use of Convolutional Neural Networks (CNN) on multiple perspectives and SnapNet [65] which uses snapshots of the cloud to generate RGB and depth images to address the problem of information loss. These methods ensure excellent image segmentation performance, but the 3D features transposition remains a challenging task, producing large loss of spatial information.
2. Volumetric representation: Volumetric representation consists in the transformation of the unstructured 3D cloud into a regular spatial grid, a process also called voxelisation. The information as distributed on the regular grid is then exploited to train a standard neural network to perform the segmentation task. The most popular architectures are VoxNet [66] which uses CNN to predict classes directly on the spatial grid, OctNET [67] and SEGCloud [68] which introduced the methods of spatial partition such as K-d tree or Octree. These methods require large amounts of computing memory and produce reasonable performance on small point clouds. They are therefore unfortunately still unsuitable for complex scenarios.
3. Spherical representation: This type of representation retains more geometrical and spatial information compared to multiview representation. The most important works in this regard include SqueezeNet [69] and RangeNet++ [70] especially for application on real time lidar data segmentation. However, they still have to face several issues such as discretisation errors and occlusion problems.
4. Lattice representation: Lattice representation converts a point cloud into discrete elements such as sparse permutohedral lattices. This representation can control the sparsity of the extracted features and it reduces memory requirement and compu-

tational cost compared to simple voxelisation. Some of the main studies include SPLATNet [71], LatticeNet [72] and MinkowskiNet [73].

2.2.2. Point-Based Methods

Point-based methods do not introduce any intermediate representation, and they work directly with point clouds. This direct approach leverage on the full use of the characteristics of the raw cloud data and consider all the geometrical and spatial information. They seem the most promising but are still in development and they still have to face several critical issues. Overall, these methods could be divided into four groups:

1. Pointwise methods: The pioneering work for this method is PointNet [74] which learns per-point features using shared Multi-Layer Perceptrons (MLPs) and global features using symmetrical polling functions. Since MLP cannot capture local geometry, a lot of networks based on PointNet have been developed recently. These methods are generally based on neighbouring feature pooling such as PointNet++ [75].
2. Convolution methods: These methods propose an effective convolution operator directly for point clouds. PointCNN [76] is an example of a network based on parametric continuous convolution layers and kernel function as parameterised by MLPs. Another example is ConvPoint [77] which proposed a point-wise convolution operator and convolution weights determined by the Euclidean distances to kernel points.
3. RNN-based methods: Recurrent Neural Network (RNN) are used recently for the segmentation of point clouds, in particular to capture inherent context features. Based on PointNet, they first transform a block of points into multi-scale blocks or grid blocks. Then the features extracted by PointNet are fed into a Recurrent Consolidation Units (RCU) to obtain the output-level context. One of the most popular networks in this regard is 3DCNN-DQN-RNN [78].
4. Graph-based methods: Graph Neural Network (GNN) is a type of network which directly operates on graph structure. Several methods leverage on graphs to capture richer geometrical information, for example DGCNN [79].

Point-based methods seem to be the most promising in the future as evidenced, amongst others, by the great interest it generated in recent research. However, this study will focus more on the deployment of a workflow for dense point cloud semantic segmentation based on two-dimensional data as integrated within the traditional photogrammetric workflow. On one hand, this approach allows us to exploit the tried-and-tested results in 2D image processing while on the other hand it allows the automatic creation of a directly segmented and classified point cloud. In addition, the interaction between point clouds and images could converge in a hybrid point-image method that may improve the performance of both approaches in the future.

2.3. Reprojection of 2D Semantic Segmentation into the 3D Space

In the case of multiview deep learning approaches for 3D semantic segmentation, two main steps may be distinguished: (i) the labelling of the two-dimensional images related to the 3D scene, and (ii) the (re)projection of such labels from the images to the 3D shape or point cloud. Since numerous techniques and methods are already developed for 2D image segmentation with promising results and accuracy [80], the most challenging and critical step in this framework is the reprojection step. This operation introduces inevitably a loss of spatial and geometrical information, and, in many cases, involves a loss of accuracy on the overall performance.

In the last years, several methods have been proposed to address these problems. In [81], the authors proposed a 2D-to-3D based label propagation approach to create 3D training data by utilising existing datasets such as ImageNet and LabelMe. The proposed method consists of two major novel components, Exemplar SVM based label propagation, which effectively addresses the cross-domain issue, and a graphical model based contextual refinement incorporating 3D constraints.

For similar purposes the method developed in [82] propagates object label from 2D image to a sparse point cloud by matching a group of points that corresponds to the area within the 2D bounding box in the image. Furthermore, [42] proposed a semantic photogrammetry workflow similar to the one proposed in this paper, in which the label back-projection is based on the projection matrix which connects the 3D with the 2D space. Using this approach, all of the images contribute to the labelling projection on the cloud with a weighted winner procedure. Although the proposed method is similar, the authors only described their method briefly with few quantitative analysis.

Our previous work described in [40] presented an approach for the segmentation of 3D building facade based on orthophoto and the corresponding depth maps. The XY coordinates of each pixel in the orthophoto was used to determine the corresponding planimetric coordinates of the point in the point cloud and finally a winner-takes-all approach was applied to annotate the 3D points with the respective 2D pixel class.

In [83], the authors proposed an approach for label propagation in RGB-D video sequences, in which each unlabelled frame is segmented using an intermediate 3D point cloud representation obtained from the camera pose and depth information of two keyframes. For similar purposes some studies deal with the 3D to 2D projection as can be seen for example in [84]. In this paper, a CFR model was proposed which is able to transfer the labels from a sparse 3D point cloud to the image pixels by leveraging the calibration and the registration of a camera and laser scanner system, estimated using structure-from-motion. Finally, the authors in [85] developed a method to map the semantic label of 3D point clouds into street view images. The images are over-segmented into super-pixels, and each image plane super-pixel is associated with a collection of labelled 3D points using the generic camera model.

3. Proposed Method

Figure 1 presents a flowchart of the developed workflow. It starts with image acquisition following standard photogrammetric procedure. The acquired images were then processed using the previously trained DL network to semantically segment them according to the predetermined classes. The output of this process is class labels for each pixel for each input image. Using these segmented images, class masks were then generated which was later on used as constraints during the dense image matching process. The final result would be a semantically segmented 3D dense point clouds directly out from the photogrammetric process without need for further labelling or annotation.

In the case of this paper, image acquisition of a building facade was conducted using terrestrial images as a proof of concept for the semantic photogrammetry method. The building used in this case is the main facade of the Zoological Museum of Strasbourg, France. The dimensions of this facade is roughly 40×10 m. Note that the building was built in the 19th century and therefore presents a typical architecture of the era; indeed, it is part of the UNESCO World Heritage site of Neustadt since 2017. This heritage aspect is another challenge for the DL networks, since heritage architectural elements are more complex and thus more difficult to identify [23]. In this case, the terrestrial images were taken using a Canon EOS 6D DSLR camera with a 24 mm fixed lens.

A total of 33 images were acquired and processed using the open source Apero-Micmac software suite [47]. As an additional comparison, they were also independently processed using the commercial software Agisoft Metashape. The use of Apero-Micmac in this study is prioritised since almost if not all theoretical aspects of this open source suite can be determined and more importantly verified, whereas the same cannot be said of commercial solutions for understandable reasons related to trade secrets. That being said, Metashape also employs a bundle adjustment computation process [49] and an SGM-like dense matching approach [53].

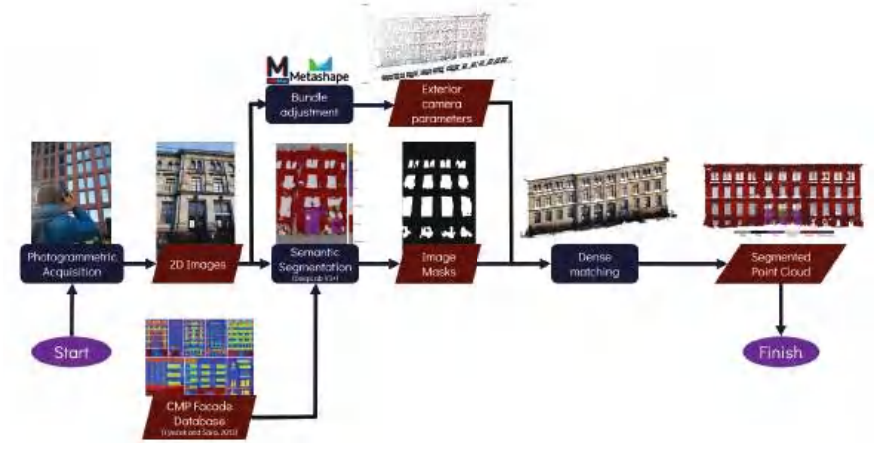


Figure 1. Developed workflow for the proposed semantic photogrammetry process.

Parallel to the computation of the image orientation parameters, a neural network was applied on the input images to semantically segment them. To this end, a DL network of the DeepLabV3+ architecture [86] pre-trained using a ResNet-18 network [87]. Using the pre-trained network, further training was performed using an open dataset prepared by the Center for Machine Perception (CMP) of the Czech Technical University [88]. This dataset consists of 606 rectified images of building facades with varying types of architecture. This process of transfer learning was deemed adequate to perform the 2D semantic segmentation of the case study presented in this paper. Furthermore, the images were classified into six classes: "pillar", "door", "facade", "window", "shops" and "background". The "shops" class refers to business signs and plaques. Note that this setup is more or less identical to a previous study as referred in [40].

Once the semantic segmentation was performed on all the input images, a simple script enabled the extraction of pixels pertaining to each class. Image masks were created for each class and for each image (Figure 2). These masks were then integrated into the photogrammetric process by applying them during the dense matching step. At this stage, it is assumed that the exterior orientation parameters acquired from the bundle adjustment are of a good quality. Dense matching was thereafter performed separately for each of the six classes using the image masks as constraints. The result is six distinct 3D dense point clouds which will naturally inherit the classes of the respective input 2D masks.

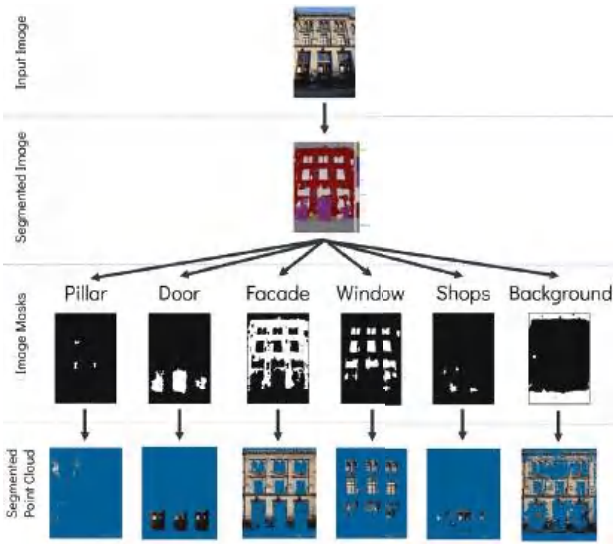


Figure 2. Creation of class-dependent image masks from the segmented image and its application in dense matching to generate semantically classified point clouds.

4. Experimental Results and Assessments

In the following section the case study on the Strasbourg Zoological Museum will be presented. A visual description of some of the results can be seen in Figure 3, in which dense point cloud generated by Micmac is presented, along with the manually segmented ground truth and the prediction results. The outcome of the same method applied in Metashape is also presented in said figure. It should be noted that the ground truth displayed in Figure 3 is created from manual segmentation of Micmac point cloud. A separate ground truth was also created for the Metashape point cloud.

In order to perform quantitative assessment on the results, several metrics were chosen to measure the performance of the proposed method. The semantic segmentation metrics of precision, recall and the aggregate F1 score were used in this regard. In addition, the Intersection over Union (IoU) score was also used to assess the results. As has been previously mentioned, for each photogrammetric software a separate ground truth was created. These ground truth data were created from combining all the separate point clouds generated by the method as described in Figures 1 and 2, and then manually labelled.

Table 1 shows the confusion matrix for the proposed semantic photogrammetry method applied to the software Micmac. Note that the assessment does not include the "background" class which was not considered particularly pertinent overall (see, however, a technical application in Section 5.3). In general, the proposed method seems to show promising results judging from the number of correctly classified points. Similarly Table 2 shows the same matrix for Metashape. In addition, Figure 4 presents a comparison between the statistics obtained from both Micmac and Metashape. In both cases, the proposed method was able to perform well in detecting and segmenting important classes such as windows and doors. The good performance on the facade class is nevertheless expected since it constitutes the majority of labels in any building-related semantic segmentation. The results for the "shops" class, in this case defined as panels and business signs, seem to be better in Metashape than Micmac. This point may be further explained by the fact that Metashape generated a much denser point cloud than Micmac. Indeed, this issue might be related to the fact that Micmac employs a much stricter post-filtering of dense matching than Metashape in relation to problematic areas such as little or textureless objects and shadows [18].

Figure 3. Visual illustration of some results from the experiment: (**a**) raw unclassified dense point cloud generated by Micmac, (**b**) manually segmented ground truth, (**c**) result of the semantic segmentation on Micmac dense point cloud and (**d**) result of the same procedure applied to Metashape dense point cloud.

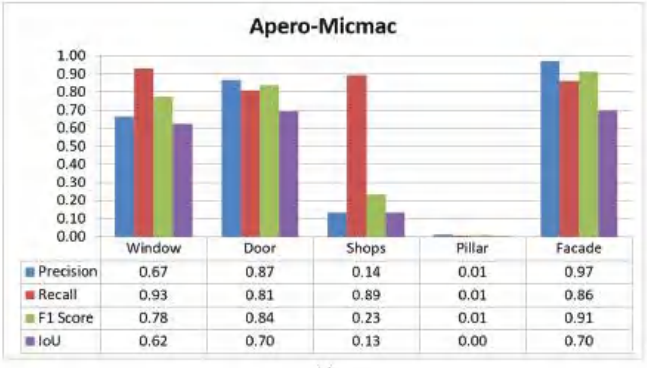
The results for the building openings (i.e., windows and doors) are especially encouraging because this has often been known to be a particular problem in building semantic segmentation, especially those using point-based approaches [89]. These results become even more interesting in light of the many potential applications for the automatic detection of building openings, such as automatic indoor–outdoor point cloud registration [90] or BIM creation [31]. For this reason, a further comparison was performed between these results and our implementation of PointNet++, which shall be detailed in Section 5.1. Furthermore, a comparison against another approach developed in a prior work shall also be explained in the next section.

Table 1. Confusion matrix for the semantic segmentation on Micmac dense point cloud.

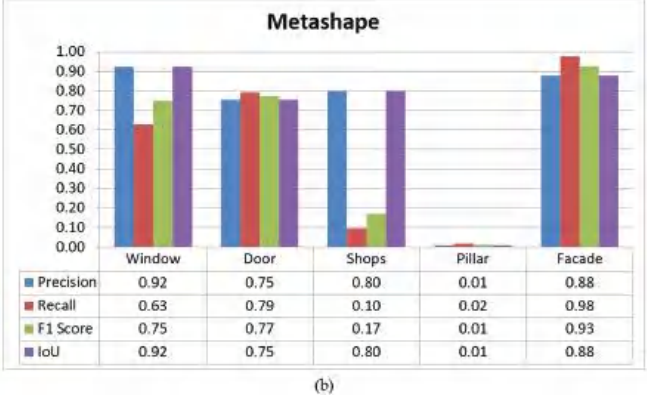
		Ground Truth					
	Class	**Window**	**Door**	**Shops**	**Pillar**	**Facade**	***Total***
Predicted	**Window**	**548,886**	3920	688	14,927	254,941	*823,362*
	Door	1566	**152,538**	0	0	21,483	*175,587*
	Shops	682	25,738	**6216**	0	13,326	*45,962*
	Pillar	15	0	0	**124**	9535	*9674*
	Facade	38,903	6121	66	6910	**1,876,172**	*1,928,172*
	Total	*590,052*	*188,317*	*6970*	*21,961*	*2,175,457*	***2,982,757***

Table 2. Confusion matrix for the semantic segmentation on Metashape dense point cloud.

		Ground Truth					
	Class	**Window**	**Door**	**Shops**	**Pillar**	**Facade**	***Total***
Predicted	**Window**	**3,104,942**	22,286	6467	6697	217,949	*3,358,341*
	Door	42,294	**819,405**	179,411	87	44,619	*1,085,816*
	Shops	6427	0	**28,283**	0	531	*35,241*
	Pillar	181,417	0	0	**2621**	54,961	*238,999*
	Facade	1,595,515	190,429	82,260	130,134	**14,612,534**	*16,610,872*
	Total	*4,930,595*	*1,032,120*	*296,421*	*139,539*	*14,930,594*	***21,329,269***



(a)



(b)

Figure 4. Performance statistics for the proposed method applied to dense point clouds generated by (**a**) Micmac and (**b**) Metashape.

5. Discussion

5.1. Comparison to Previous Work

In previous work detailed in Murtiyoso et al. (2021) [40], we presented another approach to reprojection-based semantic segmentation. While the neural network was prepared in a similar manner, in this paper the pixel class prediction was performed on the orthophoto of a building facade instead of the input images as presented in this paper. This method produced good results also for the building openings, but was severely limited by the fact that an orthophoto and a depth map are required as inputs. This may prove problematic in the case of more complex building architectures, hence the development of the semantic photogrammetry method as described here. In this section, a comparison is performed between the method proposed in this research and the one described in our previous work.

Furthermore, in another experiment conducted almost in parallel to the development of the methods in this paper, an implementation of the PointNet++ architecture was done for the Zoological Museum dataset [91]. This enables a further comparison to a point-based 3D segmentation method in order to better assess the results of this study. For the PointNet++ implementation, the 3D point cloud of the four facades was acquired. All of the facades were then manually labelled into classes. Three facades were then used to train the neural network, with the fourth and final facade used as a test data. In this case, the same main facade as the one used in this paper and in [40] was used.

For the purposes of this comparison, only three classes ("window", "door" and "facade") shall be compared since both the "shops" and "pillar" classes were grossly underrepresented in the training data for PointNet++. This is owed to the fact that within

the Zoological Museum dataset these two classes do not present adequate data, whereas they are not negligible in the CMAP image dataset used for the training in this paper.

Figure 5 describes the comparison between these methods in a histogram representation. As can be observed from the figure, the proposed method shows a clear advantage in regards to PointNet++. Indeed, for PointNet++ the "door" class is virtually non-existent while the "window" class IoU score is less than 0.5. As has been mentioned before, this is a known issue in direct point-based 3D segmentation. The main reasons are usually related to inadequacy in terms of training data and point features, especially in the case of building openings. Compared to our previous approach in [40], the proposed semantic photogrammetry method presented an improvement with regards to the two building opening classes while the detection of facade remained better in this previous approach. However, this previous method is very limited to certain buildings with mostly flat facades and few architectural ornaments. The need to acquire not only the orthophoto but also a depth map to reproject the labels to the 3D point cloud may also present additional problems. This would be the case especially in heritage buildings with more complex types of architecture.

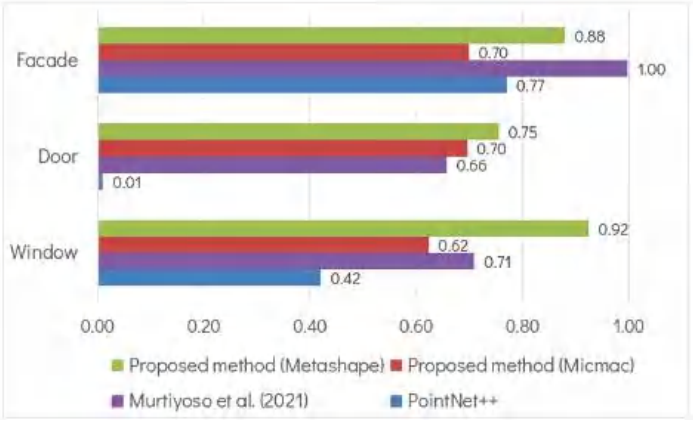


Figure 5. Comparison of IoU scores of the proposed method to other previous work.

5.2. Comparison to Other Studies

Finally, in order to further assess the results obtained especially in the case of building openings, a comparison was also performed to other studies which use AI-based semantic segmentation to perform the detection of openings, i.e., windows and/or doors. Four papers were identified, all of which were based on either an ML or DL, and are fairly recent. In Malinverni et al. (2019) [89], the authors used DGCNN to perform the task. Building upon this, Pierdicca et al. (2020) modified the base DGCNN architecture [41]. Matrone et al. (2020) [34] presented results not only from this modified DGCNN, but also the inclusion of 3D features during training. Finally, Grilli et al. (2020) [92] presented some results from their implementation of Random Forest (RF) algorithm.

Throughout these studies, the "window" and "facade" class was the only one common to all of them. Figure 6 shows therefore the comparison on the performance of each method in a histogram form. For comparison purposes, values for Grilli et al. (2020) and Matrone et al. (2020) represent the average of the several datasets described in those papers. Furthermore, results of the modified DGCNN with 3D features in Matrone et al. (2020) was chosen for this comparison. Similarly, the values for the proposed method present an average of results from both Micmac and Metashape. It should be noted that this comparison is only intended as a general overview, since for each study not only the method is different but also the nature of the case studies, the training data and their distribution of class labels as well as the determination of which classes were included

during the segmentation. From Figure 6, the semantic photogrammetry method proposed here seems to have an advantage at least for the "window" class.

Overall, the proposed method registered better scores compared to the four other studies using AI for the semantic segmentation of building openings. It is worth noting that the four studies included in the comparison are all based on point-based semantic segmentation, i.e., direct segmentation of the 3D point cloud. In the majority of these cases, classes representing building openings e.g., windows are often underrepresented, as can be seen in our own implementation of PointNet++ described previously in Section 5.1. On the other hand, facade or walls are mostly overrepresented, although in some of the cited studies the authors further divide the facade into several other classes, e.g., mouldings and vaults. This is reflected by the results from the three DL-based approaches of Malinverni et al. (2019), Matrone et al. (2020) and Pierdicca et al. (2020), as shown in Figure 6. However, using more classical Random Forest ML-based approach, Grilli et al. (2020) were able to achieve better results in the case of windows. The implemented semantic photogrammetry approach was able to outperform all other studies for the detection of building openings, while reaching a comparable result to RF in the case of facades.

Furthermore, it may be argued that in these other cases, the source of the point cloud is irrelevant due to the point-wise nature of the segmentation. Using the proposed approach, we argue that both the much more available training data for 2D segmentation and the introduction of the DL process into the photogrammetric workflow directly contribute to the observed performance. It is also interesting to note that the current implementation of semantic photogrammetry as described in this paper involves a small training dataset for DL standards (606 images), and further improvements and adaptations of this proof of concept may be envisaged in the future.

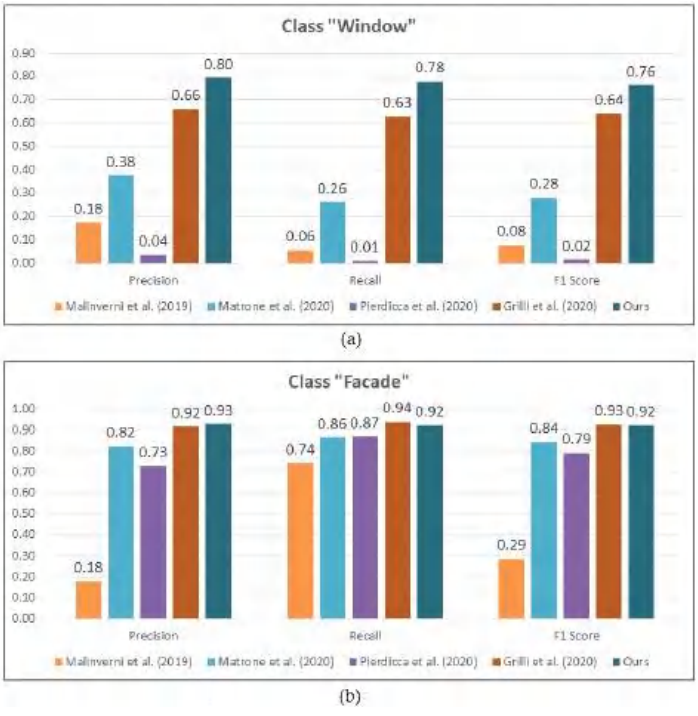


Figure 6. Comparison to other studies for the class (**a**) "window" and (**b**) "facade".

5.3. Example of Direct Application: Point Cloud Cleaning

In order to show the potential of the developed approach, an example of direct application can be seen in Figure 7. In this figure, the semantic photogrammetry approach was used to automatically mask unwanted objects in a scene, directly from the 2D images input. Concretely, this involves the inversion of the masks for the "background" class, thus excluding objects not considered as of interest. Furthermore, this approach for automatic point cloud cleaning not only excludes unwanted object classes, but may also reduce overall processing time during dense image matching. This is because the masks by virtue of its constraining effect reduces the area of interest to be matched. Quantitative assessment has shown that this method manages to achieve a 0.86 F1 score for the non-background classes (all combined).

Figure 7. Example of concrete application of the proposed method in photogrammetric point cloud cleaning: (**a**) original image, (**b**) mask of all classes except "background", (**c**) mask applied to the original image and (**d**) 3D point cloud from dense image matching using the masked image.

6. Conclusions and Future Investigations

This paper presents an approach to introduce AI-based semantic segmentation into the photogrammetric workflow, in an attempt to develop a semantic photogrammetry method. The proposed method takes benefit from the abundance of 2D image label data and reliable AI-based methods available today, in contrast to the scant availability of 3D labelled point clouds especially for large scale applications. With semantic segmentation performed on the 2D input images, a processing strategy based on the creation of 2D image masks were developed. The image masks created correspond to the class labels, and create therefore separate point clouds for each class.

The proposed method was implemented in both Apero-Micmac and Metashape. While the comparison of these two pieces of software in their capacity as photogrammetric solutions is beyond the scope of the paper, it has been shown that the quality of dense matching also plays a role in the final quality of the result. Furthermore, the post filtering process also plays a role as it determines the level of noise, i.e., false positives in the final dense point cloud. This relation between semantic photogrammetry and dense image matching quality has not been sufficiently investigated and may be an interesting subject for a future work.

Nevertheless, this paper attempted to present a proof of concept to the possibility to use AI in photogrammetric task. In the case study and comparisons, this was demonstrated in the case of building facade segmentation. The method has shown that the initial hypothesis of using the vastly more available labelled 2D training data is beneficial, as highlighted in the comparisons. Especially for the very interesting application of building opening detection, the proposed method has performed well. On the other side, this has also shown the limitation of the current implementation of the approach. Indeed, underrepresented classes, e.g., shop signs and pillars still pose problems although this is a more general problem with any method of semantic segmentation.

Based on the results obtained in the experiments, the developed method of semantic photogrammetry show much promise. It is also interesting to investigate its potentials for implementations in other settings, e.g., aerial photogrammetry, building interior modeling or even low-cost spherical photogrammetry. Evidently different scenes will require different sorts of DL learning; however, the overall semantic photogrammetry approach may be easily transposed on these different scenes thereafter.

Other points for improvement include the generation of 2D training data more suited to the encountered situation. For example, in this study the CMP database was used to train the neural network. This image database consists of rectified images, i.e., images already processed to have a perpendicular point of view. This does not exactly correspond to the input images in the experiments, which were close-range photogrammetry images. Methods to automatically create more suitable training data for close-range photogrammetry are also under investigation, with preliminary results described in [62].

Author Contributions: Conceptualization, A.M. (Arnadi Murtiyoso), P.G. and T.L.; methodology, A.M. (Arnadi Murtiyoso); software, A.M. (Arnadi Murtiyoso) and E.P.; validation, P.G., T.L. and A.M. (Andrea Masiero); formal analysis, A.M. (Arnadi Murtiyoso); investigation, A.M. (Arnadi Murtiyoso) and E.P.; resources, P.G., T.L. and A.M. (Andrea Masiero); data curation, A.M. (Arnadi Murtiyoso); writing—original draft preparation, A.M. (Arnadi Murtiyoso) and E.P.; writing—review and editing, P.G., T.L. and A.M. (Andrea Masiero); visualization, A.M. (Arnadi Murtiyoso); supervision, P.G., T.L. and A.M. (Andrea Masiero); project administration, P.G. and T.L.; funding acquisition, P.G. and T.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the French National Research Agency (ANR) under the BIOM (Building Indoor/Outdoor Modeling) project, grant number ANR-17-CE23-0003.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors would like to thank Camille Lhenry for her implementation of PointNet++ for the Zoological Museum dataset and Bastien Wirtz for his help during data acquisition.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Grilli, E.; Menna, F.; Remondino, F. A Review of Point Clouds Segmentation and Classification Algorithms. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2017**, *XLII-2/W3*, 339–344. [CrossRef]
2. Mölg, N.; Bolch, T. Structure-from-motion using historical aerial images to analyse changes in glacier surface elevation. *Remote Sens.* **2017**, *9*, 1021. [CrossRef]
3. Abate, D.; Murtiyoso, A. Bundle adjustment accuracy assessment of unordered aerial dataset collected through Kite platform. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *XLII-2/W17*, 1–8. [CrossRef]
4. Grussenmeyer, P.; Hanke, K.; Streilein, A. Architectural Photogrammetry. In *Digital Photogrammetry*; Kasser, M., Egels, Y., Eds.; Taylor and Francis: London, UK, 2002; pp. 300–339.
5. Granshaw, S.I. Bundle Adjustment Methods in Engineering Photogrammetry. *Photogramm. Rec.* **1980**, *10*, 181–207. [CrossRef]
6. Grussenmeyer, P.; Al Khalil, O. Solutions for exterior orientation in photogrammetry: A review. *Photogramm. Rec.* **2002**, *17*, 615–634. [CrossRef]
7. Gruen, A. Adaptive least squares correlation: A powerful image matching technique. *S. Afr. J. Photogramm. Remote Sens. Cartogr.* **1985**, *14*, 175–187.
8. Börlin, N.; Grussenmeyer, P. Bundle adjustment with and without damping. *Photogramm. Rec.* **2013**, *28*, 396–415. [CrossRef]
9. Wu, C.; Agarwal, S.; Curless, B.; Seitz, S.M. Multicore bundle adjustment. In Proceedings of the CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011; Volume 1, 3057–3064. [CrossRef]
10. Szeliski, R. *Computer Vision: Algorithms and Applications*; Springer: Berlin/Heidelberg, Germany, 2010; Volume 5, p. 832.
11. Lowe, D.G. Distinctive image features from scale invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
12. Hirschmüller, H. Semi-Global Matching Motivation, Developments and Applications. In Proceedings of the Photogrammetric Week, Stuttgart, Germany, 9–13 September 2011; pp. 173–184.
13. Furukawa, Y.; Ponce, J. Accurate, dense, and robust multi-view stereopsis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2009**, *32*, 1362–1376. [CrossRef]
14. Barazzetti, L.; Previtali, M.; Roncoroni, F. Can we use low-cost 360 degree cameras to create accurate 3D models? *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2018**, *XLII-2*, 69–75. [CrossRef]
15. Kim, J.; Lee, S.; Ahn, H.; Seo, D.; Park, S.; Choi, C. Feasibility of employing a smartphone as the payload in a photogrammetric UAV system. *ISPRS J. Photogramm. Remote Sens.* **2013**, *79*, 1–18. [CrossRef]

16. Wand, M.; Berner, A.; Bokeloh, M.; Jenke, P.; Fleck, A.; Hoffmann, M.; Maier, B.; Staneker, D.; Schilling, A.; Seidel, H.P. Processing and interactive editing of huge point clouds from 3D scanners. *Comput. Graph.* **2008**, *32*, 204–220. [CrossRef]
17. Meng, F.; Zha, H. An Easy Viewer for Out-of-core Visualization of Huge Point-sampled Models. In Proceedings of the IAN Proceedings 2nd International Symposium on 3D Data Processing, Visualization and Transmission 2004, Thessaloniki, Greece, 9 September 2004; pp. 207–214. [CrossRef]
18. Murtiyoso, A.; Grussenmeyer, P. Documentation of heritage buildings using close-range UAV images: Dense matching issues, comparison and case studies. *Photogramm. Rec.* **2017**, *32*, 206–229. [CrossRef]
19. Campanaro, D.M.; Landeschi, G.; Dell'Unto, N.; Leander Touati, A.M. 3D GIS for cultural heritage restoration: A 'white box' workflow. *J. Cult. Herit.* **2016**, *18*, 321–332. [CrossRef]
20. Murtiyoso, A.; Veriandi, M.; Suwardhi, D.; Soeksmantono, B.; Harto, A.B. Automatic Workflow for Roof Extraction and Generation of 3D CityGML Models from Low-Cost UAV Image-Derived Point Clouds. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 743. [CrossRef]
21. Barsanti, S.G.; Remondino, F.; Fenández-Palacios, B.J.; Visintini, D. Critical factors and guidelines for 3D surveying and modelling in Cultural Heritage. *Int. J. Herit. Digit. Era* **2014**, *3*, 141–158. [CrossRef]
22. Nex, F.; Gerke, M.; Remondino, F.; Przybilla, H.J.; Bäumker, M.; Zurhorst, A. ISPRS benchmark for multi-platform photogrammetry. *Isprs Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *II-3/W4*, 135–142. [CrossRef]
23. Matrone, F.; Lingua, A.; Pierdicca, R.; Malinverni, E.S.; Paolanti, M.; Grilli, E.; Remondino, F.; Murtiyoso, A.; Landes, T. A benchmark for large-scale heritage point cloud semantic segmentation. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2020**, *43*, 1419–1426. [CrossRef]
24. Poux, F.; Neuville, R.; Nys, G.A.; Billen, R. 3D point cloud semantic modelling: Integrated framework for indoor spaces and furniture. *Remote Sens.* **2018**, *10*, 1412. [CrossRef]
25. Fabbri, K.; Zuppiroli, M.; Ambrogio, K. Heritage buildings and energy performance: Mapping with GIS tools. *Energy Build.* **2012**, *48*, 137–145. [CrossRef]
26. Seker, D.Z.; Alkan, M.; Kutoglu, H.; Akcin, H.; Kahya, Y. Development of a GIS Based Information and Management System for Cultural Heritage Site; Case Study of Safranbolu. In Proceedings of the FIG Congress 2010, Sydney, Australia, 11–16 April 2010.
27. Gröger, G.; Plümer, L. CityGML—Interoperable semantic 3D city models. *ISPRS J. Photogramm. Remote Sens.* **2012**, *71*, 12–33. [CrossRef]
28. Biljecki, F.; Stoter, J.; Ledoux, H.; Zlatanova, S.; Çöltekin, A. Applications of 3D city models: State of the art review. *ISPRS Int. J. Geo-Inf.* **2015**, *4*, 2842–2889. [CrossRef]
29. Volk, R.; Stengel, J.; Schultmann, F. Building Information Modeling (BIM) for existing buildings—Literature review and future needs. *Autom. Constr.* **2014**, *38*, 109–127. [CrossRef]
30. Macher, H.; Landes, T.; Grussenmeyer, P. Point clouds segmentation as base for as-built BIM creation. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2015**, *II-5/W3*, 191–197. [CrossRef]
31. Macher, H.; Landes, T.; Grussenmeyer, P. From Point Clouds to Building Information Models: 3D Semi-Automatic Reconstruction of Indoors of Existing Buildings. *Appl. Sci.* **2017**, *7*, 1030. [CrossRef]
32. Bassier, M.; Bonduel, M.; Genechten, B.V.; Vergauwen, M. Octree-Based Region Growing and Conditional Random Fields. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2017**, *XLII-2/W8*, 25–30. [CrossRef]
33. Murtiyoso, A.; Grussenmeyer, P. Virtual disassembling of historical edifices: Experiments and assessments of an automatic approach for classifying multi-scalar point clouds into architectural elements. *Sensors* **2020**, *20*, 2161. [CrossRef]
34. Matrone, F.; Grilli, E.; Martini, M.; Paolanti, M.; Pierdicca, R.; Remondino, F. Comparing machine and deep learning methods for large 3D heritage semantic segmentation. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 535. [CrossRef]
35. Granshaw, S.I. Photogrammetric terminology: Fourth edition. *Photogramm. Rec.* **2020**, *35*, 143–288. [CrossRef]
36. Jinqiang, W.; Basnet, P.; Mahtab, S. Review of machine learning and deep learning application in mine microseismic event classification. *Min. Miner. Depos.* **2021**, *15*, 19–26. [CrossRef]
37. Maalek, R.; Lichti, D.D.; Ruwanpura, J.Y. Automatic recognition of common structural elements from point clouds for automated progress monitoring and dimensional quality control in reinforced concrete construction. *Remote Sens.* **2019**, *11*, 1102. [CrossRef]
38. Murtiyoso, A.; Grussenmeyer, P. Automatic heritage building point cloud segmentation and classification using geometrical rules. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. -ISPRS Arch.* **2019**, *XLII-2/W15*, 821–827. [CrossRef]
39. Kirillov, A.; He, K.; Girshick, R.; Rother, C.; Dollar, P. Panoptic segmentation. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 9396–9405. [CrossRef]
40. Murtiyoso, A.; Lhenry, C.; Landes, T.; Grussenmeyer, P.; Alby, E. Semantic Segmentation for Building Façade 3D Point Cloud From 2D Orthophoto Images Using Transfer Learning. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2021**, *XLIII-B2-2*, 201–206. [CrossRef]
41. Pierdicca, R.; Paolanti, M.; Matrone, F.; Martini, M.; Morbidoni, C.; Malinverni, E.S.; Frontoni, E.; Lingua, A.M. Point Cloud Semantic Segmentation Using a Deep Learning Framework for Cultural Heritage. *Remote Sens.* **2020**, *12*, 1005. [CrossRef]
42. Stathopoulou, E.K.; Remondino, F. Semantic photogrammetry—Boosting image-based 3D reconstruction with semantic labeling. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *XLII-2/W9*, 685–690. [CrossRef]
43. Heipke, C.; Rottensteiner, F. Deep learning for geometric and semantic tasks in photogrammetry and remote sensing. *Geo-Spat. Inf. Sci.* **2020**, *23*, 10–19. [CrossRef]

44. Stathopoulou, E.K.; Remondino, F. Multi-view stereo with semantic priors. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. -SPRS Arch.* **2019**, *XLII-2/W15*, 1135–1140. [CrossRef]
45. Grilli, E.; Battisti, R.; Remondino, F. An advanced photogrammetric solution to measure apples. *Remote Sens.* **2021**, *13*, 3960. [CrossRef]
46. Kernell, B. Improving Photogrammetry Using Semantic Segmentation. Ph.D. Thesis, Linköping University, Linköping, Sweden, 2018.
47. Rupnik, E.; Daakir, M.; Pierrot Deseilligny, M. MicMac—A free, open-source solution for photogrammetry. *Open Geospat. Data Softw. Stand.* **2017**, *2*, 14. [CrossRef]
48. Schenk, T. *Introduction to Photogrammetry*; Department of Civil and Environmental Engineering and Geodetic Science, The Ohio State University: Columbus, OH, USA, 2005; pp. 79–95.
49. Murtiyoso, A.; Grussenmeyer, P.; Börlin, N.; Vandermeerschen, J.; Freville, T. Open Source and Independent Methods for Bundle Adjustment Assessment in Close-Range UAV Photogrammetry. *Drones* **2018**, *2*, 3. [CrossRef]
50. Wolf, P.; DeWitt, B.; Wilkinson, B. *Elements of Photogrammetry with Applications in GIS*, 4th ed.; McGraw-Hill Education: New York, NY, USA 2014; p. 696.
51. Luhmann, T.; Robson, S.; Kyle, S.; Boehm, J. *Close-Range Photogrammetry and 3D Imaging*, 2nd ed.; De Gruyter: Berlin, Germany 2014; p. 684.
52. Hirschmüller, H. Accurate and efficient stereo processing by semi-global matching and mutual information. In Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 20–25 June 2005; pp. 807–814.
53. Remondino, F.; Spera, M.G.; Nocerino, E.; Menna, F.; Nex, F. State of the art in high density image matching. *Photogramm. Rec.* **2014**, *29*, 144–166. [CrossRef]
54. Murtiyoso, A.; Grussenmeyer, P.; Suwardhi, D.; Awalludin, R. Multi-Scale and Multi-Sensor 3D Documentation of Heritage Complexes in Urban Areas. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 483. [CrossRef]
55. Bedford, J. *Photogrammetric Applications for Cultural Heritage*; Historic England: Swindon, UK, 2017; p. 128.
56. Kalinichenko, V.; Dolgikh, O.; Dolgikh, L.; Pysmennyi, S. Choosing a camera for mine surveying of mining enterprise facilities using unmanned aerial vehicles. *Min. Miner. Depos.* **2020**, *14*, 31–39. [CrossRef]
57. Wenzel, K.; Rothermel, M.; Fritsch, D.; Haala, N. Image acquisition and model selection for multi-view stereo. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2013**, *XL-5/W1*, 251–258. [CrossRef]
58. Verhoeven, G.; Taelman, D.; Vermeulen, F. Computer Vision-Based Orthophoto Mapping of Complex Archaeological Sites: The Ancient Quarry of Pitaranha (Portugal-Spain). *Archaeometry* **2012**, *54*, 1114–1129. [CrossRef]
59. Bassier, M.; Vergauwen, M.; Van Genechten, B. Automated Classification of Heritage Buildings for As-Built BIM using Machine Learning Techniques. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2017**, *IV-2/W2*, 25–30. [CrossRef]
60. Poux, F.; Hallot, P.; Neuville, R.; Billen, R. Smart Point Cloud: Definition and Remaining Challenges. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* **2016**, *IV-2/W1*, 119–127. [CrossRef]
61. Xie, Y.; Tian, J.; Zhu, X.X. Linking Points with Labels in 3D: A Review of Point Cloud Semantic Segmentation. *IEEE Geosci. Remote Sens. Mag.* **2020**, *8*, 38–59. [CrossRef]
62. Pellis, E.; Masiero, A.; Tucci, G.; Betti, M.; Grussenmeyer, P. Towards an Integrated Design Methodology for H-Bim. In Proceedings of the Joint International Event 9th ARQUEOLÓGICA 2.0 and 3rd GEORES, Valencia, Spain, 26–28 April 2021; pp. 389–398. [CrossRef]
63. Zhang, K.; Hao, M.; Wang, J.; de Silva, C.W.; Fu, C. Linked dynamic graph CNN: Learning on point cloud via linking hierarchical features. *arXiv* **2019**, arXiv:1904.10014.
64. Su, H.; Maji, S.; Kalogerakis, E.; Learned-Miller, E. Multi-view convolutional neural networks for 3D shape recognition. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 945–953. [CrossRef]
65. Boulch, A.; Guerry, J.; Le Saux, B.; Audebert, N. SnapNet: 3D point cloud semantic labeling with 2D deep segmentation networks. *Comput. Graph. (Pergamon)* **2018**, *71*, 189–198. [CrossRef]
66. Maturana, D.; Scherer, S. VoxNet: A 3D Convolutional Neural Network for Real-Time Object Detection. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Sustems (IROS), Hamburg, Germany, 28 September–2 October 2015; pp. 922–928.
67. Riegler, G.; Ulusoy, A.O.; Geiger, A. OctNet: Learning deep 3D representations at high resolutions. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, 21–26 July 2017; pp. 6620–6629. [CrossRef]
68. Tchapmi, L.P.; Choy, C.B.; Armeni, I.; Gwak, J.; Savarese, S. SEGCloud: Semantic Segmentation of 3D Point Clouds. In Proceedings of the 2017 International Conference on 3D Vision (3DV), Qingdao, China, 10–12 October 2017,
69. Iandola, F.N.; Han, S.; Moskewicz, M.W.; Ashraf, K.; Dally, W.J.; Keutzer, K. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size. *arXiv* **2016**, arXiv:1602.07360.
70. Milioto, A.; Vizzo, I.; Behley, J.; Stachniss, C. RangeNet ++: Fast and Accurate LiDAR Semantic Segmentation. In Proceedings of the 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, 3–8 November 2019; pp. 4213–4220. [CrossRef]

71. Su, H.; Jampani, V.; Sun, D.; Maji, S.; Kalogerakis, E.; Yang, M.H.; Kautz, J. SPLATNet: Sparse Lattice Networks for Point Cloud Processing. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2530–2539. [CrossRef]
72. Rosu, R.A.; Schütt, P.; Quenzel, J.; Behnke, S. LatticeNet: Fast Point Cloud Segmentation Using Permutohedral Lattices. *arXiv* **2020**, arXiv:1912.05905.
73. Choy, C.B.; Gwak, J.; Savarese, S. 4D Spatio-Temporal ConvNets: Minkowski Convolutional Neural Networks. *arXiv* **2019**, arXiv:1904.08755.
74. Qi, C.R.; Su, H.; Mo, K.; Guibas, L.J. Pointnet: Deep learning on point sets for 3d classification and segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 652–660.
75. Qi, C.R.; Yi, L.; Su, H.; Guibas, L.J. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–7 December 2017; pp. 5099–5108.
76. Wang, S.; Suo, S.; Ma, W.C.; Pokrovsky, A.; Urtasun, R. Deep Parametric Continuous Convolutional Neural Networks. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2589–2597. [CrossRef]
77. Boulch, A. ConvPoint: Continuous Convolutions for Point Cloud Processing. *Comput. Graph.* **2020**, *88*, 24–34. [CrossRef]
78. Liu, F.; Li, S.; Zhang, L.; Zhou, C.; Ye, R.; Wang, Y.; Lu, J. 3DCNN-DQN-RNN: A Deep Reinforcement Learning Framework for Semantic Parsing of Large-Scale 3D Point Clouds. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017, pp. 5679–5688. [CrossRef]
79. Wang, Y.; Sun, Y.; Liu, Z.; Sarma, S.E.; Bronstein, M.M.; Solomon, J.M. Dynamic graph Cnn for learning on point clouds. *ACM Trans. Graph.* **2019**, *38*, 1–12. [CrossRef]
80. Minaee, S.; Boykov, Y.Y.; Porikli, F.; Plaza, A.J.; Kehtarnavaz, N.; Terzopoulos, D. Image Segmentation Using Deep Learning: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**. [CrossRef] [PubMed]
81. Wang, Y.; Ji, R.; Chang, S.F. Label propagation from imagenet to 3D point clouds. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 3135–3142. [CrossRef]
82. Tasaka, K.; Yanagihara, H.; Lertniphonphan, K.; Komorita, S. 2D TO 3D Label Propagation for Object Detection in Point Cloud. In Proceedings of the 2018 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), San Diego, CA, USA, 23–27 July 2018; pp. 1–6.
83. Reza, M.A.; Zheng, H.; Georgakis, G.; Kosecka, J. Label propagation in RGB-D video. In Proceedings of the IEEE International Conference on Intelligent Robots and Systems, Vancouver, BC, Canada, 24–28 September 2017; pp. 4917–4922. [CrossRef]
84. Xie, J.; Kiefel, M.; Sun, M.T.; Geiger, A. Semantic Instance Annotation of Street Scenes by 3D to 2D Label Transfer. *arXiv* **2016**, arXiv:1511.03240.
85. Babahajiani, P.; Fan, L.; Kämäräinen, J.K.; Gabbouj, M. Urban 3D segmentation and modelling from street view images and LiDAR point clouds. *Mach. Vis. Appl.* **2017**, *28*, 679–694. [CrossRef]
86. Chen, L.C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [CrossRef]
87. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
88. Tyleček, R.; Šára, R. Spatial pattern templates for recognition of objects with regular structure. *Lect. Notes Comput. Sci. (Incl. Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinform.)* **2013**, *8142 LNCS*, 364–374. [CrossRef]
89. Malinverni, E.S.; Pierdicca, R.; Paolanti, M.; Martini, M.; Morbidoni, C.; Matrone, F.; Lingua, A. Deep learning for semantic segmentation of point cloud. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *XLII-2/W15*, 735–742. [CrossRef]
90. Assi, R.; Landes, T.; Murtiyoso, A.; Grussenmeyer, P. Assessment of a Keypoints Detector for the Registration of Indoor and Outdoor Heritage Point Clouds. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2019**, *XLII-2/W15*, 133–138. [CrossRef]
91. Landes, T.; Macher, H.; Murtiyoso, A.; Lhenry, C.; Alteirac, V.; Lallement, A.; Kastendeuch, P. Detection and 3D Reconstruction of Urban Trees and Façade Openings by Segmentation of Point Clouds: First Experiment with PointNet++. In Proceedings of the International Symposium on Applied Geoinformatics, Riga, Latvia, 2–3 December 2021.
92. Grilli, E.; Remondino, F. Machine Learning Generalisation across Different 3D Architectural Heritage. *ISPRS Int. J. Geo-Inf.* **2020**, *9*, 379. [CrossRef]

sensors

MDPI

Article

SSA with CWT and *k*-Means for Eye-Blink Artifact Removal from Single-Channel EEG Signals

Ajay Kumar Maddirala and Kalyana C. Veluvolu *

School of Electronics Engineering, College of IT Engineering, Kyungpook National University, Daegu 41566, Korea; maddirala@knu.ac.kr
* Correspondence: veluvolu@ee.knu.ac.kr

Abstract: Recently, the use of portable electroencephalogram (EEG) devices to record brain signals in both health care monitoring and in other applications, such as fatigue detection in drivers, has been increased due to its low cost and ease of use. However, the measured EEG signals always mix with the electrooculogram (EOG), which are results due to eyelid blinking or eye movements. The eye-blinking/movement is an uncontrollable activity that results in a high-amplitude slow-time varying component that is mixed in the measured EEG signal. The presence of these artifacts misled our understanding of the underlying brain state. As the portable EEG devices comprise few EEG channels or sometimes a single EEG channel, classical artifact removal techniques such as blind source separation methods cannot be used to remove these artifacts from a single-channel EEG signal. Hence, there is a demand for the development of new single-channel-based artifact removal techniques. Singular spectrum analysis (SSA) has been widely used as a single-channel-based eye-blink artifact removal technique. However, while removing the artifact, the low-frequency components from the non-artifact region of the EEG signal are also removed by SSA. To preserve these low-frequency components, in this paper, we have proposed a new methodology by integrating the SSA with continuous wavelet transform (CWT) and the *k*-means clustering algorithm that removes the eye-blink artifact from the single-channel EEG signals without altering the low frequencies of the EEG signal. The proposed method is evaluated on both synthetic and real EEG signals. The results also show the superiority of the proposed method over the existing methods.

Keywords: electroencephalogram (EEG); electrooculogram (EOG); singular spectrum analysis (SSA); continuous wavelet transform (CWT); *k*-means clustering

Citation: Maddirala, A.K.; Veluvolu, K.C. SSA with CWT and *k*-Means for Eye-Blink Artifact Removal from Single-Channel EEG Signals. *Sensors* **2022**, *22*, 931. https://doi.org/10.3390/s22030931

Academic Editor: Manuel José Cabral dos Santos Reis

Received: 21 October 2021
Accepted: 21 January 2022
Published: 25 January 2022

1. Introduction

Electroencephalogram (EEG) signals represent the electrical activity of the brain and are measured by placing electrodes over the scalp. The EEG signals are often used to understand brain functions such as mental state (or cognitive state) and brain disorders such as epilepsy and stroke [1–5]. However, the recorded EEG signals are always contaminated by physiological artifacts, such as electrooculogram (EOG), electromyogram (EMG) and electrocardiogram (ECG). Unlike other artifacts, the EOG artifact that is a result of the eye-blink/movement activity and always contaminates the EEG signal. As the eye-blink is an uncontrollable and involuntary activity and occurs once every 5 s (as in [6]), we refer to the EOG artifact as an eye-blink artifact in this paper. Therefore, the removal of these artifacts forms an important stage before analyzing the EEG signals [6]. Hence, methods such as linear filters have been used for eye-blink artifact removal from EEG signals. In general, the eye-blink artifact strongly contaminates the low-frequency spectrum of EEG (0.5–12 Hz) [7]. Therefore, the use of linear filters for the removal of eye-blink artifacts alters the valuable information from the EEG signal. Later, a regression-based method was proposed to remove artifacts from multichannel EEG signals [8]. In this method, the artifact weighting coefficients are computed from the EOG channels, which are

recorded separately. However, such fixed coefficients may not fully remove the eye-blink artifacts from the EEG signals.

Blind source separation (BSS) techniques such as independent component analysis (ICA) and canonical correlation analysis (CCA) techniques have been used to remove artifacts from the multichannel EEG signals [9–13]. The ICA technique was extensively used to remove eye-blink artifacts from EEG signals as compared to the CCA method [12,13]. Several other techniques were also integrated with ICA for efficient removal of eye-blink artifacts from the multichannel EEG signals [14–17]. The artifact subspace reconstruction (ASR) method was also proposed to remove the artifact from the EEG signals [18,19]. The performance of this method depends on the user-defined cut-off parameter k. Even though a detailed study was conducted for selecting the cut-off parameter in [19], inappropriate selection of this parameter may result in the loss of EEG information.

Recently, the demand for in-home health monitoring has been increasing due to the increase in chronic illnesses and population aging [20]. Several studies have employed portable EEG devices for various applications, including analysis of cognitive state in stroke survivors, sleep disorders, driver fatigue and event-related potential (ERP)-based BCI applications [2,21–23]. To reduce the burden and to minimize the stress on the patient, recently portable EEG devices with a reduced number of EEG channels, including single EEG channel equipment [24,25], have been developed. Therefore, the existing ICA and ASR techniques that are popular for multichannel settings cannot be used to remove eye-blink artifacts from single-channel EEG signals. Therefore, there is a need for new methods that are customized for processing single-channel EEG signals.

An adaptive filter is one of the possible solutions to process single-channel EEG signals. The use of adaptive filters to remove eye-blink artifacts from the EEG signals was first discussed in [26]. However, the adaptive filters require reference signals to remove the eye-blink artifacts from single-channel EEG data. Therefore, in [27], the adaptive filter is combined with discrete wavelet transform (DWT) to solve this problem. In this method, the reference signal (an approximated eye-blink artifact) needed for the adaptive filter is estimated from the contaminated EEG signal using DWT. After that, the estimated eye-blink artifact signal is used as a reference signal to the adaptive filter to remove the eye-blink artifact from the EEG signal. Recently, the Savitzky–Golay (SG) filter was also used to estimate the reference signal needed for an adaptive filter [28]. Very recently, the Variational Mode Extraction (VME) and DWT techniques were combined to remove eye-blink artifacts from single-channel EEG signals [29]. In this method, first, the eye-blink artifact interval is identified using VME. Next, a DWT algorithm is employed to filter the contaminated interval of the EEG signal. Although this method does not significantly alter the non-artifact regions of the EEG signal, the eye-blink artifact component is partially removed from the contaminated EEG signal. Along with these methods, a data-driven decomposition method, namely an ensemble empirical mode decomposition with adaptive noise, is also proposed to remove eye-blink artifacts from a single-channel EEG signal [30]. However, this method alters the non-artifact regions of the EEG signal.

Singular spectrum analysis (SSA) is a subspace-based technique used to extract the low-frequency, oscillating and noise components from uni-variate time-series data [31,32]. Recently, the SSA technique has been applied for processing the biomedical signals [33–36]. The application of SSA for eye-blink artifact removal from single-channel EEG signals was first studied in [37]. However, identifying the desired signal subspace (eigenvectors) is a critical step in classical SSA. Therefore, new criteria were proposed in [38] to identify the eigenvectors that are used to reconstruct the desired signal. In [38], the SSA is combined with an adaptive filter to enhance the performance of the adaptive filter over the method in [27]. Recently, in [39], with new grouping criteria, the adaptive SSA technique is combined with ANC (SSA+ANC) and the method showed better performance over the method in [38]. Moreover, SSA is used as a means to apply ICA on single-channel EEG signals [40,41]. Very recently, SSA has been used as a smoothing filter in [42] to remove the eye-blink artifact from the EEG signal. In this method, the user has to adjust the thresh-

old for faithful separation of the eye-blink artifact from the EEG signal. In other words, the performance of the method is sensitive to the user-defined threshold.

Even though the SSA is able to extract the eye-blink artifact efficiently, it also removes the EEG low-frequency information (0.5–12 Hz) from the non-artifact regions. Removing these components may affect the subsequent analysis of the EEG signal. Recently, the effect of pre-processing methods on EEG results has been studied in [43] and it concludes that the selection of artifact removal strategy affects the end application results. Therefore, care should be taken while designing the artifact removal method. Therefore, in this paper, we proposed a new technique by combining SSA with continuous wavelet transform (CWT) and *k*-means algorithms so that it removes the eye-blink artifact from single-channel EEG signal without altering the non-artifact regions of the EEG signal. The proposed method exploited the strengths of both SSA and the CWT in removing the artifact. Unlike the method in [42], where time-domain features are used, the proposed method used frequency-domain features of the signal to remove the eye-blink artifact. Moreover, a frequency-based threshold is defined for SSA to identify the artifact subspace, and such threshold will act as the cut-off frequency as in a low-pass filter. The performance of the proposed method (which we call SSA-CWT) is evaluated on synthetic and real single-channel EEG signals. The results show its superiority over existing methods.

The rest of the paper is organized as follows: The performance measures to evaluate the efficiency of the proposed and existing method are defined in Section 2. The framework of the proposed method is discussed in Section 3. The simulation results and their discussions are presented in Section 4 and Section 5, respectively. Section 6 concludes the paper.

2. Performance Metrics

In this section, we have employed several few performance metrics to evaluate the performance of the proposed method on a synthetic EEG dataset. We define four commonly used performance measures to evaluate the performance of the methods on synthetic EEG data: the relative root mean square error (RRMSE), the canonical correlation analysis (CC), artifact reduction ratio (λ) and mean absolute error (MAE). To evaluate the performance of the proposed method on real EEG datasets, we first identify the non-artifact and artifact intervals of the real EEG signal manually. Then RRMSE and CC between the non-artifact interval of contaminated and corrected EEG signals is computed.

Consider the N sampled contaminated signal $\mathbf{x} = \mathbf{s} + p.\mathbf{a}$, where $\mathbf{s}$ and $\mathbf{a}$ are the true EEG and the EOG artifact signals, respectively and p is an artifact mixing constant. The following performance metrics are defined as follows:

2.1. Relative Root Measure Square Error (RRMSE)

The RRMSE measure is often used to evaluate the performance of artifact removal methods on synthetic EEG data. The RRMSE between the two signals $\mathbf{a}$ and $\hat{\mathbf{a}}$ can be defined as

$$RRMSE = \sqrt{\frac{\sum_{n=1}^{N} [a(n) - \hat{a}(n)]^2}{\sum_{n=1}^{N} a^2(n)}} \times 100(\%) \quad (1)$$

where $\mathbf{a}$ and $\hat{\mathbf{a}}$ represent the ground truth eye-blink and the estimated eye-blink artifacts, respectively. The relationship between the signal-to-noise ratio (SNR) and the artifact mixing constant p is given by

$$SNR = \frac{RMS(\mathbf{s})}{RMS(p\mathbf{a})}$$

$$RMS(\mathbf{s}) = \sqrt{\frac{1}{N} \sum_{n=1}^{N} s^2(n)}$$

when the constant p is small, the EOG artifact is small and the SNR of the EEG signal is high. The low RRMSE value indicates a good estimation of artifacts by the method. Here, the RRMSE is computed between the true eye-blink and the estimated eye-blink artifact to understand the efficacy of the proposed method in estimating the artifact from the contaminated EEG signal.

2.2. Correlation Coefficient (CC)

It is a statistical-based measure, which shows the strong relationship between the two signals. The *CC* measure is also used to evaluate the performance of an artifact removal technique. The CC between the two signals $\mathbf{a}$ and $\hat{\mathbf{a}}$ can be defined as

$$CC = \frac{cov(\mathbf{a}, \hat{\mathbf{a}})}{\sigma_{\mathbf{a}\mathbf{a}}\sigma_{\hat{\mathbf{a}}\hat{\mathbf{a}}}} \tag{2}$$

where $cov(\cdot)$ represents the covariance between the two signals $\mathbf{a}$ and $\hat{\mathbf{a}}$ and $\sigma_{(\cdot)}$ variance of the signal itself. The *CC* value close to one indicates a good estimation of eye-blink artifact from the contaminated EEG data.

2.3. Artifact Reduction Ratio (λ)

Along with the above-defined two performance measures, we also employed a performance metric that quantifies the percentage reduction in artifacts and is defined as

$$\lambda = \left(1 - \frac{R_{clean} - R_{after}}{R_{clean} - R_{before}}\right) \times 100 \tag{3}$$

where R_{clean} is set to 1 and the R_{before} is the correlation between the true EEG and the contaminated EEG signals and R_{after} is the correlation between the true EEG and the estimated EEG signals. For a good artifact removal method, this value should be high.

2.4. Mean Absolute Error (MAE)

This metric is employed to evaluate the performance of the proposed method in the frequency domain. It is defined as the sum of the absolute of the difference between the true EEG signal power spectrum, $P_{\mathbf{s}}$, and the corrected EEG signal power spectrum $P_{\hat{\mathbf{s}}}$ in a particular band. The MAE between the spectrums of the true and corrected EEG signals is defined as

$$MAE = \frac{\sum_{i=1}^{K} |P_{\mathbf{s}}(i) - P_{\hat{\mathbf{s}}}(i)|}{K} \tag{4}$$

where K represents the number of frequency bins in a specific band. The MAE value is expected to be very small for a good artifact removal method.

2.5. Precision and Accuracy

Along with these performance measures, we have also defined two measures associated to binary classification, precision and accuracy, to detect how precisely and accurately the proposed method identifies (detected) the artifact and non-artifact intervals of the EEG signals. The performance measures, precision and accuracy are defined as

$$\text{Precision} = \frac{TP}{(TP + FP)} \tag{5}$$

$$\text{Accuracy} = \frac{TP + TN}{(TP + TN + FP + FN)} \tag{6}$$

where TP, TN, FP and FN are true positive, true negative, false positive and false negative, respectively. The true positive indicates that the artifact removal method correctly predicted (detected) the positive class (artifact interval) and true negative indicates that the method correctly detected the negative class (non-artifact interval). Similarly, false positive and

false negatives represents that the method incorrectly detected the positive and negative classes, respectively.

3. Eye-Blink Artifact Removal from Single-Channel EEG Signals

The key components of the proposed method for eye-blink artifact removal is shown in Figure 1. It is a two-step approach: first, an eye-blink artifact is extracted from the contaminated single-channel EEG signal using SSA. Next, the extracted eye-blink artifact is denoised in the non-artifact region using CWT and $k-$ means algorithms.

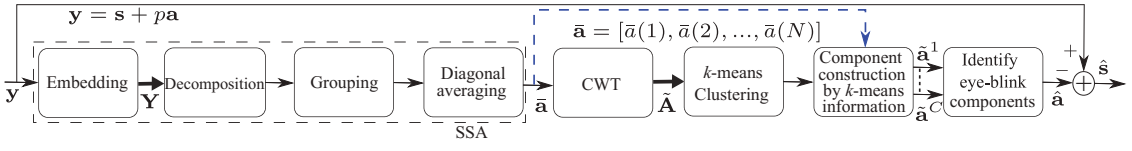


Figure 1. Block diagram of proposed method for eye-blink artifact removal from single EEG signals.

SSA is a data-driven technique employed to process the single-channel (uni-variate) time-series data [31,32]. Basically, the SSA technique comprises the following four steps: embedding, decomposition, grouping and diagonal averaging. Let us consider the contaminated EEG signal $\mathbf{y}$, which is a result of the mixing model shown as follows:

$$\mathbf{y} = \mathbf{s} + p\mathbf{a} \tag{7}$$

where $\mathbf{s}$ and $\mathbf{a}$ are the ground truth EEG and the eye-blink artifact signals, respectively, and p is an artifact mixing constant that changes the signal-to-noise ratio (SNR) of the measured EEG signal $\mathbf{y}$. When p is small (<1), the artifact contribution is less and results in a high SNR of the EEG signal $\mathbf{y}$ and vice-versa for $p > 1$. The key steps of SSA are as follows: in the embedding step of SSA, the given N sampled single-channel EEG signal $\mathbf{y} = [y(1), y(2), \ldots, y(N)]$ is mapped into multivariate data matrix $\mathbf{Y}$.

$$\mathbf{Y} = \begin{bmatrix} y(1) & y(2) & \ldots\ldots & \ldots\ldots & y(K) \\ y(2) & y(3) & \ldots\ldots & \ldots\ldots & y(K+1) \\ \vdots & \vdots & \ldots\ldots & \ldots\ldots & \vdots \\ y(M) & y(M+1) & \ldots\ldots & \ldots\ldots & y(N) \end{bmatrix} \tag{8}$$

where M represents the window length and $K = N - M + 1$. The matrix in (8) is called the Hankel matrix, as its anti-diagonal elements are constant (same). From (7), we can write $\mathbf{Y} = \mathbf{S} + \mathbf{A}$ (assuming that $p = 1$), where $\mathbf{S}$ and $\mathbf{A}$ represent the trajectory matrices of the ground truth EEG and eye-blink artifact signals, respectively. Note that we have considered the artifact mixing constant $p = 1$ for a simple explanation.

In the decomposition step of SSA, the trajectory matrix $\mathbf{Y}$ is decomposed into M trajectory matrices, for example, $\mathbf{Y}_1, \mathbf{Y}_2, \ldots, \mathbf{Y}_M$. Hence, the singular value decomposition (SVD) of $\mathbf{Y} = \mathbf{U}\mathbf{D}\mathbf{V}^T$ will be performed, where $\mathbf{D}$ represents the diagonal matrix whose elements are singular values and $\mathbf{U}$ and $\mathbf{V}$ are left and right singular matrices, whose columns are the eigenvectors of covariance matrix $\mathbf{C} = \mathbf{Y}\mathbf{Y}^T$ and $\mathbf{C} = \mathbf{Y}^T\mathbf{Y}$, respectively. However, direct decomposition of $\mathbf{Y}$ using SVD will increase the computational complexity. Therefore, the eigen decomposition of the covariance matrix of $\mathbf{C} = \mathbf{Y}\mathbf{Y}^T$ will be performed first.

Let us consider that $\lambda_1, \lambda_2, \ldots, \lambda_M$ and $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_M$ represent the eigenvalues and the eigenvectors of the covariance matrix $\mathbf{C} = \mathbf{Y}\mathbf{Y}^T$. Moreover, we assume that the eigenvalues are sorted in the descending order of their amplitudes, $\lambda_1 \geq \lambda_2 \geq, \ldots, \geq \lambda_M \geq 0$. Then, the jth trajectory matrix $\mathbf{Y}_j$ can be represented as

$$\mathbf{Y}_j = \sqrt{\lambda_j}\mathbf{u}_j\mathbf{v}_j^T \qquad j = 1, 2, \ldots, M \tag{9}$$

from the SVD of $\mathbf{Y}$, $\mathbf{v}_j = \mathbf{Y}^T\mathbf{u}_j/\sqrt{\lambda_j}$. Substituting $\mathbf{v}_j$ in (9), then the jth trajectory matrix $\mathbf{Y}_j$ can be represented as

$$\mathbf{Y}_j = \mathbf{u}_j\mathbf{u}_j^T\mathbf{Y} \tag{10}$$

The terms $\mathbf{u}_j\mathbf{u}_j^T$ in (10) form a subspace to reconstruct the jth component from the given signal $\mathbf{y}$.

The main goal in the grouping step of SSA is to construct the eye-blink associated trajectory matrix $\mathbf{A}$ from M trajectory matrices $\mathbf{Y}_j, j = 1, 2, \ldots, M$. Basically, we try to identify the appropriate eigenvectors by which we can construct an eye-blink artifact-associated trajectory matrix $\mathbf{A}$. In the classical SSA technique, the eigenvectors that are used to construct the eye-blink artifact are identified based on the strength of the eigenvalues (eigen spectrum) of a covariance matrix $\mathbf{C}$ [36]. However, in this work, we have identified these eigenvectors based on the local mobility or Hjorth mobility [44], which is a signal complexity measure of each eigenvector [38]. Here, the hypothesis is that the local mobility of the eigenvectors corresponding to the eye-blink artifact is low and is high for eigenvectors associated with EEG signals. Therefore, the pre-defined threshold has to be set to identify the eigenvectors associated to the eye-blink artifact. In fact, finding the eigenvectors associated with the artifact is similar to identifying the artifact signal subspace from the given signal space. The parameter for identifying the eye-blink artifact subspace is computed as follows: as the eigenvector holds the variation of the data, first, M sampled sinusoidal signal of frequency f is generated. Next, the local mobility of the sinusoidal signal is computed and it will be used as a threshold. As the threshold, which is used to identify the artifact subspace, is proportional to the frequency, it is denoted with variable f. The threshold parameter f of SSA will be acting as a cut-off frequency as in the case of a low-pass filter. After identifying the eigenvectors (basis functions) associated with an eye-blink artifact, the trajectory matrix corresponding to an eye-blink artifact ($\bar{\mathbf{A}}$) is computed using (10).

In fact, the computed trajectory matrix $\bar{\mathbf{A}}$ that resulted from the grouping step of SSA will not hold the Hankel structure. In the diagonal averaging step of SSA, the anti-diagonal elements are replaced with their average, and the uni-variate signal $\bar{\mathbf{a}}$ will be constructed using (11), as follows:

$$\bar{a}(n) = \begin{cases} \frac{1}{n}\sum_{i=1}^{n}\bar{\mathbf{A}}(i, n-i+1) & \text{for } 1 \geq n < M \\ \frac{1}{M}\sum_{i=1}^{M}\bar{\mathbf{A}}(i, n-i+1) & \text{for } M \geq n \leq K \\ \frac{1}{N-n+1}\sum_{i=n-K+1}^{N-K+1}\bar{\mathbf{A}}(i, n-i+1) & \text{for } K < n \leq N \end{cases} \tag{11}$$

The extracted eye-blink artifacts $\bar{\mathbf{a}}$ from the SSA method contain low-frequency EEG components. The direct subtraction of the extracted eye-blink artifact ($\bar{\mathbf{a}}$) from the contaminated signal $\mathbf{y}$ results in a loss of low-frequency components in the reconstructed EEG signal. Therefore, denoising of these components from $\bar{\mathbf{a}}$ has to be performed before it is subtracted from the contaminated EEG signal $\mathbf{y}$.

Denoising the EEG Components from the Extracted Eye-Blink Artifact ($\bar{\mathbf{a}}$)

In order to denoise the EEG components in the extracted eye-blink artifact $\bar{\mathbf{a}}$, we proposed a new methodology. In this method, the time-frequency representation of $\bar{\mathbf{a}}$,

which is the output of the SSA block, is performed using CWT, and it results in a matrix $\tilde{\mathbf{A}}$ of size $L \times N$ and is denoted by

$$|\tilde{\mathbf{A}}| = \begin{bmatrix} \tilde{a}(1,1) & \ldots\ldots & \tilde{a}(1,j) & \ldots\ldots & \tilde{a}(1,N) \\ \tilde{a}(2,1) & \ldots\ldots & \tilde{a}(2,j) & \ldots\ldots & \tilde{a}(2,N) \\ \vdots & \vdots & \ldots\ldots & \ldots\ldots & \vdots \\ \tilde{a}(L,1) & \ldots\ldots & \tilde{a}(L,j) & \ldots\ldots & \tilde{a}(L,N) \end{bmatrix} = [\tilde{\mathbf{a}}_1, \ldots, \tilde{\mathbf{a}}_j, \ldots, \tilde{\mathbf{a}}_N]$$

where L is the number of frequencies for which CWT is computed. Each column vector $\tilde{\mathbf{a}}_j$ $(j = 1, 2, \ldots, N)$ of $|\tilde{\mathbf{A}}|$ represents the feature vector of jth sample of $\bar{\mathbf{a}}$. Next, each column vector of $|\tilde{\mathbf{A}}|$ is clustered using k-means clustering algorithm with C number of clusters. Then, k-means algorithm provides the labels for each feature vector of $|\tilde{\mathbf{A}}|$. These labels inform to which cluster a particular feature vector (indirectly the sample of $\bar{\mathbf{a}}$) has fallen. With this clustering information, we construct C number of signals using (12)

$$\tilde{a}^i(j) = \begin{cases} \bar{a}(j) & \text{if } \tilde{\mathbf{a}}_j \in C_i,\ i = 1, 2, \ldots, C \,\&\, j = 1, 2, \ldots, N \\ 0 & \text{if } \tilde{\mathbf{a}}_j \notin C_i \end{cases} \tag{12}$$

Here, $\tilde{\mathbf{a}}_j$ represents the jth column vector of matrix $|\tilde{\mathbf{A}}|$. After decomposing the signal $\bar{\mathbf{a}}$ into C number of signals, say $\tilde{\mathbf{a}}^1$, $\tilde{\mathbf{a}}^2$, ……, $\tilde{\mathbf{a}}^C$ using (12), then, the fractal dimension (FD) [45] of each component is computed to identify the eye-blink artifact associated component. The estimated eye-blink artifact ($\hat{\mathbf{a}}$) is identified based on the FD; usually, it is low for denoised eye-blink artifacts. Finally, the corrected EEG signal ($\hat{\mathbf{s}}$) is obtained by subtracting the estimated eye-blink artifact $\hat{\mathbf{a}}$ from $\mathbf{y}$.

4. Results

To evaluate the performance of the proposed and the existing methods, we have constructed synthetically contaminated EEG signals from fatigue EEG data [46,47].

4.1. Construction of Synthetically Contaminated EEG Signal and Eye-Blink Artifact

We have considered 10 subjects' EEG data from the Fatigue EEG database [46,47]. Each subject performed a driving task on a static simulator. The EEG data were recorded in two phases normal and fatigue states using a 32-channel electrode cap with a sampling frequency of 1000 Hz. More details about the EEG data are discussed in [46,47]. In the construction of a true EEG signal for the simulation study, first, the raw EEG data measured from Fp_1 channel of ten subjects is down sampled to 250 Hz from 1000 Hz. Next, the baseline drift and the high-frequency components in the EEG data are removed using a band-pass filter with cut-off frequencies of 1 and 45 Hz. However, for synthetic simulation, a 10 s artifact-free EEG epoch is segmented from the filtered EEG data. These artifact-free EEG epochs are served as true EEG signals $(\mathbf{s})$ for a synthetic simulation study. The synthetic eye-blink artifact data were constructed as follows: first, we identified the eye-blink artifact region manually and segmented it from the EEG signal. Next, zeros were padded to the segmented eye-blink component on both sides such that the length of the signal is 10 s. In order to remove the EEG remnants present on the eye-blink component, MATLAB *smooth* command was used. This results in the ground truth eye-blink artifact signal $(\mathbf{a})$. We have constructed five such eye-blink artifacts from five subjects. Using these five eye-blink artifacts and ten EEG signals, we constructed a total of 50 synthetically contaminated EEG signals $(\mathbf{y})$. However, we assumed that the contaminated EEG signal is additive mixing of both the true EEG signal and the eye-blink artifact, i.e., $\mathbf{y} = \mathbf{s} + p\mathbf{a}$. Here, the artifact mixing constant p changes the SNR of the EEG signal. When the artifact mixing constant is $p > 1$, the eye-blink artifact contribution in the contaminated EEG signal is high, and as a result, the SNR of the EEG signal is low. When $p < 1$, the eye-blink artifact contribution in the contaminated EEG signal is low, and as a result, the SNR of the EEG signal is high. Figure 2 shows the synthetically constructed ground truth EEG, the eye-blink artifact and the contaminated EEG signals for $p = 0.5$.

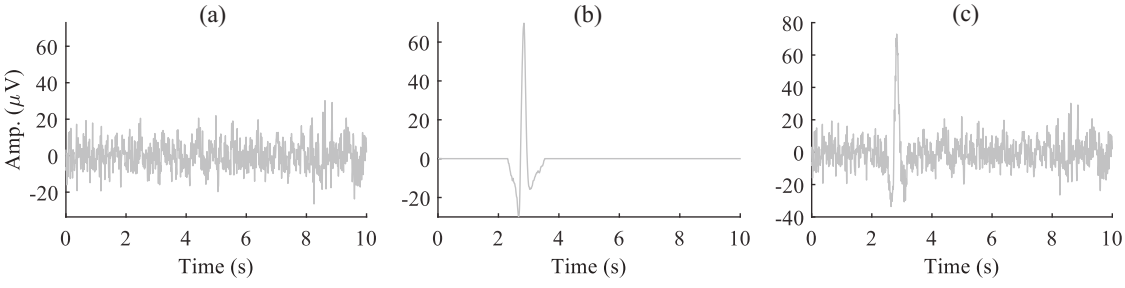


Figure 2. (**a**,**b**) Synthetically constructed ground truth EEG signal ($\mathbf{s}$) and the EOG artifact (**a**), respectively, and (**c**) the contaminated EEG signal $\mathbf{x} = \mathbf{s} + p\mathbf{a}$ for the artifact mixing constant $p = 0.5$.

4.2. Parameter Settings for Proposed and Existing Methods

The faithful reconstruction of eye-blink artifact components from the contaminated EEG signal depends on the SSA window length and the parameter f that identifies the artifact subspace. Therefore, we have performed simulations to select these parameters. Figure 3 shows the effect of the parameter f and the window length M in extracting the eye-blink artifact. We have identified the region of the eye-blink component (the artifact region only) from the fifty extracted eye-blink artifact signals by SSA and computed the mean eye-blink artifact component. Figure 3a–c shows the mean eye-blink artifact component (the artifact region) of $\bar{\mathbf{a}}$ obtained by the SSA method for window lengths $M = 22, 32$ and 64 and the parameter $f = 4, 6, 8, 10$ and 12 Hz. We have noticed from Figure 3a–c that the performance of SSA with $f = 4$ Hz is low for different window lengths $M = 22, 32$ and 64. However, the performance of SSA with $M = 32$ and 64 is stable for $f = 6, 8, 10$ and 12 Hz, as evident from Figure 3d. The RRMSE curves were plotted with respect to the mean ground truth eye-blink artifact **a**. Based on the results in Figure 3d, the parameters of the SSA method, f and the window length M are set to 8 Hz and 64, respectively, to obtain better performance. For the proposed denoising methodology, Morlet wavelet transform has been used to represent the eye-blink artifact obtained by SSA into its time–frequency feature matrix, which is then given as input to the k-means clustering algorithm. In order to map the eye-blink artifact component into its time–frequency representation, we compute the wavelet coefficients in the range of 1 to 12 Hz with an increment of 0.25 Hz. This results in a feature matrix of size 45×2500. Such representation maps each sample of the eye-blink artifact estimated by SSA into a high-dimensional feature vector of size 45×1. It was clear from Figure 1 that the number of components ($\tilde{\mathbf{a}}^1, \tilde{\mathbf{a}}^2, \ldots, \tilde{\mathbf{a}}^C$) constructed using k-means information also increased when the number of clusters (C) increases. As the eye-blink artifact is a strong component, setting the number of clusters to 2 displayed better performance on short EEG epochs. Hence, we set the number of clusters to 2 for the proposed method. Based on the recommendations in [42], the parameters of the k-means+SSA method, the window length and thresholds T_h and T_{SSA} are set to 125, 1.4 and 0.01, respectively. In the case of SSA+ANC, we identified better performance with window length 40. Whereas in the case of the VME-DWT method, the α parameter is set to 1000 and the other parameters are fixed as in [29].

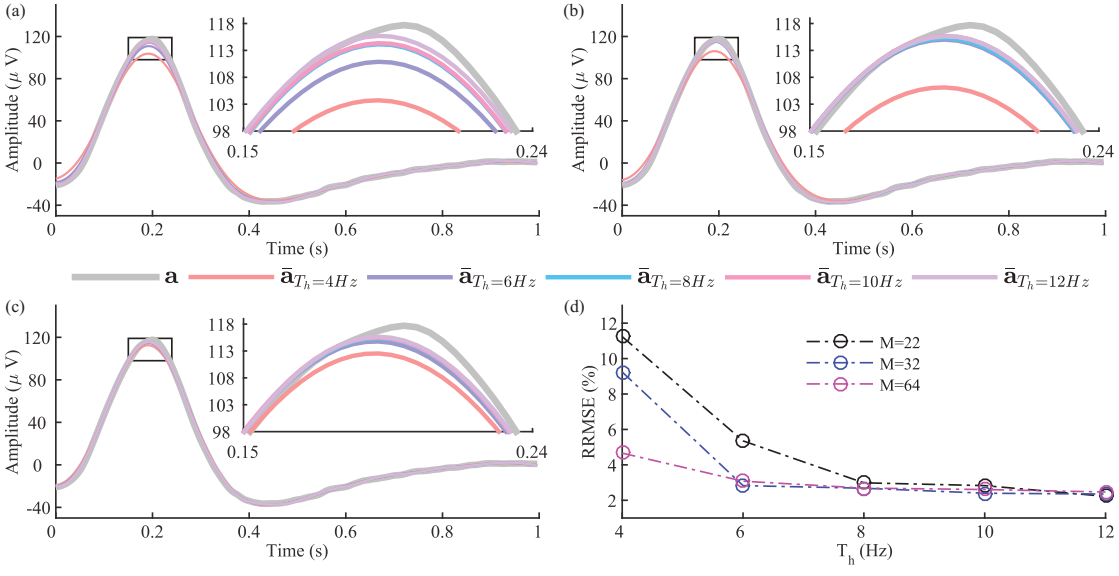


Figure 3. The estimated eye-blink artifacts $\bar{\mathbf{a}}$ (the artifact region only) by SSA with different thresholds and the window lengths (**a**) $M = 22$, (**b**) $M = 32$, and (**c**) $M = 64$. (**d**) Performance of SSA in terms of RRMSE for varying window length M and thresholds ($f = 4, 6, 8, 10$ and 12 Hz). The RRMSEs were calculated with respect to the ground truth eye-blink signal, **a**.

4.3. Results with Synthetic EEG Signals

The time–frequency representation of the extracted eye-blink artifact $\bar{\mathbf{a}}$, in Figure 4a obtained by SSA, is shown in Figure 4b. As the eye-blink component is a strong component in $\bar{\mathbf{a}}$, also evident from Figure 4a, the feature vectors of the time–frequency matrix (Figure 4b) between 2.5 and 3.5 s are significantly different. It is clear from the clustering information, shown in Figure 4c that all of the feature vectors (the columns of time-frequency map) corresponding to the eye-blink artifact region belong to cluster 2. The features vectors that correspond to the non-artifact region belong to cluster 1. By computing (12), we have obtained two signals $\tilde{\mathbf{a}}^1$ and $\tilde{\mathbf{a}}^2$, (as $C = 2$). We have computed the FD of these two components to identify the eye-blink artifact. As the eye-blink artifact is a low-frequency component, we expect its corresponding FD to be a low value. Finally, the denoised eye-blink component is identified based on their FD value. The estimated eye-blink artifact and the corrected EEG signals using the proposed and the existing methods are shown in Figure 5. Even though the SSA and SSA+ANC methods extracted the eye-blink artifact very well, they also extracted the low-frequency EEG information from the non-artifact regions, as shown in Figure 5a. Although VME-DWT does not alter the non-artifact regions, it removed the eye-blink artifact partially (see circled region), whereas the *k*-means+SSA method removes valuable EEG information (see the circled region in the fourth row). In contrast, it is also clear from Figure 5b that there is no loss of EEG information with the proposed method. The RRMSE, the CC, the artifact reduction ratio (λ) and MAE values shown in Figure 5 also reveal the superiority of the proposed method over the existing methods. We also computed the power spectrums of the true EEG, the contaminated EEG and the corrected EEG signals to observe any spectral changes in the EEG signal after the artifact removal. Figure 5c–g shows the superposition plots of the true EEG, contaminated EEG and the corrected EEG signals using all methods. It can be observed from the power spectrum plots of the true and corrected EEG signals that the proposed method almost preserves the low-frequency information of the EEG signal as compared with the existing methods.

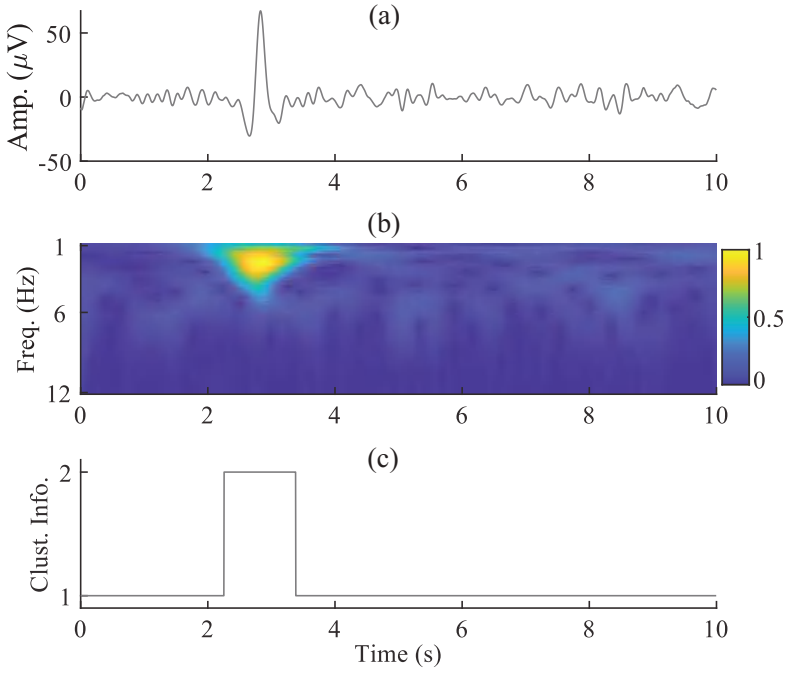


Figure 4. (**a**) The extracted eye-blink artifact ($\bar{\mathbf{a}}$) by SSA, (**b**) its time–frequency representation using CWT (normalized plot) and (**c**) clustering information.

We have applied the proposed method to remove the eye-blink artifacts from 50 synthetically contaminated EEG signals. Figure 6 shows the RRMSE, the CC, the artifact reduction ratio (λ) and the MAE plots obtained by the application of the existing and the proposed techniques over 50 EEG records for different artifact mixing constants (p). As discussed earlier, the artifact mixing constant p alters the SNR of the EEG signal. When $p > 1$, the SNR of the EEG signal is low, whereas the SNR of the EEG signal is high for $p < 1$. Removing the eye-blink artifact is a challenging task when its contribution in the contaminated EEG signal is low (i.e., $p < 1$). The relation between p and SNR of the EEG signal is inversely proportional. The RRMSE and the CC values are computed with respect to the ground truth eye-blink artifacts. Whereas, the artifact reduction ratio (λ) and MAE values are computed with respect to the ground-truth EEG signals. It is clear from Figure 6a–d that in all conditions, the mean RRMSE, the CC, artifact reduction ratio and MAE values of the proposed method show better performance over SSA, SSA+ANC and VME-DWT methods. Although the VME-DWT showed comparative performance with the proposed method (see MAE plot) for $p < 1$, its performance is poor for $p \geq 1$. Furthermore, the performance of the proposed method is better as compared to *k*-means+SSA for $p < 1$ condition. Although the performance of the *k*-means+SSA method is comparable with the proposed method for $p \geq 1$, its performance is not stable due to the threshold parameters T_h and T_{SSA}.

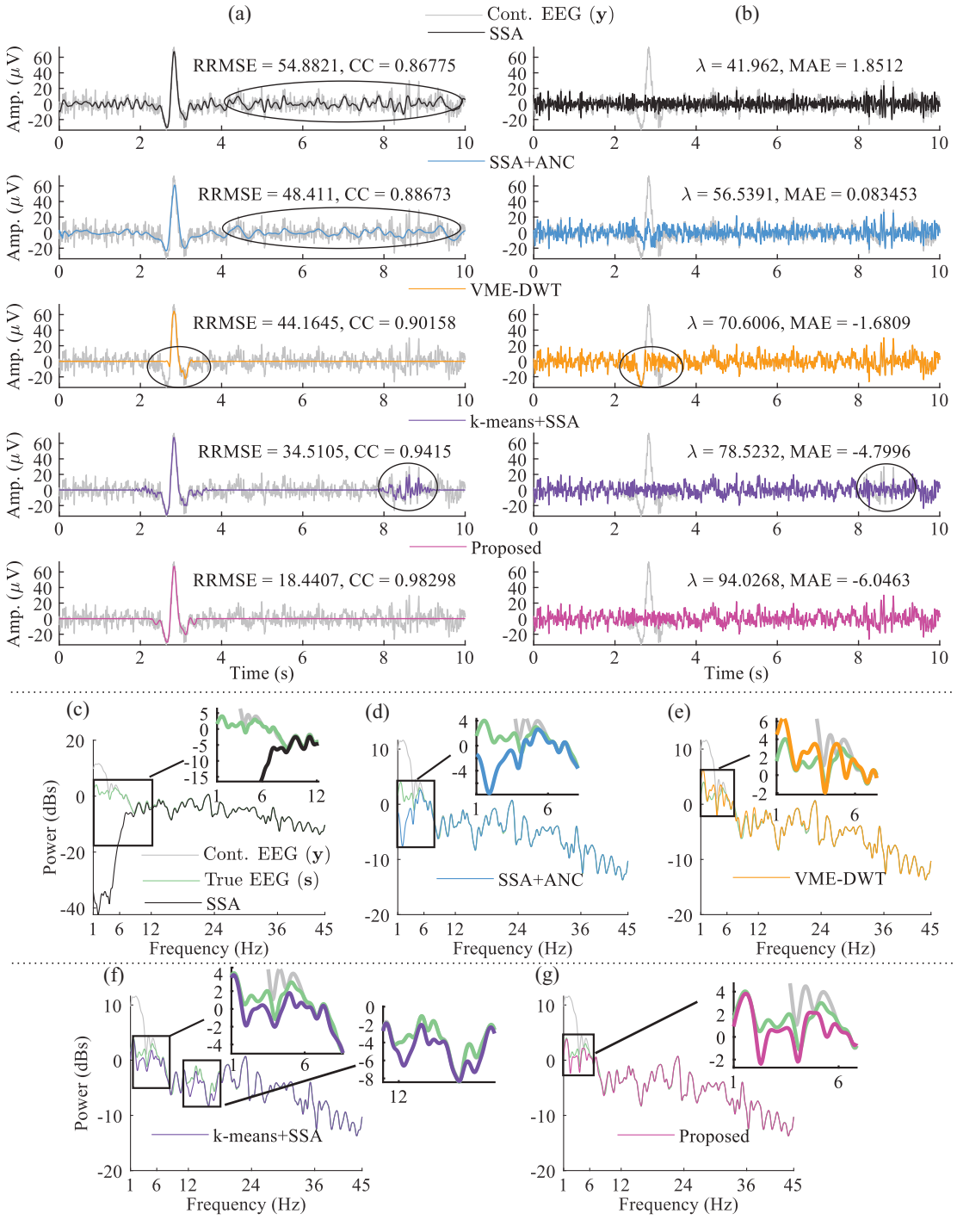


Figure 5. (**a**) The estimated eye-blink artifact (**â**), (**b**) the corrected EEG signals (**ŝ**) using all methods, for the artifact mixing constant $p = 0.5$. (**c**–**g**) the power spectrums of the true EEG (**s**), the contaminated EEG (**y**), and the corrected EEG signals of all methods.

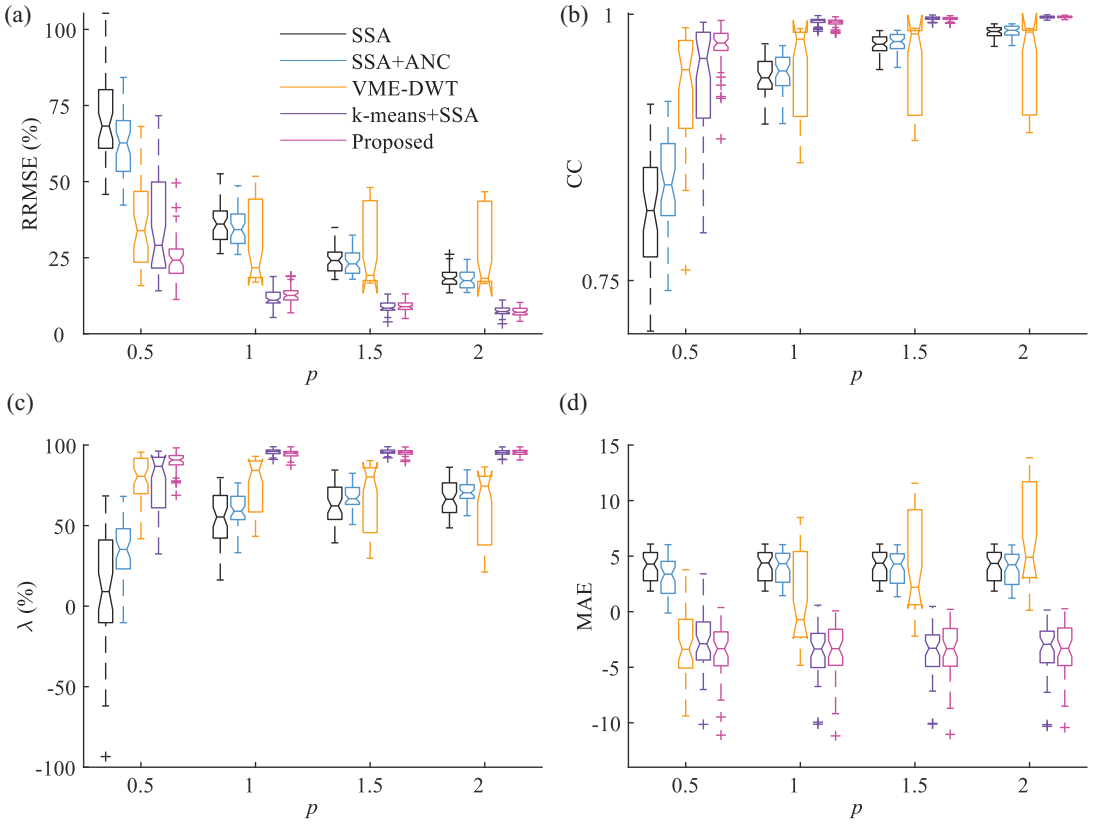


Figure 6. Performance of the existing and the proposed methods in terms of (**a**) RRMSE, (**b**) CC, (**c**) artifact removal ratio (λ) and (**d**) MAE (in log scale) with respect to the artifact mixing constant p. (The artifact mixing constant p is $\propto 1/SNR$) of the signal.

4.4. Results with Real EEG Signals

To evaluate the performance of the proposed method on the real EEG signals from the Fatigue EEG dataset (Fatigue EEG DB), we have segmented 50 EEG epochs of length 10 s from ten subjects' lengthy EEG records [46,47]. Note that the data are re-sampled to 250 from 1000 Hz. Similarly, from the EEG Motor Movement/Imagery Database (EEG-MMI DB), an EEG epoch of 10 s from the lengthy EEG signal (eyes open task) obtained from 65 subjects is segmented [48,49]. The sampling frequency of this dataset is 160 Hz. For both datasets, the segmentation of the EEG epoch is performed such that at least one eye-blink artifact component is present in the segmented EEG epoch. From these two datasets, we have constructed in total 105 EEG epochs of length 10 s and evaluated the performance of the proposed and existing methods. In fact, for real EEG signals there will be no ground-truth EEG to evaluate the performance. Hence, we manually indicated the non-artifact and artifact intervals of each record and computed RRMSE and CC values.

The estimated eye-blink artifact and the corrected EEG signals (Fatigue EEG data) with all the methods are shown in Figure 7a,b. From Figure 7a, we can see that the low-frequency components are still present in eye-blink artifacts obtained by the SSA and SSA+ANC methods (see the non-artifact region between 1–4 s). As a result, low-frequency EEG information is removed from the corrected EEG signal obtained by the SSA and SSA+ANC, as shown in Figure 7b, whereas the VME-DWT method partially removed the eye-blink artifact and altered the non-artifact region in the time interval 2–3 s. The *k*-means+SSA

method also altered the non-artifact region of the EEG signals in time interval 1–4 s (as indicated by circles in Figure 7b). The corrected EEG signal obtained by the *k*-means+SSA method and the contaminated EEG signals do not match in the non-artifact region (see 1–4 s in Figure 7b time interval). However, the corrected EEG signal obtained by the proposed method perfectly matches with the non-artifact region of the contaminated EEG signal, as shown in Figure 7b. The RRMSE and CC values shows the superiority of the proposed method over the existing methods.

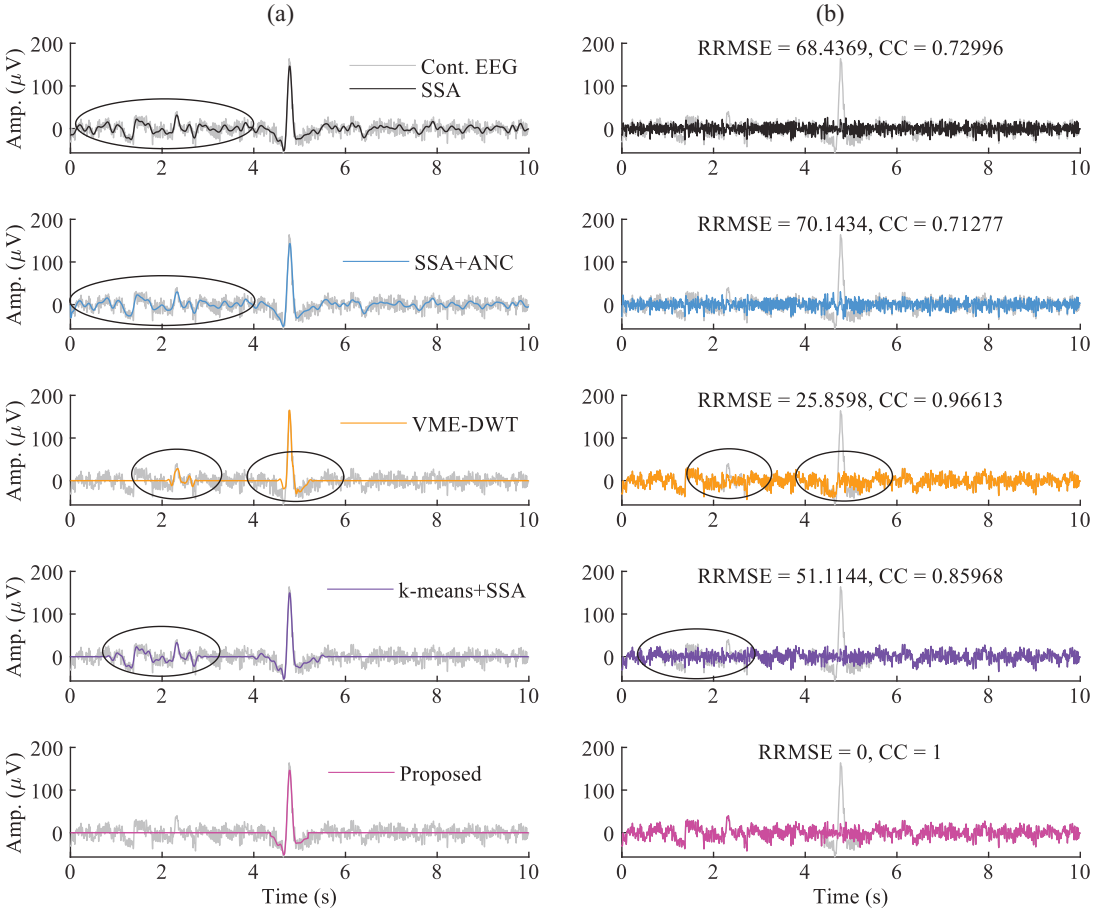


Figure 7. (**a**) The estimated eye-blink artifact ($\hat{\mathbf{a}}$) and (**b**) the corrected EEG signals ($\hat{\mathbf{s}}$) from the contaminated EEG signal (**y**) using the existing and the proposed methods.

As we do not have ground truth EEG signals for real EEG datasets, it is difficult to assess the performance of the proposed and existing methods in the frequency domain (power spectrum). However, the manually identified non-artifact and artifact regions of the EEG epoch are used to evaluate the performance of all methods in-terms of RRMSE and CC values. Table 1 shows the RRMSE and CC (mean $\pm$ standard deviation) values of the proposed method. Moreover, two binary classifier performance measures, such as precision and accuracy are also computed to evaluate the performance. Table 2 shows the mean precision and accuracy values of VME-DWT, *k*-means+SSA and proposed methods. It is also evident from Tables 1 and 2 that the proposed method shows superior performance over the existing methods.

Table 1. RRMSE and CC ($\mu \pm \sigma$) comparison between the non-artifact interval of contaminated and corrected EEG signals.

Measures and Methods	Fatigue EEG DB		EEG-MMI DB	
	RRMSE	**CC**	**RRMSE**	**CC**
SSA	63.6077 ± 11.6133	0.7576 ± 0.1068	71.4361 ± 11.9799	0.6831 ± 0.1106
SSA+ANC	61.4721 ± 9.3272	0.7815 ± 0.0764	61.5249 ± 11.2345	0.7775 ± 0.0862
VME-DWT	6.7885 ± 13.3722	0.9885 ± 0.0283	5.9036 ± 10.9759	0.9922 ± 0.0164
k-means+SSA	16.3888 ± 16.4607	0.9713 ± 0.0598	16.0701 ± 13.7886	0.9770 ± 0.0361
Proposed	**4.9198 ± 7.4213**	**0.9960 ± 0.0139**	**2.9976 ± 7.3030**	**0.9969 ± 0.0104**

Table 2. Comparison of precision and accuracy ($\mu \pm \sigma$) of the proposed method with existing methods for eye-blink detection on two real EEG datasets.

Measures and Methods	Fatigue EEG DB		EEG-MMI DB	
	Precision (%)	**Accuracy (%)**	**Precision (%)**	**Accuracy (%)**
VME-DWT	80.0445 ± 14.9771	93.8336 ± 5.0842	72.0040 ± 14.1527	92.8067 ± 4.9993
k-means-DWT	55.5252 ± 12.4375	82.7320 ± 8.7630	57.5738 ± 11.3917	86.8750 ± 7.8568
Proposed	**96.1604 ± 4.3639**	**94.2760 ± 6.3941**	**98.8142 ± 3.4201**	**95.4538 ± 2.6401**

5. Discussion

Even though the SSA and SSA+ANC methods extract the eye-blink artifact component efficiently, they also alter the low-frequency component of the EEG signal in the non-artifact region (from Figures 5a and 7a). However, subtracting the estimated eye-blink artifact directly from the contaminated EEG signal will also remove the low-frequency components (0.5–12 Hz) of the EEG signal. This can be a cause of concern in applications such as driver fatigue detection, where the spectral energy of low-frequency EEG components is used to detect the fatigue level [50]. The use of low-frequency EEG components to detect hand movements of subjects with spinal cord injury has been studied in [51,52]. In a recent study, it is found that the low-frequency EEG oscillations could be used as a biomarker of stroke injury and recovery [53]. Moreover, eye-blink component features (the frequency, amplitude and phase) are also used in applications such as control of hand exoskeleton for the paralyzed hand [54–56]. Therefore, in order to preserve these important low-frequency components at the pre-processing step, we combined SSA with CWT and *k*-means algorithms. The results show that the proposed method preserves these components while removing the eye-blink artifact. As the eye-blink artifact is a high amplitude component in the EEG signal (particularly in pre-frontal EEG channels), the proposed method has exploited this inherent feature to remove the eye-blink artifact without altering the original EEG components. Although the VME-DWT method does not alter the non-artifact intervals of EEG, it failed to remove the eye-blink artifact completely. Even though the *k*-means+SSA method displayed comparable performance as compared to the proposed method for a few EEG records, for cases where the eye-blink artifact is stronger, the proposed method fared well in overall performance. In this present study, we have only considered pre-frontal EEG channel signals. However, it can be expected from the results that the performance of the proposed method will be degraded further when the amplitude of the eye-blink artifact that is mixed in the EEG signal is low and this will be our topic of future research. For example, the eye-blink artifact contribution is low on fronto-central EEG channels FC_x.

6. Conclusions

In this paper, we combined SSA with CWT and the *k*-means algorithms to preserve the low-frequency EEG information in the artifact removal process. As the eye-blink artifact

appears as a slow-time varying and strong component in the contaminated EEG signal, the proposed method exploited this feature to remove eye-blink artifacts from a single-channel EEG signal. The proposed method is evaluated on one synthetic and two real EEG datasets, and results show superior performance over existing techniques. Results also show the advantage of integrating SSA with CWT and *k*-means for eye-blink artifact removal from single-channel EEG signal. Since the present study considered the artifact removal from pre-frontal channel EEG signals, with the integration of available artifact detection algorithms, the proposed method could be employed for online applications where the pre-frontal EEG channel is used. Results show that the proposed method was successful in removing the eye blink artifact without the loss of original EEG information. Although the classification problem using the proposed method was not studied in the paper, we foresee that the proposed method will offer good performance in the final application.

Author Contributions: K.C.V. devised the protocol and setup. A.K.M. performed all the experiments, wrote the manuscript and prepared all the figures. K.C.V. performed proofreading and corrections for this manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Brain Pool Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT (2019H1D3A1A01068799).

Data Availability Statement: The EEG data and the MATLAB codes employed in this article will be made available by the authors upon request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Kutafina, E.; Heiligers, A.; Popovic, R.; Brenner, A.; Hankammer, B.; Jonas, S.M.; Mathiak, K.; Zweerings, J. Tracking of Mental Workload with a Mobile EEG Sensor. *Sensors* **2021**, *21*, 5205. [CrossRef] [PubMed]
2. Aminov, A.; Rogers, J.M.; Johnstone, S.J.; Middleton, S.; Wilson, P.H. Acute single channel EEG predictors of cognitive function after strok. *PLoS ONE* **2017**, *12*, e0185841. [CrossRef]
3. Guo, Z.; Pan, Y.; Zhao, G.; Cao, S.; Zhang, J. Detection of driver vigilance level using *EEG* Signals and driving contexts. *IEEE Trans. Reliab.* **2017**, *67*, 370–380. [CrossRef]
4. Noachtar, S.; Rémi, J. The role of *EEG* Epilepsy: A critical review. *Epilepsy Behav.* **2009**, *15*, 22–33. [CrossRef] [PubMed]
5. Wilkinson, C.M.; Burrell, J.I.; Kuziek, J.W.; Thirunavukkarasu, S.; Buck, B.H.; Mathewson, K.E. Predicting stroke severity with a 3-min recording from the Muse portable EEG system for rapid diagnosis of stroke. *Sci. Rep.* **2020**, *10*, 18465. [CrossRef]
6. Hagemann, D.; Naumann, E. The effects of ocular artifacts on (lateralized) broadband power in the *EEG*.*Clin. Neurophysiol.* **2001**, *112*, 215–231. [CrossRef]
7. Halder, S.; Bensch, M.; Mellinger, J.; Bogdan, M.; Kübler, A.; Birbaumer, N.; Rosenstiel, W. Online artifact removal for brain-computer interfaces using support vector machines and blind source separation. *Comput. Intell. Neurosci.* **2007**, *2007*, 82069. [CrossRef]
8. Schlögl, A.; Keinrath, C.; Zimmermann, D.; Scherer, R.; Leeb, R.; Pfurtscheller, G. A fully automated correction method of *EOG* Artifacts *EEG* Recordings. *Clin. Neurophysiol.* **2007**, *118*, 98–104. [CrossRef] [PubMed]
9. Jung, T.; Makeig, S.; Humphries, C.; Lee, T.; Mckeown, M.J.; Iragui, V.; Sejnowski, T.J. Removing electroencephalographic artifacts by blind source separation. *Psychophysiology* **2000**, *37*, 163–178. [CrossRef]
10. Vigário, R.; Sarela, J.; Jousmiki, V.; Hamalainen, M.; Oja, E. Independent component approach to the analysis of EEG and MEG recordings. *IEEE Trans. Biomed. Eng.* **2000**, *47*, 589–593. [CrossRef] [PubMed]
11. Delorme, A.; Sejnowski, T.; Makeig, S. Enhanced detection of artifacts in *EEG* data using higher-order statistics and independent component analysis. *Neuroimage* **2007**, *34*, 1443–1449. [CrossRef] [PubMed]
12. De Clercq, W.; Vergult, A.; Vanrumste, B.; Van Paesschen, W.; Van Huffel, S. Canonical Correlation Analysis Applied to Remove Muscle Artifacts from the Electroencephalogram. *IEEE Trans. Biomed. Eng.* **2006**, *53*, 2583–2587. [CrossRef] [PubMed]
13. Gao, J.; Zheng, C.; Wang, P. Online removal of muscle artifact from electroencephalogram signals based on canonical correlation analysis. *Clin. EEG Neurosci.* **2010**, *41*, 53–59. [CrossRef] [PubMed]
14. Castellanos, N.P.; Makarov, V.A. Recovering *EEG* Brain Signals: Artifact suppression with wavelet enhanced independent component analysis. *J. Neurosci. Methods* **2006**, *158*, 300–312. [CrossRef] [PubMed]
15. Wang, G.; Teng, C.; Li, K.; Zhang, Z.; Yan, X. The Removal of *EOG* Artifacts *EEG* Signals Using independent component analysis and multivariate empirical mode decomposition.*IEEE J. Biomed. Health Inform.* **2016**, *20*, 1301–1308. [CrossRef] [PubMed]
16. Issa, M.F.; Juhasz, Z. Improved *EOG* Artifact Removal Using Wavelet Enhanced Independent Component Analysis. *Brain Sci.* **2019**, *9*, 355. [CrossRef]

17. Mammone, N.; Morabito, F.C. Enhanced automatic wavelet independent component analysis for electroencephalographic artifact removal. *Entropy* **2014**, *16*, 6553–6572. [CrossRef]
18. Chang, C.Y.; Hsu, S.H.; Pion-Tonachini, L.; Jung, T.P. Evaluation of artifact subspace reconstruction for automatic *EEG* artifact removal. In Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 18–21 July 2018; pp. 1242–1245.
19. Chang, C.Y.; Hsu, S.H.; Pion-Tonachini, L.; Jung, T.P. Evaluation of artifact subspace reconstruction for automatic artifact components removal in multi-channel *EEG* recordings. *IEEE Trans. Biomed. Eng.* **2019**, *67*, 1114–1121. [CrossRef]
20. Mshali, H.; Lemlouma, T.; Moloney, M.; Magoni, D. A survey on health monitoring systems for health smart homes. *Int. J. Ind. Ergon.* **2018**, *66*, 26–56. [CrossRef]
21. Koley, B.; Dey, D. An ensemble system for automatic sleep stage classification using single channel *EEG* Signal. *Comput. Biol. Med.* **2012**, *42*, 1186–1195. [CrossRef]
22. Ogino, M.; Mitsukura, Y. Portable drowsiness detection through use of a prefrontal single-channel electroencephalogram. *Sensors* **2018**, *18*, 4477. [CrossRef] [PubMed]
23. Ogino, M.; Kanoga, S.; Muto, M.; Mitsukura, Y. Analysis of prefrontal single-channel *EEG* Data Portable auditory ERP-based brain-computer interfaces. *Front. Hum. Neurosci.* **2019**, *13*, 250. [CrossRef] [PubMed]
24. Grosselin, F.; Navarro-Sune, X.; Vozzi, A.; Pandremmenou, K.; De Vico Fallani, F.; Attal, Y.; Chavez, M. Quality assessment of single-channel *EEG* Wearable Devices. *Sensors* **2019**, *19*, 601. [CrossRef] [PubMed]
25. Rogers, J.M.; Johnstone, S.J.; Aminov, A.; Donnelly, J.; Wilson, P.H. Test-retest reliability of a single-channel, wireless *EEG* System. *Int. J. Psychophysiol.* **2016**, *106*, 87–96. [CrossRef]
26. He, P.; Wilson, G.; Russell, C. Removal of ocular artifacts from electro-encephalogram by adaptive filtering. *Med. Biol. Eng. Comput.* **2004**, *42*, 407–412. [CrossRef]
27. Peng, H.; Hu, B.; Shi, Q.; Ratcliffe, M.; Zhao, Q.; Qi, Y.; Gao, G. Removal of ocular artifacts in *EEG*—An improved approach combining DWT and ANC for portable applications. *IEEE J. Biomed. Health Inform.* **2013**, *17*, 600–607. [CrossRef]
28. Abd Rahman, F.; Othman, M. Real time eye blink artifacts removal in electroencephalogram using savitzky-golay referenced adaptive filtering. In Proceedings of the International Conference for Innovation in Biomedical Engineering and Life Sciences, Putrajaya, Malaysia, 6–8 December 2015; pp. 68–71.
29. Shahbakhti, M.; Beiramvand, M.; Nazari, M.; Broniec-Wójcik, A.; Augustyniak, P.; Rodrigues, A.S.; Wierzchon, M.; Marozas, V. VME-DWT: An efficient algorithm for detection and elimination of eye blink from short segments of single *EEG* Channel. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2021**, *29*, 408–417. [CrossRef]
30. Wu, Q.; Zhang, W.; Wang, Y.; Zhang, W.; Liu, X. Research on removal algorithm of *EOG* artifacts in single-channel EEG signals based on *CEEMDAN-BD*. *Comput. Methods Biomech. Biomed. Eng.* **2021**, *24*, 1368–1379. [CrossRef]
31. Golyandina, N.; Nekrutkin, V.; Zhigljavsky, A.A. *Analysis of Time Series Structure: SSA and Related Techniques*; CRC Press: Boca Raton, FL, USA, 2001.
32. Ghil, M.; Allen, M.; Dettinger, M.; Ide, K.; Kondrashov, D.; Mann, M.; Robertson, A.W.; Saunders, A.; Tian, Y.; Varadi, F.; et al. Advanced spectral methods for climatic time series. *Rev. Geophys.* **2002**, *40*, 3-1–3-41. [CrossRef]
33. Teixeira, A.; Tomé, A.; Lang, E.; Gruber, P.; da Silva, A.M. Automatic removal of high-amplitude artefacts from single-channel electroencephalograms. *Comput. Methods Programs Biomed.* **2006**, *83*, 125–138. [CrossRef]
34. Sanei, S.; Lee, T.K.M.; Abolghasemi, V. A New Adaptive Line Enhancer Based on Singular Spectrum Analysis. *IEEE Trans. Biomed. Eng.* **2012**, *59*, 428–434. [CrossRef] [PubMed]
35. Maddirala, A.K.; Shaik, R.A. Motion artifact removal from single channel electroencephalogram signals using singular spectrum analysis. *Biomed. Signal Process. Control* **2016**, *30*, 79–85. [CrossRef]
36. Mukhopadhyay, S.K.; Krishnan, S. A singular spectrum analysis-based model-free electrocardiogram denoising technique. *Comput. Methods Programs Biomed.* **2020**, *188*, 105304. [CrossRef] [PubMed]
37. Teixeira, A.R.; Tome, A.M.; Lang, E.W.; Gruber, P.; Martins da Silva, A. On the use of clustering and local singular spectrum analysis to remove ocular artifacts from electroencephalograms. In Proceedings of the 2005 IEEE International Joint Conference on Neural Networks, Montreal, QC, Canada, 31 July–4 August 2005; Volume 4, pp. 2514–2519.
38. Maddirala, A.K.; Shaik, R.A. Removal of EOG artifacts from single channel *EEG* Signals combined singular spectrum analysis and adaptive noise canceler. *IEEE Sens. J.* **2016**, *16*, 8279–8287. [CrossRef]
39. Noorbasha, S.K.; Sudha, G.F. Removal of *EOG* Artifacts Single Channel *EEG*—An efficient model combining overlap segmented ASSA and ANC. *Biomed. Signal Process. Control* **2020**, *60*, 101987. [CrossRef]
40. Maddirala, A.K.; Shaik, R.A. Separation of Sources from Single-Channel *EEG* Signals using independent component analysis. *IEEE Trans. Instrum. Meas.* **2018**, *67*, 382–393. [CrossRef]
41. Cheng, J.; Li, L.; Li, C.; Liu, Y.; Liu, A.; Qian, R.; Chen, X. Remove diverse artifacts simultaneously from a single-channel *EEG* Based *SSA* and *ICA*: A Semi-Simulated Study. *IEEE Access* **2019**, *7*, 60276–60289. [CrossRef]
42. Maddirala, A.K.; Veluvolu, K.C. Eye-blink artifact removal from single channel EEG with k-means and SSA. *Sci. Rep.* **2021**, *11*, 11043. [CrossRef]
43. Robbins, K.A.; Touryan, J.; Mullen, T.; Kothe, C.; Bigdely-Shamlo, N. How sensitive are *EEG* Results Preprocessing Methods: A Benchmarking Study. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2020**, *28*, 1081–1090. [CrossRef]
44. Hjorth, B. *EEG* analysis based on time domain properties. *Clin. Neurophysiol.* **1970**, *29*, 306–310. [CrossRef]

45. Sevcik, C. A procedure to Estimate the Fractal Dimension of Waveforms. *arXiv* **2010**, arXiv:nlin.CD/1003.5266.
46. Qiu, T. *Data for: Research on Fatigue Driving Detection Based on Adaptive Multi-Scale Entropy*; Mendeley Data: 2019. [CrossRef]
47. Luo, H.; Qiu, T.; Liu, C.; Huang, P. Research on fatigue driving detection using forehead EEG based on adaptive multi-scale entropy. *Biomed. Signal Process. Control* **2019**, *51*, 50–58. [CrossRef]
48. Schalk, G.; McFarland, D.; Hinterberger, T.; Birbaumer, N.; Wolpaw, J. BCI2000: A general-purpose brain-computer interface (*BCI*) System. *IEEE Trans. Biomed. Eng.* **2004**, *51*, 1034–1043. [CrossRef] [PubMed]
49. Goldberger, A.L.; Amaral, L.A.; Glass, L.; Hausdorff, J.M.; Ivanov, P.C.; Mark, R.G.; Mietus, J.E.; Moody, G.B.; Peng, C.K.; Stanley, H.E. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* **2000**, *101*, e215–e220. [CrossRef]
50. Jap, B.T.; Lal, S.; Fischer, P.; Bekiaris, E. Using EEG spectral components to assess algorithms for detecting fatigue. *Expert Syst. Appl.* **2009**, *36*, 2352–2359. [CrossRef]
51. Ofner, P.; Schwarz, A.; Pereira, J.; Wyss, D.; Wildburger, R.; Müller-Putz, G.R. Attempted arm and hand movements can be decoded from low-frequency *EEG* from persons with spinal cord injury. *Sci. Rep.* **2019**, *9*, 7134. [CrossRef]
52. Ofner, P.; Schwarz, A.; Pereira, J.; Müller-Putz, G.R. Upper limb movements can be decoded from the time-domain of low-frequency *EEG*. *PLoS ONE* **2017**, *12*, e0182578. [CrossRef]
53. Cassidy, J.M.; Wodeyar, A.; Wu, J.; Kaur, K.; Masuda, A.K.; Srinivasan, R.; Cramer, S.C. Low-Frequency Oscillations Are a Biomarker of Injury and Recovery After Stroke. *Stroke* **2020**, *51*, 1442–1450. [CrossRef]
54. Witkowski, M.; Cortese, M.; Cempini, M.; Mellinger, J.; Vitiello, N.; Soekadar, S.R. Enhancing brain-machine interface (*BMI*) Control A Hand Exoskeleton Using Electrooculography (*EOG*). *J. Neuroeng. Rehabil.* **2014**, *11*, 165. [CrossRef]
55. Soekadar, S.R.; Witkowski, M.; Vitiello, N.; Birbaumer, N. An *EEG/EOG*-Based hybrid brain-neural computer interaction (BNCI) system to control an exoskeleton for the paralyzed hand. *Biomed. Eng./Biomed. Tech.* **2015**, *60*, 199–205. [CrossRef]
56. Huang, Q.; Zhang, Z.; Yu, T.; He, S.; Li, Y. An *EEG-/EOG* hybrid brain-computer interface: Application on controlling an integrated wheelchair robotic arm system. *Front. Neurosci.* **2019**, *13*, 1243. [CrossRef] [PubMed]

sensors

MDPI

Article

Vital Signal Detection Using Multi-Radar for Reductions in Body Movement Effects

Ah-Jung Jang, In-Seong Lee and Jong-Ryul Yang *

Department of Electronic Engineering, Yeungnam University, Gyeongsan 38541, Korea; dkwjd289@yu.ac.kr (A.-J.J.); dldlstjd0322@yu.ac.kr (I.-S.L.)
* Correspondence: jryang@yu.ac.kr; Tel.: +82-53-810-2495

Abstract: Vital signal detection using multiple radars is proposed to reduce the signal degradation from a subject's body movement. The phase variation in the transceiving signals of continuous-wave radar due to respiration and heartbeat is generated by the body surface movement of the organs monitored in the line-of-sight (LOS) of the radar. The body movement signals obtained by two adjacent radars can be assumed to be the same over a certain distance. However, the vital signals are different in each radar, and each radar has a different LOS because of the asymmetric movement of lungs and heart. The proposed method uses two adjacent radars with different LOS to obtain correlated signals that reinforce the difference in the asymmetrical movement of the organs. The correlated signals can improve the signal-to-noise ratio in vital signal detection because of a reduction in the body movement effect. Two radars at different frequencies in the 5.8 GHz band are implemented to reduce direct signal coupling. Measurement results using the radars arranged at angles of 30°, 45°, and 60° showed that the proposed method can detect the vital signals with a mean accuracy of 97.8% for the subject moving at a maximum velocity of 53.4 mm/s.

Keywords: body movement cancelation; continuous wave; Doppler radar; multiple radars; vital signal detection; heartbeat; respiration

Citation: Jang, A.-J.; Lee, I.-S.; Yang, J.-R. Vital Signal Detection Using Multi-Radar for Reductions in Body Movement Effects. *Sensors* **2021**, *21*, 7398. https://doi.org/10.3390/s21217398

Academic Editor: Manuel José Cabral dos Santos Reis

Received: 7 October 2021
Accepted: 3 November 2021
Published: 7 November 2021

1. Introduction

The extraction of respiration and heartbeat signals from the variations in transceiving signal characteristics is a promising remote vital signal detection technology because it can escape the physical restraint of a contact sensor and be used in various applications [1–3]. A continuous wave (CW) Doppler radar, which monitors the Doppler frequency change in the CW signal caused by the thoracic movement from respiration and the heart's periodic movement, can acquire remote vital signals using a simple hardware configuration [4,5].

The effect of a subject's body movement must be removed for vital signal detection technology using a CW Doppler radar sensor to be commercialized for industrial and medical applications. A human body movement normally involves a larger displacement than movement caused by respiration (with a displacement of 4–12 mm) and heartbeat (with a displacement of 0.2–0.5 mm) on the body surface [6]. When a human body moves, a signal saturation may occur in a sensitive receiver for heartbeat detection because of its limited dynamic range, making it impossible to detect any signals [7]. Even when the radar has a sufficiently wide dynamic range, the frequency components of body movement can occupy a similar band to the frequencies resulting from respiration and heartbeat. As these components act as noise, they can deteriorate the signal-to-noise ratio (SNR) in vital signal detection or make the detection of vital signals impossible [7].

Previous studies on mitigating performance degradation in vital signal detection due to body movement can be divided into techniques for improving radar hardware configuration and signal-processing techniques [8–19]. Previous studies on radar hardware configuration separate the body movement and vital signals by measuring the directivity of body movement. Some studies use a plurality of radars to remove Doppler shifts

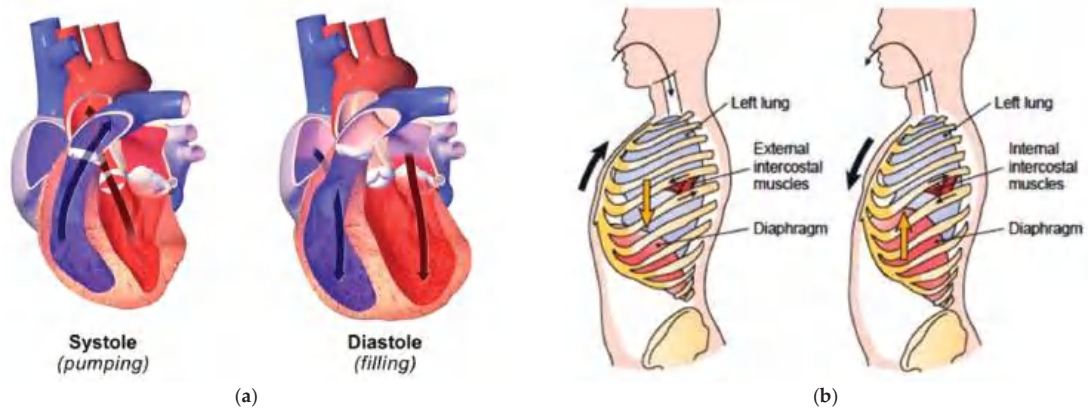
due to body movement by arranging them at positions facing each other around the subject and comparing the phase change in the baseband signals obtained from each radar [8–12]. However, these studies have a limitation in that it is difficult to consider the same movement in each radar because of the interference between radars and a change in the polarity of the received signal. Previous studies have shown that body movement can be canceled by fusion techniques using additional sensors, but they have a limitation in terms of their increased system complexity and implementation cost [13–16]. A signal-processing technique for minimizing the effect of body movement on radars is based on the compensation of the dominant baseband signal characteristics generated by body movement [17–19]. They have a limitation in that compensating for the effect of body movement because the cancelation performance can depend on the windowing size and time period of polynomial fitting. Although previous studies on removing the effect of body movement in vital signal detection using radars have been conducted in various directions, a technique for removing the effect of body movement has not been sufficiently explored.

In this study, the method of placing two independent radar sensors at the front for a certain angle of line-of-sight (LOS) is proposed to effectively compensate for the body movement characteristics and sensitively detect only vital signals based on the asymmetrical movement of internal organs. In the proposed radar configuration, two radars with different LOS are arranged in the same direction within a shorter distance than the wavelength of the operating frequency. The two radars use different operating frequencies to minimize direct coupling. It is assumed that the vital signals obtained by the two radars are different because of the asymmetric movement of organs, but the signal from body movement is approximately the same for each radar. The proposed configuration can improve the SNR of vital signal detection by removing the baseband signals from body movement. Section 2 describes the displacement difference due to the asymmetric movement of the heart and lungs, along with the proposed configuration and operating principle using multiple radars. The digital signal processing and hardware configuration to improve the SNR using a correlation between the two baseband signals of the radars are presented in Section 3. The measurement results and analyses are discussed in Section 4. Section 5 presents the conclusions of this study.

2. Proposed Configuration Using Multi-Radars

2.1. Physiological Movement of the Heart and Lungs

The heart and lungs inside a human body move in asymmetric directions with repeated contraction and expansion, as shown in Figure 1. The human heart in Figure 1a, which is divided into four parts (left and right atria and left and right ventricles) generates different displacements of the chest wall due to its different volumes and pressures [20–22]. In addition to the non-uniform characteristics of the human tissue layer consisting of various organs, muscles, and bones, the inherent directional movement of heart muscles caused by its various parts is asymmetrically monitored on the surface of the human body. As shown in Figure 1b, the lung movement during respiration is accompanied by the movement of the surrounding intercostal muscles, diaphragm, and the lung itself, resulting in a larger asymmetrical movement. The movement due to respiration, accompanied by the movement of the ribcage, is also generated anisotropically because the body volume depends on the contraction and expansion of the lungs [23,24]. The asymmetric movements of the human heart and lungs imply that a radar sensor to detect vital signals from varying surface displacements can be positioned along a specific direction to increase its detection sensitivity.



(**a**) (**b**)

Figure 1. Asymmetrical movements of human organs: (**a**) pumping and filling of blood in the heart; (**b**) expansion and contraction in the lungs [25,26].

The asymmetric movement of the heart and lungs on the surface of the human body was experimentally verified using two 5.8-GHz CW Doppler radars, as shown in Figure 2, to monitor respiration and heartbeat signals from a periodic displacement. The configuration of the CW Doppler radar module is described in Section 3. Two radars operating at the same frequency are placed between the subject in a line at a distance of 0.8 m to ensure their LOS is between the subject's front and rear. It was assumed that the vital signals obtained from each radar have dominant characteristics caused by the position of the human body. Unlike previous studies, in which radars were also placed on the left and right sides of the subject, in this study, there were no additional radars on either side to exclude the effect of minute movements of the subject's arms [9]. Figure 3 shows the vital signals that were simultaneously measured by the radars. Both the measured data displayed respiration and heartbeat signals at the same frequency, but the signal powers were measured differently between the two datasets, even though all components and conditions in the radars were identical. The respiration measured from the front radar (located at the front of the subject) had a higher intensity than that from the rear radar (located on the back of the subject). However, the heartbeat measured from the rear radar had a higher intensity than that from the front radar, even though the noise signal near DC was higher in the rear radar. The measurement results in Figure 3 show that the respiration and heartbeat signals measured by the radars, which detect the vital signal from the displacement of the human body surface, have an asymmetrical movement, as shown in several previous studies [8–10]. It is unreasonable to insist that the rear radar is more advantageous for heartbeat detection than the front radar based on the results in Figure 3. The SNR for heartbeat detection could decrease even though the amplitude of the heartbeat signal increased in the rear radar because it could increase the harmonic components of respiration and the noise near DC by increasing the nonlinearity of the received signal. The SNR could improve by using a signal correlation between the front and rear radars, because the body movement could be canceled by a displacement compensation using the quadrature signals of the configuration shown in Figure 3; however, the compensation performance is limited when the received signal from the rear radar is too small to detect the vital signals of the subject. The vital signals from the rear radar generally have lower power due to the small displacements based on the asymmetric movements of human organs when compared to the front radar, and their attenuation is more affected in the rear radar by the clothing conditions of the subject and the distance between the subject and the radar.

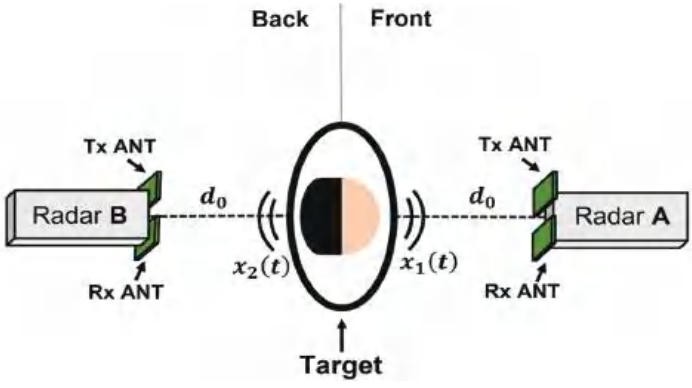


Figure 2. Preceding experiment showing that the respiration and heartbeat signals obtained from the radar can be different depending on the direction of the line-of-sight to the subject.

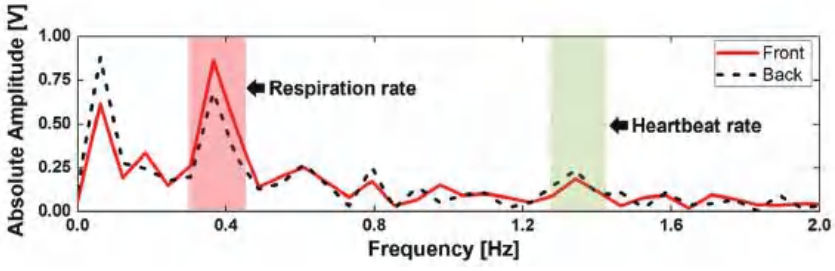


Figure 3. Frequency spectrum of the vital signals simultaneously measured from the two radars placed in front of and behind the subject.

2.2. Proposed Configuration for Vital Signal Detection Using Multiple Radars

A configuration using multiple radars, as shown in Figure 4, is proposed to increase the SNR of vital signal detection while considering the asymmetrical movements of human organs. The radar modules in the 5.8 GHz frequency band, which operate independently and consist of transmitting and receiving front-end and baseband circuits, are separately located at the same angle (θ) to the left and right and at the same distance (d_0) from the subject. The operating frequencies of the two radars are set to be different within the frequency band to reduce the degradation from a direct signal coupling between them and are arranged at an angle of 30° or more for a sufficient separation distance to reduce the increase in noise due to the blocker signal. When the operating frequencies of the two radars are different in the frequency band, the transmitted signals $T_1(t)$ and $T_2(t)$ from the two radars can be expressed as follows:

$$T_k(t) = A_{\mathrm{Tk}} \cdot \cos[2\pi f_k t + \theta_k(t)],\ k = 1.2, \tag{1}$$

where k is an index to discriminate the radar module, f_k is the operating frequency of each radar, A_{Tk} is the amplitude of the transmitted signals, and $\Delta\theta_k(t)$ is the phase noise generated from the signal source at the operating frequency. The vital signals generated by the asymmetrical movements of human organs are differently monitored in the two radars because of their different LOSs, and the received signals $R_k(t)$ in each radar can be expressed, except for the non-ideal characteristics such as the multipath and signal coupling, as follows:

$$R_k(t) = A_{\mathrm{Rk}} \cdot \cos\left[2\pi f_k t - \frac{4\pi d_0}{\lambda_k} - \frac{4\pi x_k(t)}{\lambda_k} \pm \frac{4\pi b(t)}{\lambda_k} + \theta_k\left(t - \frac{2d_0}{c}\right)\right],\ k = 1.2 \tag{2}$$

where A_{Rk} denotes the amplitudes of the received signals, c represents the propagation velocity of light in air, λ_k denotes the wavelength of the operating frequency, $x_k(t)$ denotes the displacement of the vital signals, and $b(t)$ denotes the displacement caused by human body movement. Assuming that the human body moves only in the forward and backward directions, the displacement $b(t)$ because of this movement can be equally expressed in both radars. The $\pm$ sign is used to indicate the human body movement direction, and the + and – signs, respectively, indicate movements approaching and moving away from the radar. Owing to the asymmetric movements of the heart and lungs and the different radar-operating frequencies, $x_k(t)$ can be expressed by distinguishing the amplitudes, phases, and frequencies as follows:

$$x_k(t) = x_{rk}(t) + x_{hk}(t) = m_{rk}\cos(\omega_r t + \varphi_{rk}) + m_{hk}\cos(\omega_h t + \varphi_{hk}) \quad (3)$$

where $x_{rk}(t)$ and $x_{hk}(t)$ are the displacements from respiration and heartbeat, respectively; m_{rk} and m_{hk} denote the magnitudes of respiration and heartbeat, respectively; w_r and w_h denote the angular frequencies of respiration and heartbeat, respectively; and φ_{rk} and φ_{hk} denote the phases of respiration and heartbeat, respectively. Although the radars are located at the same distance from the subject, the magnitudes m_{rk} and m_{hk} and the phases φ_{rk} and φ_{hk} are differently indicated because of the asymmetrical movement of the organs. The angular frequencies w_r and w_h can be assumed to be a single-frequency component because respiration and heartbeat signals at the surface of the human body are dominated by changes in the volume of the chest cavity and the left ventricle's movement, respectively [20,24]. After a down-conversion with a quadrature mixer and filtering with the low-pass filters, the baseband signals in the in-phase (I) and quadrature (Q) channels can be obtained as:

$$I_k(t) = A_{Ik}\cdot\cos\left[\frac{4\pi d_0}{\lambda_k} + \frac{4\pi x_k(t)}{\lambda_k} \mp \frac{4\pi b(t)}{\lambda_k} + \cdot\theta_k(t)\right] + DC_{Ik},\ k = 1.2 \quad (4)$$

$$Q_k(t) = A_{Qk}\cdot\sin\left[\frac{4\pi d_0}{\lambda_k} + \frac{4\pi x_k(t)}{\lambda_k} \mp \frac{4\pi b(t)}{\lambda_k} + \cdot\theta_k(t)\right] + DC_{Qk},\ k = 1.2 \quad (5)$$

where A_{Ik} and A_{Qk} are the amplitudes in I/Q channels, $\Delta\theta_k(t)$ is the residual phase noise, which is neglected in short-range applications because of the range correlation effect, and DC_{IK} and DC_{QK} are the DC offset voltages in I/Q channels, which are generated by stationary clutters in the experimental environment and direct coupling between the transmitting and receiving signals [27]. The signal processing may require demodulation to extract the vital signal $x(t)$ in the trigonometric functions in Equations (4) and (5). A mathematical demodulation technique such as arcsine or arctangent demodulation is not suitable for the proposed configuration with multiple radar modules because DC offset voltages caused by the presence of the subject and surrounding clutter are difficult to remove [27,28]. The circle fitting method, which can extract the displacement through the circle trajectory, shown as a graph in the I/Q plot, can demodulate the dominant displacement of baseband signals, even in an environment with a DC offset [29]. However, when $b(t)$ caused by the movement of the human body is dominantly shown in the circle trajectory, the circle fitting method has limits to vital signal detection using the proposed radar configuration because $x(t)$ may be lost in the demodulated signals [8]. The complex signal demodulation (CSD) used in the proposed configuration is useful for detecting small displacements from vital signals in an environment with a DC offset. The complex signal $S_k(t)$ can be expressed as follows:

$$S_k(t) = I_k'(t) + j\cdot Q_k'(t) = A_k\exp\left[j\left(\frac{4\pi d_0}{\lambda_k} + \frac{4\pi x_k(t)}{\lambda_k} \mp \frac{4\pi b(t)}{\lambda_k}\right)\right],\ k = 1.2 \quad (6)$$

where $I_k'(t)$ and $Q_k'(t)$ are baseband signals after compensating for the I/Q imbalance, and A_k is the amplitude of the complex signal. The DC offset is negligible in the I/Q

imbalance compensation and the CSD method because it is significantly reduced in $S_k(t)$ by the amplitude of the baseband signals.

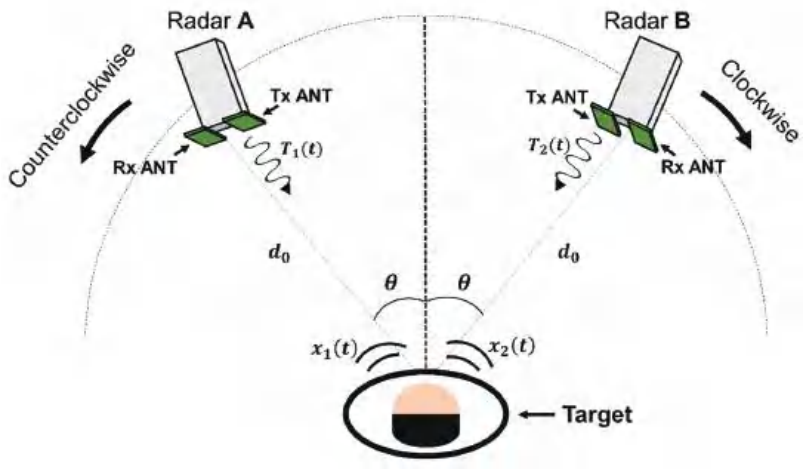


Figure 4. Proposed radar configuration using multiple radars based on the asymmetrical movement of human organs.

A signal processing technique is proposed to improve the SNR in vital signal detection using the correlation between two radar signals. By normalizing $S_k(t)$ and extracting only the phase using a natural logarithm, the phase of the baseband signal obtained in the conventional CSD method can be expressed as follows:

$$P_k(t) = \frac{4\pi d_0}{\lambda_k} + \frac{4\pi x_k(t)}{\lambda_k} \mp \frac{4\pi b(t)}{\lambda_k}. \qquad k = 1.2 \tag{7}$$

The phase difference $P_D(t)$ between two operating frequencies obtained from each radar can be expressed as

$$P_D(t) = P_1(t) - P_2(t) = 4\pi(d_0 \mp b(t))\cdot\left(\frac{1}{\lambda_1} - \frac{1}{\lambda_2}\right) + 4\pi\left(\frac{x_1(t)}{\lambda_1} - \frac{x_2(t)}{\lambda_2}\right) \tag{8}$$

As shown in Equation (8), the effect of $b(t)$ in the proposed configuration is not entirely diminished because of the different operating frequencies, but it is smaller than that in the single CW radar configuration. The amplitude of the periodic vital signal is as prominent as the asymmetrical movement of organs due to the non-identical amplitudes of the vital signals from each radar. When an identical operating frequency is used in the two radars, the effect of the human body movement can be removed, as the second term in Equation (8) will remain due to the asymmetric movement, but the first term will cancel out [8–10]. However, the displacement of the movement in Equation (8) is difficult to remove if the wavelength difference between the two radars is large, and it has a significant effect on the noise level by increasing the harmonic components due to the vital signals and the signal caused by human body movement. The mathematical expression of the signal using Equation (8) can be expressed as

$$C(t) = \exp(jP_D(t)) = \exp\left[j4\pi(d_0 \mp b(t))\cdot\left(\frac{1}{\lambda_1} - \frac{1}{\lambda_2}\right) + j4\pi\left(\frac{x_1(t)}{\lambda_1} - \frac{x_2(t)}{\lambda_2}\right)\right] \tag{9}$$

The conventional signal processing method obtains the respiration and heartbeat signals by searching the signal amplitude in the frequency band corresponding to the respiration and heartbeat using the fast Fourier Transform (FFT) of Equation (9). Respiratory and heartbeat signals can be accurately obtained for each vital signal by comparison with

the frequency detected by the reference sensor. In the proposed signal processing, the vital signals are extracted from the difference in the spectrum obtained by each FFT of the demodulated signal by the CSD. The mathematical expression of the signal processed by the proposed method can be expressed as a subtraction of the normalized complex signals $S_k{}'(t)$ shown in Equation (6) as follows:

$$F(t) = S_1'(t) - S_2'(t) = \exp\left[\frac{j4\pi}{\lambda_1}(d_0 + x_1(t) \mp b(t))\right] - \exp\left[\frac{j4\pi}{\lambda_2}(d_0 + x_2(t) \mp b(t))\right]. \quad (10)$$

The first term in the exponential function represents the DC signals caused by the distance between the radar and the subject; the DC signal level from the difference between two frequencies is reduced in the FFT results compared with that from a single radar. The effect of $b(t)$ in the proposed configuration is not entirely diminished because of the different operating frequencies, but $b(t)$ in Equation (10) can be expressed in the same form as Equation (9) after applying a complex FFT and can be reduced in the output signal $F(t)$ by a subtractive operation. However, $x(t)$ in Equation (10) cannot be reduced at the output because $x(t)$ in each radar is not the same in magnitude and phase, as shown in Equation (3). As the phases of the vital signals independently obtained from the two radars are different owing to the asymmetric movement of human organs, $x(t)$ may be integrated during the sampling period and increase beyond the signal level obtained from a single radar.

Figure 5 shows the digital signal processing in the proposed radar configuration. A simulation to verify the proposed configuration and signal processing was performed using MATLAB. It was assumed that two radars individually operating at 5.75 GHz and 5.85 GHz are located at 0.5 m from the subject. In the simulation, the magnitudes of the vital signals were set to 8 mV_{PP} at radar A and 7 mV_{PP} at radar B for respiration and 1 mV_{PP} at radar A and 0.4 mV_{PP} at radar B for heartbeat, considering the measurement data from the previous experiments [4,5,18]. The phase differences of the vital signals were set to π between radar A and B. The overall data acquisition time and sampling frequency in the simulated experiment were set to 40 s and 1 kHz, respectively. The simulation assumed that the signal caused by the human body is located at 0.003 Hz with the magnitude of 100 mV_{PP}. Figure 6 shows the normalized spectrum of the baseband signals using Equations (6) and (10) following the complex FFT. In the simulated spectrum of the single-frequency radar, the signal caused by body movement has a lower frequency than the vital signals because of its low velocity. The simulation results show that the SNR of the vital signals can be reduced by body movements with a large displacement. The simulated spectrum processed by the proposed signal processing method shows a significant reduction in the effect of body movement because of a correlation between the baseband signals of the two radars. A respiration frequency of 0.4 Hz and heartbeat frequency of 1.4 Hz, which are the same values as in the simulation condition, were effectively recovered by increasing the SNR of the vital signal detection because of the proposed signal processing. Figure 7 shows the simulated spectra of both the conventional method and the proposed method in the radar configuration. Compared to the conventional method, which increases the harmonics using a nonlinear function, the modified method can suppress the harmonic generation in the spectrum. The same frequencies of vital signals in both spectra indicate that the proposed method does not distort the demodulated signals when compared to the conventional method. Figure 8 shows the signal processing gain of the conventional and proposed signal processing methods depending on the difference in the operating frequency of the two radars. The SNR in Figure 8 was calculated from the simulation results by using the signal spectrum magnitude of the vital signals and the noise spectrum magnitude of the human movement signal. Simulation results show that the proposed signal processing method can achieve higher SNR in both respiration and heart rate compared to the conventional method. In particular, the SNR of the proposed method improved 1.3 dB for respiration signals and 0.7 dB for heartbeat signals at a frequency difference of 100 MHz in the proposed configuration. The SNRs of the conventional and proposed methods do not show a significant difference above the frequency difference of 170 MHz. These results show that

the proposed method can be effective for vital signal detection in the 5.8 GHz ISM band with a maximum frequency bandwidth of 150 MHz.

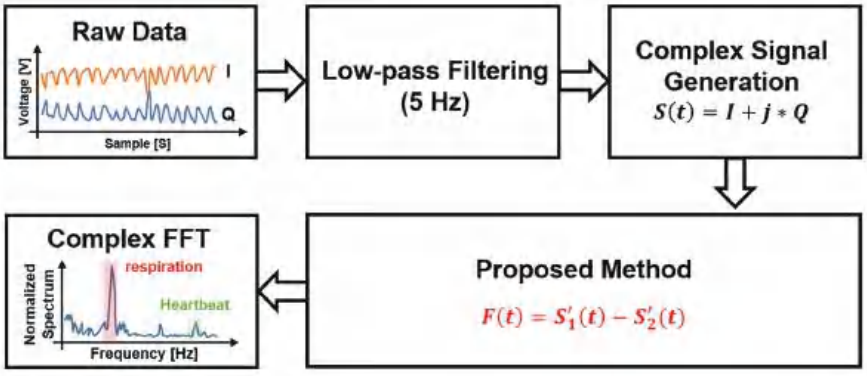


Figure 5. Digital signal processing using the proposed correlation technique for reducing the effect of human body movement.

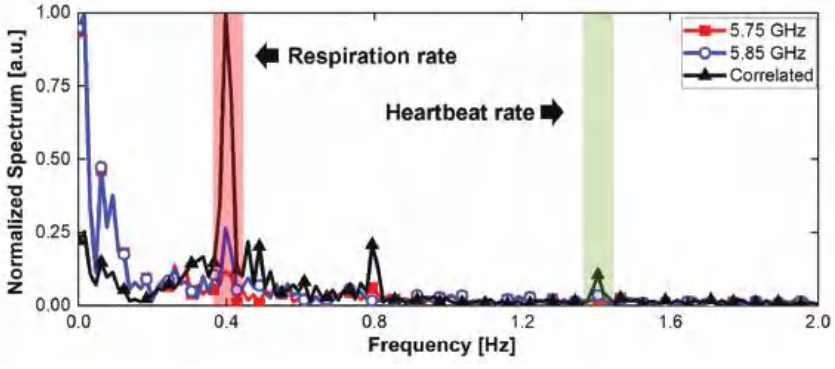


Figure 6. Normalized spectrum of baseband signals in the simulation.

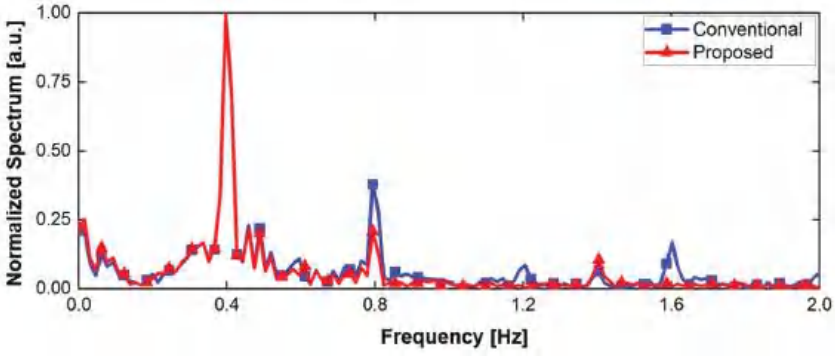


Figure 7. Simulated spectrum of the conventional and proposed signal processing methods.

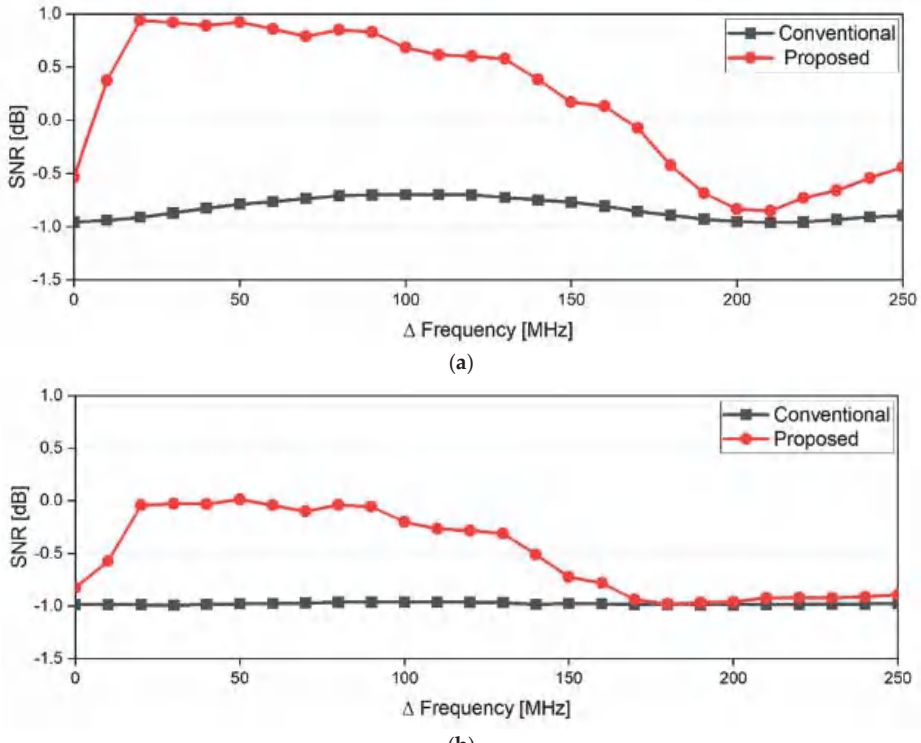


Figure 8. Signal-to-noise ratios of the conventional and proposed signal processing methods depending on the difference in the operating frequency of the two radars: the signal is the magnitude of the vital signs and the noise is the magnitude of the signal caused by the human body movement: (**a**) for respiration; (**b**) for heartbeat.

3. Measurement Environment

Two signal channel radar modules were implemented for the proposed radar configuration as shown in Figure 9 [17]. The operating frequencies of the two radars in the 5.8 GHz ISM band are individually determined with the control voltage of a voltage-controlled oscillator (VCO). In the experiment, the frequencies were set to 5.75 GHz and 5.85 GHz, with a frequency gap of 100 MHz. The transmitted powers at each radar module were measured to be 7.8 dBm at 5.75 GHz and 9.3 dBm at 5.85 GHz, respectively. The desired LOS, as shown in Figure 10, was set to the position and angle of the patch antenna with an antenna gain of 4.4 dBi, which is connected to the radar module through low-loss RF cables. The quadrature signals of the module were simultaneously obtained using a data acquisition board (NI USB-6366, National Instruments, Austin, TX, USA) with a sampling rate of 1 k samples per second in each channel of the two radars. A three-electrode ECG sensor (EKG-BTA, Vernier Software & Technology, Beaverton, OR, USA) and respiration belt (GDX-RB, Vernier Software & Technology, Beaverton, OR, USA) were used as reference sensors to compare the accuracy of vital sign detection using the proposed radar configuration. The reference data for displacement and velocity of the human body movement were measured using a laser sensor (ILR1182-30, Micro-Epsilon, Ortenburg, Germany) with a resolution of 0.1 mm and sampling rate of 50 samples per second.

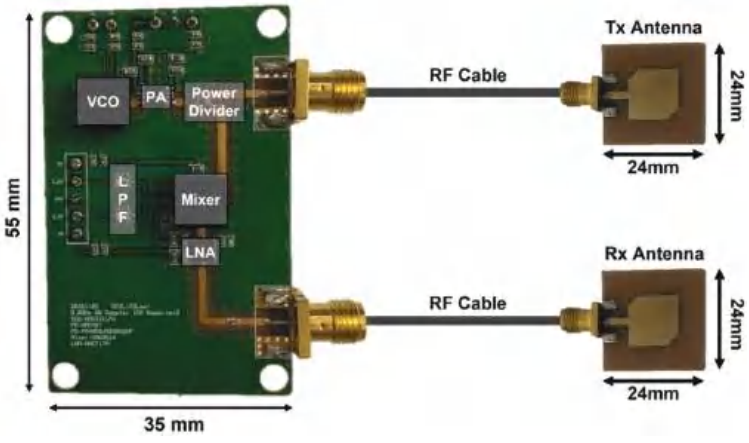


Figure 9. Implemented single-channel radar module in the 5.8 GHz ISM band.

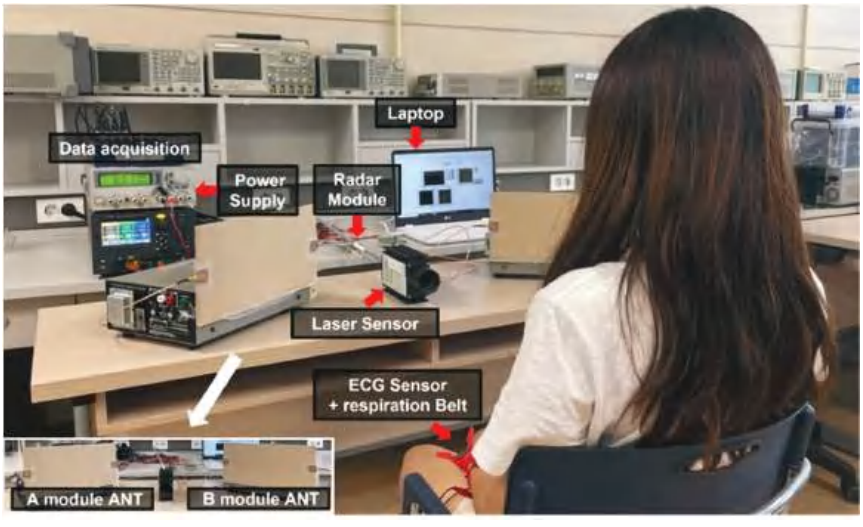


Figure 10. Experiment environment of the proposed radar configuration for vital signal detection for canceling the effect of body movement.

The two CW radars with different operating frequencies were positioned at a specific angle in the radar configuration, as shown in Figures 4 and 10. The angles in the experiment were set to 30°, 45°, and 60°, and the distance between the subject and each radar was fixed at 0.5 m. The body movement was controlled in the experiment for the given conditions, a motionless state and a random back-and-forth moving state for approximately 5 s. The movement of the subject's arms was restricted because the movement may affect the experimental results. The velocity obtained from the reference laser sensor was calibrated to the velocity at the radar module considering the different LOSs and the angles.

4. Results and Discussion

The radar experimental frequency spectrum of the respiration and heartbeat signals were simultaneously compared with the signals of the reference respiration belt and an ECG sensor. The experiment was configured so that the subject's respiration harmonics did not overlap with the heartbeat signal to prevent obscuring the detection of the heartbeat signal.

Figure 11 shows the frequency spectra of the vital signals obtained from the proposed radar configuration with a LOS angle of 30° and the reference signals obtained from the respiration belt and the ECG sensor. The peak frequency of the respiration in the proposed radar was 0.36 Hz while the reference frequency measured by the respiration belt was 0.37 Hz. The peak frequency of the heartbeat signal in the radar was 1.34 Hz while the reference ECG sensor measured 1.32 Hz. The frequency difference between the respiration harmonics and the heartbeat was 0.1 Hz or higher.

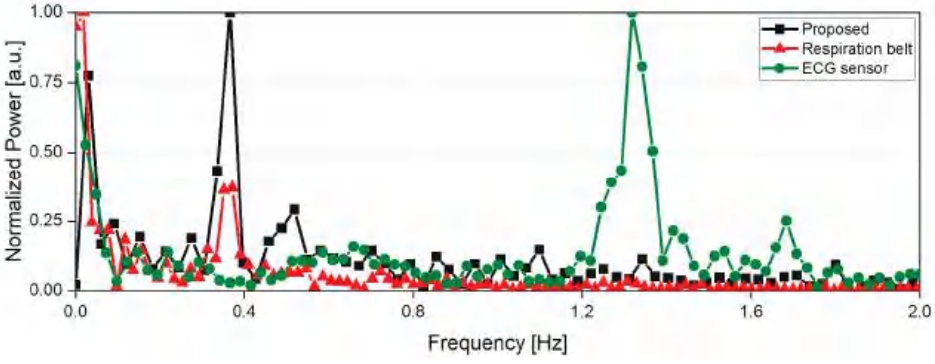


Figure 11. Frequency spectra of the vital signals measured by using the reference sensors (a respiration belt and an ECG sensor) and the proposed radar configuration at the LOS angle of 30°.

The vital signals obtained in the proposed configuration using two radars were presented as a spectrum depending on the LOS, angle, and presence of body movement. The measurement results are of two types: a normalized spectrum and a spectrum expressed with the absolute amplitude, for demonstrating the SNR improvement of the vital signal detection by the proposed radar configuration and signal processing. Figures 12–14 show the spectra measured in the experimental environment configured with a measurement angles of 30°, 45°, and 60°, respectively. In the case of motion, the subject in the experiment moved arbitrarily in the forward and backward directions, and the maximum velocity in the measurement was presented with a laser-based reference sensor because it is difficult to control the velocity and displacement of these movements. Owing to the angle set by the LOS of the subject and the radar, the body movement of the subject is presented in the direction of the LOS on the radar. The velocity of the subject's body movement was measured to be a maximum of 17.2 mm/s at 30°, 12 mm/s at 45°, and 26.7 mm/s at 60° using the reference sensor at the front of the subject, and the calculated velocity considering the angle was 19.9 mm/s at 30°, 17.0 mm/s at 45°, and 53.4 mm/s at 60°, respectively. The subject's movement was limited in the experiment because a sufficient space between the radars and the subject was not secured, and the maximum displacement because of the movement was 80 mm at 30°, 80 mm at 45°, and 30 mm at 60°, as measured by the reference sensor.

Respirations were detected from the measured signal peaks, regardless of the presence or absence of motion in the LOS angles of 30° and 60°, as shown in Figure 12a,c, Figure 13a and Figure 14a,c. However, in the radar arrangement at an angle of 45°, the peak of the respiration signal was detected only in the correlated signal by the proposed signal processing, as shown in Figure 13c. Despite its small velocity when compared to other angles, the subject's movement at 45° in Figure 13 had a significant effect on vital signal detection because of the SNR degradation caused by an increase in the noise. This shows that the displacement magnitude is more important than velocity of human body movement because the subject at an angle of 45° moved his body with a high displacement and a low velocity. For the subject's motionless condition, the heartbeat signals were measured at all angles by a correlation between the two radars using the proposed method,

as shown in Figure 12a, Figure 13a and Figure 14a; however, the signal in the single radar was measured only from radar A, which was located close to the subject's heart. Assuming that the transmitter output power and receiver sensitivity of the two radars do not have a significant difference, it can be seen that a larger heartbeat signal was received by radar A because of the asymmetry of the heart rate and position between the two radars. The heartbeat signals in an environment with the subject's motion were not obtained from each radar, as shown in Figure 12c, Figure 13c and Figure 14c. However, the signal correlated using the proposed signal processing displayed the heartbeat signals at all angles. The performance improvement by the proposed configuration and processing can be explained by the increase in the SNR because of a decrease in the noise reduction near DC. Figure 12d, Figure 13d and Figure 14d, displayed as the absolute amplitudes of the signals, show that the signals near DC caused by the motion of the subject are significantly reduced by the proposed signal processing. The spectra at angles of 30° and 45° (Figure 12b and Figure 13b) in the motionless condition show that the noise level near DC was slightly reduced by the proposed method. However, there was no significant reduction in the noise level of the spectrum at an angle of 60°, as shown in Figure 14b. The residual noise near DC in the motionless state of the subject shows that the proposed configuration and processing can only improve the performance for common noise in both radars. This shows the limitation of the proposed configuration and processing: it does not show a performance improvement for noise generated by the asymmetric clutter and multipath problem.

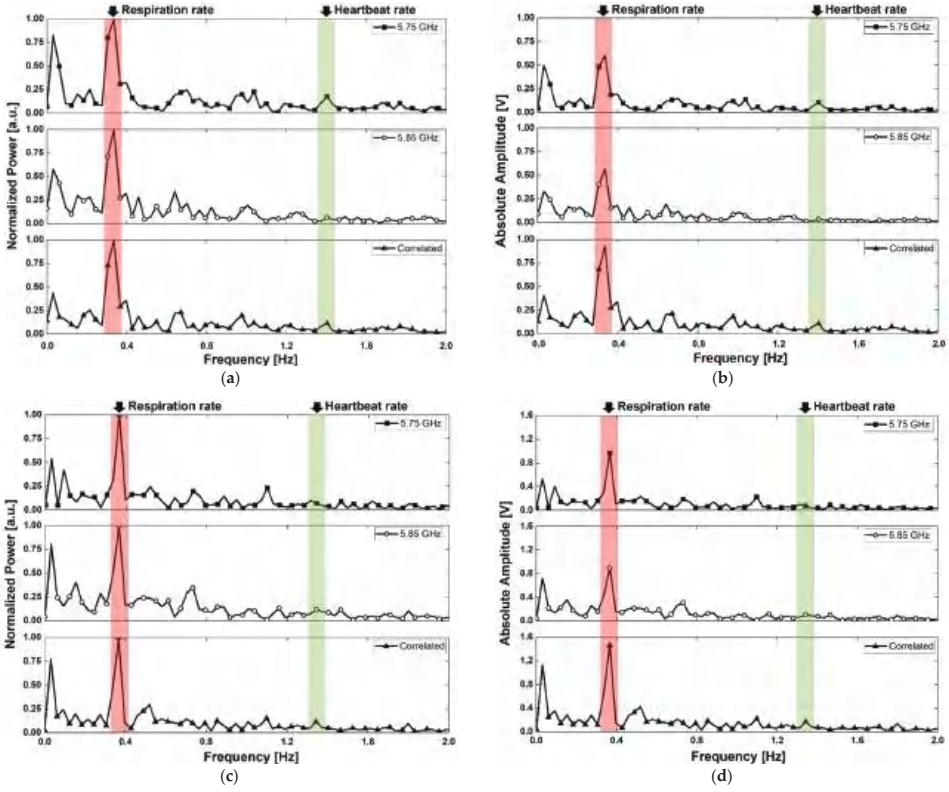


Figure 12. Measurement results using the proposed radar configuration in the experiment with a measurement angle of 30°: (**a**) normalized spectrum obtained in the motionless condition; (**b**) spectrum displayed with the absolute amplitudes of signals in the motionless condition; (**c**) normalized spectrum obtained in the presence of human body movement; and (**d**) spectrum displayed with the absolute amplitudes of the signals in the presence of human body movement.

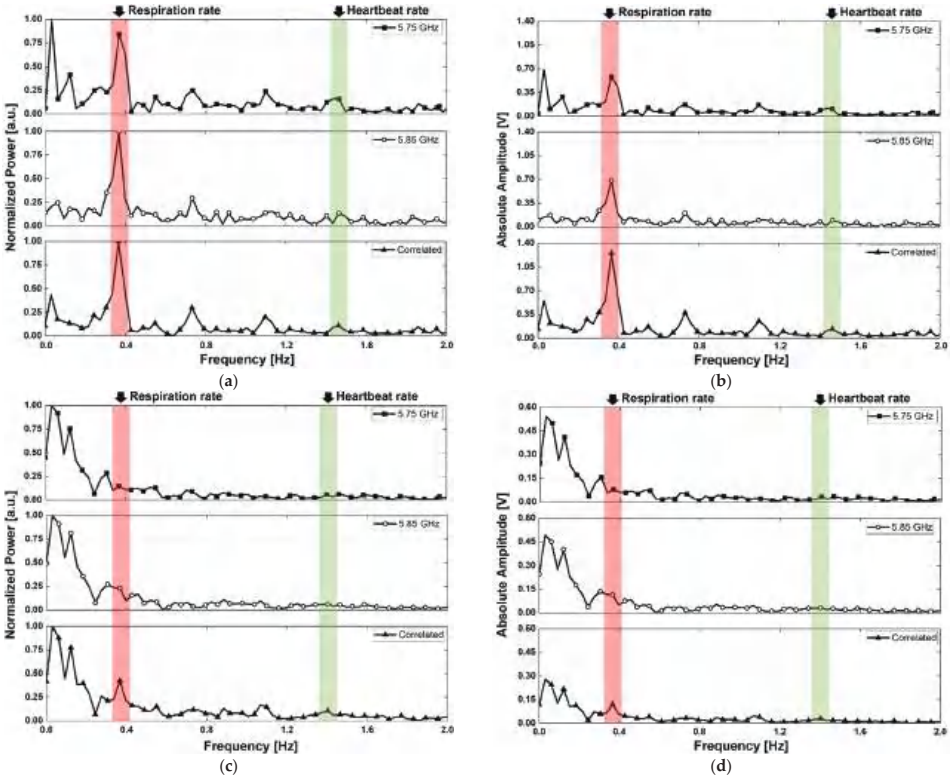


Figure 13. Measurement results using the proposed radar configuration in the experiment with a measurement angle of 45°: (**a**) normalized spectrum obtained in the motionless condition; (**b**) spectrum displayed with the absolute amplitudes of signals in the motionless condition; (**c**) normalized spectrum obtained in the presence of human body movement; (**d**) spectrum displayed with the absolute amplitudes of the signals in the presence of human body movement.

The SNR of the vital signal detection in the experiment can be expressed as the ratio of the sum of the respiration and heartbeat signals to the sum of all signals except the vital signals below 2 Hz. The SNR improvement of the vital signal detection by the proposed signal processing method for the moving human body condition was measured to be 5.6 dB for respiration and 3.3 dB for heartbeat at an angle of 30°, 5.7 dB for respiration and 4.2 dB for heartbeat at an angle of 45°, and 3.7 dB for respiration and 3.0 dB for heartbeat at an angle of 60°, respectively. The detection accuracy of the vital signal calculated using the measured peaks in Figures 12–14 was 96.8% for respiration and 98.2% for heartbeat at 30°, 96% for respiration and 99.2% for heartbeat at 45°, and 98.4% for respiration and 96% for heartbeat at 60°, respectively. Compared to previous studies on vital signal detection using the CW radars, the detection accuracy of the proposed radar configuration was lower for respiration but higher for heartbeat. In the measurement results, the detection accuracy of respiration was reduced by increasing the noise level because of the movement of the human body appearing at a lower frequency range than the respiration. The detection accuracy of heartbeat was not affected by the noise level from the motion because of the far frequency range between the noise and the heartbeat signals and increased because of the SNR improvement from the proposed signal processing. Table 1 summarizes the comparison of vital signal detection using radar technology to reduce the effect of human body movement.

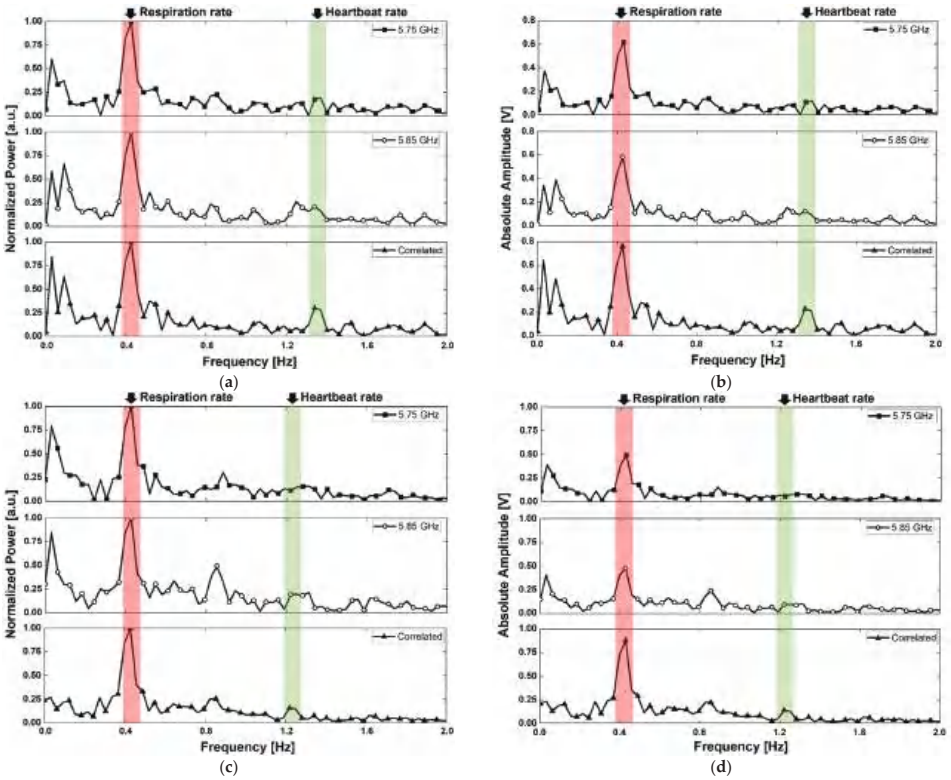


Figure 14. Measurement results using the proposed radar configuration in the experiment with a measurement angle of 60°: (**a**) normalized spectrum obtained in the motionless condition; (**b**) spectrum displayed with the absolute amplitudes of signals in the motionless condition; (**c**) normalized spectrum obtained in the presence of human body movement; (**d**) spectrum displayed with the absolute amplitudes of the signals in the presence of human body movement.

Table 1. Radar technology for vital signal detection in the presence of human body movement.

Ref.	Techniques	Maximum Body Movement [mm]	Maximum Body Velocity [mm/s]	Detection Accuracy [%]	
				Respiration	Heartbeat
[8]	CSD method using two antennas around the subject	Not mentioned	4	Not mentioned	Not mentioned
[30]	SIL [1] radar using two antennas	200	<7.7	Not mentioned	96.5
[31]	Polynomial fitting algorithm	150	≈ 0	Not mentioned	Not mentioned
[32]	Adaptive noise cancelation algorithm	155	47.6	97.9	99.1
This work	Correlation method using multiple radars	80	53.4	97.8 [2]	97.9 [2]

[1] Single Self-Injection-Locked Radar. [2] Average data from the measurement results at all angles.

The measurement results in the motion of the subject in Figure 12d, Figure 13d and Figure 14d show that the proposed configuration and processing increase the absolute amplitudes of the vital signals and decrease the motion-induced noise. The amplitude and phase of the vital signals obtained from the two radars should be different because of the asymmetrical movement of the human organs to simultaneously realize a decrease in noise and an increase in the vital signals. The different amplitudes and phases between the vital signals obtained from the two radars show that the asymmetric movement of human organs affects vital signal detection using the radar. In particular, the phases of the vital signals $x_1(t)$ and $x_2(t)$ acquired simultaneously by two radars should not be the same to increase the amplitude of vital signals by the proposed signal processing method explained in Section 2. Figure 15 shows the phase waveforms of vital signals acquired simultaneously and independently from the two radars. The waveforms of the simultaneously sampled data show that the phases of the vital signals have a difference of 180° in the proposed configuration. This is caused by the asymmetrical characteristics of vital signals in the radar because of the presented phase differences in the measurement results, regardless of the LOS angles.

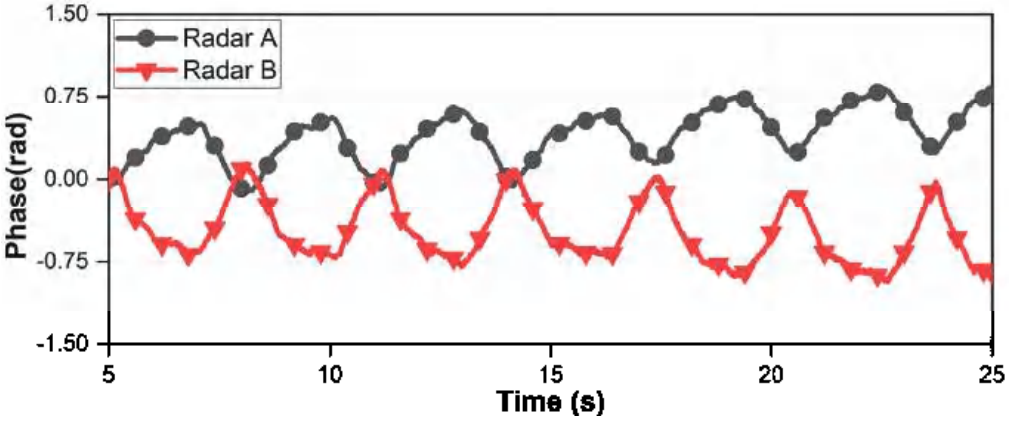


Figure 15. Phase waveform of the vital signals simultaneously measured in each radar.

The proposed radar configuration was demonstrated by strictly controlling the variables and factors that can affect the vital signal detection. Therefore, it has limitations for extension to general applications. However, the configuration limitations on the location and the arrangement between the radars and the subject can be solved with a modified radar operation that can detect the distance difference between each radar and the subject, such as FSK or FMCW radars. This study shows that the radar can detect vital signals based on the asymmetric movements of the internal organs using the proposed configuration and signal processing techniques.

5. Conclusions

A configuration and signal processing method using two radars operating at different frequencies are proposed for detecting vital signals during the subject's body movement. Based on the asymmetrical organ movements caused by the vital signals monitored in the CW Doppler radar, the proposed radar configuration includes two radars spaced apart in front of the subject with different LOSs at the same distance. The operating frequencies of two radars were individually set to 5.75 GHz and 5.85 GHz in the 5.8 GHz ISM for reducing a direct coupling between them. A signal processing method was proposed for effectively extracting correlated vital signals from the complex baseband signals received from the two radars. The proposed signal processing had the same demodulation performance as the conventional method without using a nonlinear function, which increases the harmonics. A

SNR improvement was observed in vital signal detection during human body movement, and the stable accuracy was enhanced for asymmetrical organ movements in the proposed method. The proposed configuration based on the asymmetrical movement of organs was verified by placing two radars at a distance of 0.5 m from the subject for different LOSs at angles of 30°, 45°, and 60° from the center of the subject. For the motionless condition, the respiration and heartbeat were obtained from the signals detected by the radar located closer to the heart on the left side of the subject, and the signals from the two radars were correlated using the proposed method. In the presence of body movement with a maximum velocity of 53 mm/s, respiration and heartbeat could be detected only from the correlated signals obtained using the proposed method. The noise reduction in the low-frequency range by the proposed method shows that it can reduce the effect of human movement. The improvement in SNR and the detection accuracy of both respiration and heartbeat detection by the proposed configuration and method were measured to be more than 3 dB and 96% at all three angles, respectively.

Author Contributions: Conceptualization, J.-R.Y.; methodology, A.-J.J. and J.-R.Y.; software, A.-J.J., I.-S.L.; validation, A.-J.J., I.-S.L. and J.-R.Y.; formal analysis, A.-J.J., I.-S.L. and J.-R.Y.; investigation, A.-J.J.; resources, J.-R.Y.; data curation, A.-J.J.; writing—original draft preparation, A.-J.J. and J.-R.Y.; writing—review and editing, J.-R.Y.; visualization, A.-J.J.; supervision, J.-R.Y.; project administration, J.-R.Y.; funding acquisition, J.-R.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the 2020 Yeungnam University Research Grant (No. 220A380046).

Institutional Review Board Statement: Ethical review and approval were waived for this study because the low transmitting output signal does not affect the human body.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to the privacy of the subjects.

Acknowledgments: The authors thank Ji-In Jeong for designing the patch antenna for implementing the CW radar and Jae Yong Shim for assisting with the mathematical analysis of the operating principle.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Baboli, M.; Singh, A.; Soll, B.; Boric-Lubecke, O.; Lubecke, V. Wireless sleep apnea detection using continuous wave quadrature Doppler radar. *IEEE Sens. J.* **2020**, *20*, 538–545. [CrossRef]
2. Castro, I.D.; Mercuri, M.; Patel, A.; Puers, R.; Van Hoof, C.; Torfs, T. Physiological driver monitoring using capacitively coupled and radar sensors. *Appl. Sci.* **2019**, *9*, 3994. [CrossRef]
3. Li, C.; Lubecke, V.; Boric-Lubecke, O.; Lin, J. A review on recent advances in Doppler radar sensors for noncontact healthcare monitoring. *IEEE Trans. Microwave Theory Tech.* **2013**, *61*, 2046–2060. [CrossRef]
4. Kim, J.-Y.; Park, J.-H.; Jang, S.-Y.; Yang, J.-R. Peak detection algorithm for vital sign detection using Doppler radar sensors. *Sensors* **2019**, *19*, 1575. [CrossRef] [PubMed]
5. Sim, J.Y.; Park, J.-H.; Yang, J.-R. Vital-signs detector based on frequency-shift keying radar. *Sensors* **2020**, *20*, 5516. [CrossRef]
6. Droitcour, A.D.; Boric-Lubecke, O. *Doppler Radar Physiological Sensing*; Wiley: Hoboken, NJ, USA, 2016; pp. 39–68.
7. Gouveia, C.; Vieira, J.; Pinho, P. A review on methods for random motion detection and compensation in bio-radar systems. *Sensors* **2019**, *19*, 604. [CrossRef] [PubMed]
8. Li, C.; Lin, J. Random body movement cancellation in doppler radar vital sign detection. *IEEE Trans. Microwave Theory Tech.* **2008**, *56*, 3143–3152.
9. Yu, X.; Li, C.; Lin, J. Two-dimensional noncontact vital sign detection using Doppler radar array approach. In Proceedings of the 2011 Conference on International Microwave Symposium, Baltimore, MD, USA, 5–10 June 2011.
10. Li, C.; Xiao, Y.; Lin, J. Experiment and spectral analysis of a low-power Ka-band heartbeat detector measuring from four sides of a human body. *IEEE Trans. Microwave Theory Tech.* **2006**, *54*, 4464–4471. [CrossRef]
11. Tang, M.-C.; Kuo, C.-Y.; Wun, D.-C.; Wang, F.-K.; Horng, T.-S. A self- and mutually injection-locked radar system for monitoring vital signs in real time with random body movement cancellation. *IEEE Trans. Microwave Theory Tech.* **2016**, *64*, 4812–4822. [CrossRef]

12. Wang, F.-K.; Horng, T.; Peng, K.; Jau, J.; Li, J.-Y.; Chen, C. Single-antenna Doppler radars using self and mutual injection locking for vital sign detection with random body movement cancellation. *IEEE Trans. Microwave Theory Tech.* **2011**, *59*, 3577–3587. [CrossRef]
13. Singh, A.; Lubecke, V. Respiratory monitoring and clutter rejection using a CW Doppler radar with passive RF tags. *IEEE Sens. J.* **2012**, *12*, 558–565. [CrossRef]
14. Gu, C.; Wang, G.; Li, Y.; Inoue, T.; Li, C. A hybrid Radar-camera sensing system with phase compensation for random body movement cancellation in Doppler vital sign detection. *IEEE Trans. Microwave Theory Tech.* **2013**, *61*, 4678–4688. [CrossRef]
15. Mostafanezhad, I.; Park, B.-K.; Boric-Lubecke, O.; Lubecke, V.; Host-Madsen, A. Sensor nodes for Doppler radar measurements of life signs. In Proceedings of the IEEE/MTT-S International Microwave Symposium, Honolulu, HI, USA, 3–8 June 2007.
16. Gu, C.; Wang, G.; Inoue, T.; Li, C. Doppler radar vital sign detection with random body movement cancellation based on adaptive phase compensation. In Proceedings of the IEEE MTT-S International Microwave Symposium, Seattle, WA, USA, 2–7 June 2013.
17. Lv, Q.; Dong, Y.; Sun, Y.; Li, C.; Ran, L. Detection of bio-signals from body movement based on high-dynamic-range Doppler radar sensor (Invited). In Proceedings of the IEEE MTT-S International Microwave Workshop Series on RF and Wireless Technologies for Biomedical and Healthcare Applications, Taipei, Taiwan, 21–23 September 2015.
18. Lee, I.-S.; Park, J.-H.; Yang, J.-R. Detrending technique for denoising in CW radar, Sensors 2021, 21, 6376. *Sensors* **2021**, *21*, 6376.
19. Li, Y.; Wang, G.; Gu, C.; Li, C. Movement-immune respiration monitoring using automatic DC-correction algorithm for CW Doppler radar system. In Proceedings of the Topical Conference on Biomedical Wireless Technologies, Networks, and Sensing Systems, Newport Beach, CA, USA, 19–23 January 2014; pp. 7–9.
20. Ramachandran, G.; Singh, M. Three-dimensional reconstruction of cardiac displacement patterns on the chest wall during the P, QRS and T-segments of the ECG by laser speckle inteferometry. *Med. Biol. Eng. Comput.* **2006**, *27*, 525–530. [CrossRef] [PubMed]
21. Laizzo, P.A. *Handbook of Cardiac Anatomy, Physiology, and Devices*; Springer International Publishing: Manhattan, NY, USA, 2015; pp. 51–79.
22. Periasamy, A.; Singh, M. Reconstruction of cardiac displacement patterns on the chest wall by laser speckle interferometry. *IEEE Trans. Med. Imaging* **1985**, *4*, 52–57. [CrossRef] [PubMed]
23. De Groote, A.; Wantier, M.; Cheron, G.; Estenne, M.; Paiva, M. Chest wall motion during tidal breathing. *Eur. J. Appl. Physiol.* **1997**, *83*, 1531–1537. [CrossRef]
24. Plathow, C.; Ley, S.; Fink, C.; Puderbach, M.; Heilmann, M.; Zuna, I.; Kauczor, H. Evaluation of chest motion and volumetry during the breathing cycle by dynamic MRI in healthy subjects. *Investigative Radiology* **2004**, *39*, 202–209. [CrossRef]
25. BruceBlaus. Medical Gallery of Blausen Medical 2014. Available online: https://commons.wikimedia.org/wiki/File:Systolevs_Diastole.png (accessed on 28 September 2021).
26. Memmler, R.L.; Cohen, B.J.; Wood, D.L.; Ravielli, A. *Memmler's The Human Body in Health and Disease*, 11th ed.; Lippincott Williams & Wilkins: Burlington, MA, USA, 2009.
27. Park, B.-K.; Boric-Lubecke, O.; Lubecke, V. Arctangent demodulation with DC offset compensation in quadrature Doppler radar receiver systems. *IEEE Trans. Microwave Theory Tech.* **2007**, *55*, 1073–1079. [CrossRef]
28. Fan, T.; Ma, C.; Gu, Z.; Lv, Q.; Chen, J.; Ye, D.; Huangfu, J.; Sun, Y.; Li, C.; Ran, L. Wireless hand gesture recognition based on continuous-wave Doppler radar sensors. *IEEE Trans. Microwave Theory Tech.* **2016**, *64*, 4012–4020. [CrossRef]
29. Park, J.-H.; Yang, J.-R. Multiphase continuous-wave Doppler radar with multiarc circle fitting algorithm for small periodic displacement measurement. *IEEE Trans. Microwave Theory Tech.* **2021**, *69*, 5135–5144. [CrossRef]
30. Tang, M.-C.; Wang, F.-K.; Horng, T.-S. Single self-injection-locked radar with two antennas for monitoring vital signs with large body movement cancellation. *IEEE Trans. Microwave Theory Tech.* **2017**, *65*, 5324–5333. [CrossRef]
31. Lv, Q.; Chen, L.; An, K.; Wang, J.; Li, H.; Ye, D.; Huangfu, J.; Li, C.; Ran, L. Doppler vital signs detection in the presence of large-scale random body movements. *IEEE Trans. Microwave Theory Tech.* **2018**, *66*, 4261–4270. [CrossRef]
32. Yang, Z.-K.; Shi, H.; Zhao, S.; Huang, X.-D. Vital sign detection during large-scale and fast body movements based on an adaptive noise cancellation algorithm using a single Doppler radar sensor. *Sensors* **2020**, *20*, 4183. [CrossRef] [PubMed]

sensors

MDPI

Article

Data Enhancement via Low-Rank Matrix Reconstruction in Pulsed Thermography for Carbon-Fibre-Reinforced Polymers

Samira Ebrahimi [1,*], Julien R. Fleuret [1], Matthieu Klein [2], Louis-Daniel Théroux [3], Clemente Ibarra-Castanedo [1,2] and Xavier P. V. Maldague [1]

[1] Computer Vision and Systems Laboratory (CVSL), Department of Electrical and Computer Engineering, Laval University, Quebec, QC G1V 0A6, Canada; julien.fleuret.1@ulaval.ca (J.R.F.); clemente.ibarra-castanedo@gel.ulaval.ca (C.I.-C.); Xavier.Maldague@gel.ulaval.ca (X.P.V.M.)
[2] Visiooimage Inc. Infrared Thermography Testing Systems, Quebec, QC G1W 1A8, Canada; matthieu.klein@visiooimage.com
[3] Centre Technologique et Aérospatial (CTA), Saint-Hubert, QC 3Y 8Y9, Canada; louis-daniel.theroux@cegepmontpetit.ca
* Correspondence: samira.ebrahimi.1@ulaval.ca

Abstract: Pulsed thermography is a commonly used non-destructive testing method and is increasingly studied for the assessment of advanced materials such as carbon fibre-reinforced polymer (CFRP). Different processing approaches are proposed to detect and characterize anomalies that may be generated in structures during the manufacturing cycle or service period. In this study, matrix decomposition using Robust PCA via Inexact-ALM is investigated as a pre- and post-processing approach in combination with state-of-the-art approaches (i.e., PCT, PPT and PLST) on pulsed thermography thermal data. An academic sample with several artificial defects of different types, i.e., flat-bottom-holes (FBH), pull-outs (PO) and Teflon inserts (TEF), was employed to assess and compare defect detection and segmentation capabilities of different processing approaches. For this purpose, the contrast-to-noise ratio (CNR) and similarity coefficient were used as quantitative metrics. The results show a clear improvement in CNR when Robust PCA is applied as a pre-processing technique, CNR values for FBH, PO and TEF improve up to 164%, 237% and 80%, respectively, when compared to principal component thermography (PCT), whilst the CNR improvement with respect to pulsed phase thermography (PPT) was 77%, 101% and 289%, respectively. In the case of partial least squares thermography, Robust PCA results improved not only only when used as a pre-processing technique but also when used as a post-processing technique; however, this improvement is higher for FBHs and POs after pre-processing. Pre-processing increases CNR scores for FBHs and POs with a ratio from 0.43% to 115.88% and from 13.48% to 216.63%, respectively. Similarly, post-processing enhances the FBHs and POs results with a ratio between 9.62% and 296.9% and 16.98% to 92.6%, respectively. A low-rank matrix computed from Robust PCA as a pre-processing technique on raw data before using PCT and PPT can enhance the results of 67% of the defects. Using low-rank matrix decomposition from Robust PCA as a pre- and post-processing technique outperforms PLST results of 69% and 67% of the defects. These results clearly indicate that pre-processing pulsed thermography data by Robust PCA can elevate the defect detectability of advanced processing techniques, such as PCT, PPT and PLST, while post-processing using the same methods, in some cases, can deteriorate the results.

Keywords: Robust PCA; RPCA; PCP; IALM; noise reduction; pulsed thermography; CFRP

Citation: Ebrahimi, S.; Fleuret, J.R.; Klein, M.; Théroux, L.-D.; Ibarra-Castanedo, C.; Maldague, X.P.V. Data Enhancement via Low-Rank Matrix in Pulsed Thermography for Carbon-Fibre-Reinforced Polymers. *Sensors* **2021**, *21*, 7185. https://doi.org/10.3390/s21217185

Academic Editor: Manuel José Cabral dos Santos Reis

Received: 30 July 2021
Accepted: 22 October 2021
Published: 29 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Due to the unique features of Carbon-fibre-reinforced polymers (CFRP)—low-density and high-performance physico-chemical properties—the interest in using these lighter products and thus replacing the conventional materials (Steel, aluminum, etc.) has increased. The increasing demand for CFRP structures in the aerospace industry is leading to

the development of enhanced more eco-efficient manufacturing [1]. Although composite materials are sensitive to impact damage during a lifetime (manufacturing, operations, or maintenance) [2], they are less prone to corrosion and cracks than other materials. Due to the different types of defects during the manufacturing process or the service life of the components, it is important to monitor their efficiency and functionality non-invasively. Among non-destructive testing techniques, infrared thermography, which involves mapping the surface temperatures, can characterize the surface and sub-surface anomalies. Pulsed thermography (PT) is a no-contact and full-field Infrared Non-Destructive Testing (IRNDT) approach based on thermal heat transfer analysis during the cooling period; after the thermal impulse, an incident to the sample's surface becomes a thermal wave due to conduction and propagates through the material. The temperature decay is recorded by the infrared camera during the cooling period. Subject to the presence of discontinuity, depending on its material and thermal properties and depth, defects will be revealed at different times. The deeper defects appear later with lower thermal contrast. In order to obtain quantitative information from thermal data, several approaches have been proposed. Manipulating thermal data makes active thermography an attractive and powerful approach for industrial control and maintenance purposes.

Moreover, effective pre-processing or post-processing can provide favorable conditions to enhance defect information extraction. Most of the pre-processing for thermal data is limited to removing the first few frames from the beginning of the sequences, cropping the image, and selecting the region of interest (ROI). Fleuret et al. [3], in their study, proved that using LatLRR (Latent Low-Rank Representation) as a post-processing tool on the best image of state-of-the-art methods provides significant improvement in detection. Khodayar et al. [4] have used the thermographic signal reconstruction (TSR) [5] approach for pre-processing to reduce the noise. They stated that principal component thermography (PCT) [6] after the noise reduction could enhance the results. Wang et al. [7] used sequence differential pre-processing, which was combined with cold image subtraction (CIS) [8], to provide better thermal data for post-processing approaches in laser infrared thermography. They evaluated the quality of the image after the combination of pre-processing with pulsed phase thermography (PPT) [9] or PCT and found that pre-processing improved some results. Ebrahimi et al. [10] showed that the low-rank matrix computed by RPCA-PCP via Inexact ALM when used with PT data does not provide optimal results; nonetheless, this method has not been investigated as a pre-processing method nor as a post-processing method. Several state-of-the-art IRNDT methods, i.e., PPT, PCT and Partial Least Square Thermography (PLST) [11,12], have been chosen to evaluate the approaches. We chose these methods due to the large number of studies that use them.

In the remainder of this paper, we review the most recent works involving RPCA and thermography. Then, we detail the many aspects of our investigations in Section 3. Section 4 demonstrates the obtained results, which we analyze and discuss in Section 5. Finally, Section 6 concludes this study.

2. Literature Review

The presence of excessive noise in raw thermal data always urges researchers to develop new IRNDT processing approaches. Although limited research work has been done on the improvement of PCA methods to deal with corrupted data, RPCA has been the most promising approach in recent years. RPCA is widely used in separating dynamic variations from the static feature of interest, such as video surveillance data analysis to extract foreground and background [13]. Infrared dim small target detection has been a hot and difficult research topic in infrared search and tracking systems. Later, Fan et al. [14] introduced a novel detection algorithm based on RPCA to solve the difficulty of small target detection.

Substantial progress has been made in moving object detection, for which RPCA has been demonstrated to be very effective. The RPCA has been used in infrared moving target tracking [15] and hyper-spectral image processing for anomaly detection [16]. Moreover,

RPCA has been used for pre-processing in the machine learning method proposed by Zhu et al. [17]. They utilized RPCA to detect regions of interest (ROIs) in a novel classification model based on the CNN model in eddy current testing (ECT), and the percentage of defects correctly identified have increased to almost 100%. Draganov et al. [18] used several decomposition techniques, such as RPCA with Go implementation (GoDec), to estimate the wild animal population using videos captured by thermographic cameras. They reported promising results in terms of accuracy and execution times. Later, they carried out a comparative analysis of the performance of several tensor decomposition algorithms, including high-order robust principal component analysis solved by the Singleton model (HoRPCA-S) [19]. They reported that among the selected methods, HoRPCA-S has a lower detection rate but high precision. Furthermore, Liang et al. [20] have demonstrated the feasibility of sparse tensor decomposition theory on an ECPT data sequence, and they concluded that Tensor RPCA (TRPCA) can extract defects with high accuracy. The same year, Li et al. [21] introduced the weighted contraction IALM (WIALM) algorithm based on low-rank matrix recovery for online applications. It has been used for tire inspection on radiographic images captured by tire X-ray inspection machines. They improved the efficiency of the algorithm by optimizing the incremental multiplier parameter. Wu et al. [22] proposed a novel hierarchical low-rank and sparse tensor decomposition method to detect anomalies in the induction thermography stream. This approach can suppress the interference of a strong background and sharpens the visual features of defects. Furthermore, it overcame the over- and under-sparseness problem suffered by similar state-of-the-art methods. Surface defect detection is important for product quality control. A visual detection method was based on low-rank and sparse matrices extracted from the RPCA approach for surface defect detection of the wind turbine blade [23]. This method in terms of robustness and accuracy outperformed several state-of-the-art methods. Recently, Wang et al. [24] proposed a methodology based on RPCA that can separate anomalies in a sparse matrix from a low-rank background for photovoltaic systems using thermography imaging. They successfully overcame the difficulties arising from real data and built an automatic online monitoring system for anomaly detection. Ebrahimi et al. [10] proposed the orthogonal inexact augmented lagrange multiplier (OIALM). This study demonstrates its efficiency for defect enhancement capabilities over mixed and various types of defects typically addressed in IRT in composite materials. In addition, Kaur et al. [25] conducted a comparative study between PCA and RPCA to evaluate their effectiveness in defect detection. They demonstrated that although PCA proved to be better in detection capability, the sparse matrix provides better detectability than the data reconstructed from the low-rank matrix. In the medical field, for 3D segmentation of lungs, Sun et al. [26] achieved good segmentation results for lungs with juxta-pleural tumours by the active shape model (ASM) based on RPCA.

Many research works have reported the applicability of IRNDT approaches, including PCT, PPT and PLST. The first implementation of the PCT was introduced by Rajic [27] for defect detection in composite materials. Lara et al. expressed that optical effects, such as heating non-uniformities, surface reflection and emissivity variations, appear on the first component, and the thermal effect will be retrieved on one of the secondary components [28]. Furthermore, the PCA is a linear decomposition function that is sensitive to over-illumination and non-uniform heating more than other types of noise. In our previous research, we proved that Robust PCT [10] can improve the detectability of deeper defects in composites. Moreover, the PLST is sensitive to gradient. Having an approach that is less sensitive to noise and applicable to other IRNDT approaches in order to improve the defect detection is always interesting. As indicated from the literature, low-rank matrices from RPCA have less noise, and in this study, we study the use of this matrix on different IRNDT approaches.

The following section introduces the methods and materials regarding this study.

3. Methods and Materials

3.1. Robust Principal Component Analysis (RPCA)

The Robust PCA problem can be solved via convex optimization that minimizes a combination of the nuclear norm and the ℓ^1-norm. The augmented Lagrange multiplier (ALM) is a method to solve this convex program. Equation (1) introduces the general method of ALM for solving constrained optimization problems [29]:

$$\min f(\mathbf{X}),\ \textit{subject to}\ \ h(\mathbf{X}) = 0 \tag{1}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R}^n \to \mathbb{R}^m$. Candès et al. [30] used a convex optimization; the formulation they have used is known as PCP. The observation matrix D is assumed to be a combination of the low-rank (A) and sparse matrix (E):

$$\mathbf{D} = \mathbf{A} + \mathbf{E} \tag{2}$$

To minimize the energy function, ℓ_0-norm is used.

$$\begin{gathered} \min_{A,E}\ rank(\mathbf{A}) + \lambda\|\mathbf{E}\|_0 \\ \textit{subject to}\ \ \mathbf{D} - \mathbf{A} - \mathbf{E} = 0 \end{gathered} \tag{3}$$

where λ is a positive and arbitrary balanced parameter to determine the contribution of **A** and **E** in minimizing the objective function. Since Equation (3) is an NP-hard problem, i.e., at least as hard as the hardest problems in non-deterministic polynomial (NP) time, Candès et al. [30] reformulated this equation into a similar convex optimization problem as follows:

$$\begin{gathered} X = (\mathbf{A}, \mathbf{E})\ \ , \min_{\mathbf{A},\mathbf{E}}(\|\mathbf{A}\|_* + \lambda\|\mathbf{E}\|_1) \\ \textit{subject to}\ \ \mathbf{D} - \mathbf{A} - \mathbf{E} = 0 \end{gathered} \tag{4}$$

where $\|A\|_*$, $\|E\|_1$ are the nuclear norm of **A** and l_1-norm of **E**, respectively. The balance parameter λ is defined as:

$$\lambda = 1/\sqrt{\max(m, n)} \tag{5}$$

The low-rank minimization due to the correlation between the frames provides a framework for background modelling. Lin et al. [31] solved Equation (4) using a generic ALM method. The Lagrange function can be defined as:

$$L(\mathbf{X}, \mathbf{Y}, \mu) = f(\mathbf{X}) + \langle \mathbf{Y}, h(\mathbf{X}) \rangle + \frac{\mu}{2}\|h(\mathbf{X})\|_F^2 \tag{6}$$

The Lagrange function of Equation (4) is defined as:

$$L(\mathbf{A}, \mathbf{E}, \mathbf{Y}, \mu) = \|\mathbf{A}\|_* + \lambda\|\mathbf{E}\|_1 + \langle \mathbf{Y}, \mathbf{D} - \mathbf{A} - \mathbf{E} \rangle + \frac{\mu}{2}\|\mathbf{D} - \mathbf{A} - \mathbf{E}\|_F^2 \tag{7}$$

where **Y** is the Lagrange multiplier and the penalty parameter μ is a positive scalar parameter. The inexact augmented Lagrange multiplier (IALM) method used to solve the RPCA problem is shown in Algorithm 1. $\mathbf{Y_0}$ has been initialized to $\mathbf{Y_0} = \mathbf{D}/J(\mathbf{D})$ [32], making the objective function value $\langle \mathbf{Y_0}, \mathbf{D} \rangle$ reasonably large. In addition, $J(\mathbf{D}) = max(\|\mathbf{A}\|_2, \lambda^{-1}\|\mathbf{Y}\|_\infty)$, where $\|.\|_\infty$ is the maximum absolute value of the input matrix.

In Step 1 of Algorithm 1, ρ is the learning rate, and μ_0 is the initialization of the penalty parameter that influences the convergence speed. In [31], it is proven that the objective function of the RPCA problem (Equation (4)), which is non-smooth, has an excellent convergence property. In addition, it has been proven that to converge to an optimal solution $(\mathbf{A}^*, \mathbf{E}^*)$ of the RPCA problem, it is necessary for μ_k to be non-decreasing and $\sum_{k=1}^{+\infty} \mu_k^{-1} = +\infty$. The proposed algorithm steps are detailed in the following table.

Algorithm 1: RPCA via IALM method

Input: Data: $\mathbf{D} \in \mathbb{R}^{m \times n}$, balance parameter λ
$\mathbf{Y}_0 = \frac{\mathbf{D}}{J(\mathbf{D})}$; $E_0 = 0$; $\mu_0 > 0$; $\rho > 1$; $k = 0$;
while *not converged* **do**
  // Lines 3-4 update A by solving $\mathbf{A}_{k+1} = \underset{\mathbf{A}}{argmin}\ L(\mathbf{A}, \mathbf{E}_k, Y_k, \mu_k)$
  $(\mathbf{U}, \mathbf{S}, \mathbf{V}) = svd(\mathbf{D} - \mathbf{E}_k + \mu_k^{-1}\mathbf{Y}_k)$;
  $\mathbf{A}_{k+1} = \mathbf{U}\mathcal{S}_{\mu_k^{-1}}[\mathbf{S}]\mathbf{V}^T$;
  // Line 5 update E by solving $\mathbf{E}_{k+1} = \underset{\mathbf{E}}{argmin}\ L(\mathbf{A}_{k+1}, \mathbf{E}, \mathbf{Y}_k, \mu_k)$
  $\mathbf{E}_{k+1} = \mathcal{S}_{\lambda\mu_k^{-1}}[\mathbf{D} - \mathbf{A}_{k+1} + \mu_k^{-1}\mathbf{Y}_k]$;
  $\mathbf{Y}_{k+1} = \mathbf{Y}_k + \mu_k(\mathbf{D} - \mathbf{A}_{k+1} - \mathbf{E}_{k+1})$;
  Update μ_k *to* μ_{k+1};
  $k \leftarrow k + 1$;
end
Output: $(\mathbf{A}_k, \mathbf{E}_k)$

3.2. State-of-the-Art

Pulsed thermography has been extensively investigated as a mean to detect defects for a wide variety of applications. Several processing techniques have been proposed and have been thoroughly reported. References [33–35] provide a detailed review of various methods. Principal component thermography (PCT) [27], pulsed phase thermography (PPT) [9] and the partial least squares thermography (PLST) [11] are among the most effective.

In this paper, a computed low-rank matrix was used prior to or after the application of PCT, PPT and PLST in the PT regime for comparative purposes.

3.2.1. PCT

PCT was introduced by Rajic et al. [6,27] based on the popular multivariate statistical method, principal component analysis (PCA) [36]. This method constructs a set of empirical orthogonal functions (EOFs), which are strong representations of complex input signals. In IRNDT, PCT tends to project data in the orthogonal space that maximizes the variance of projected data. The EOFs will represent the most critical variability of the data, respectively. In general, the given sequence can be represented with a few EOFs. Typically, the thermal sequence of thousands of frames can be replaced by a maximum of ten EOFs.

3.2.2. PPT

Pulsed phase thermography was introduced by Maldague et al. [9]. Each pixel in the thermal data sequence can be transformed using the one-dimensional discrete Fourier transform (DFT) to extract amplitude and phase information from PT data. Unlike raw thermal data, phase transform ϕ is less sensitive to environmental reflections, emissivity variations, non-uniform heating, surface geometry and orientation. The most important characteristic of this method is that it can provide qualitative and quantitative analysis. For instance, a straightforward formulation of depth estimation (z) using the thermal diffusion length μ and the blind frequency f_b is:

$$z = C_1.\mu = C_1.\sqrt{\frac{\alpha}{\pi.f_b}} \tag{8}$$

where f_b is the frequency at which a given defect has enough contrast to be detected, while C_1 is the empirical constant and calculated after a series of experiments. It has been observed that $C_1 \approx 1$ for amplitude data and a value in the range of 1.5 to 2, with $C_1 = 1.82$, are typically adopted for research similar to that presented in [37]. Therefore,

probing deeper defects using the phase makes it more interesting than the amplitude. More information regarding PPT can be found in [9].

3.2.3. PLST

PLST [12] is based on a statistical correlation method known as partial least squares regression (PLSR). PLST decomposes predictor $X(n \times N)$ and predicted $Y(n \times M)$ matrices into loading (P and Q), score (T and U) vectors and residuals (E and F). The predictor matrix corresponds to the thermal profile, while Y is defined by the observation time during which the thermal sequence was acquired. Mathematically, the PLS model is expressed as:

$$X = TP^T + E \tag{9}$$

$$Y = UQ^T + F \tag{10}$$

In order to select the appropriate number of PLS components, two parameters, i.e., the root mean square error (RMSE) and the percentage variance explained in the X matrix, must be taken into consideration.

3.3. Data Acquisition

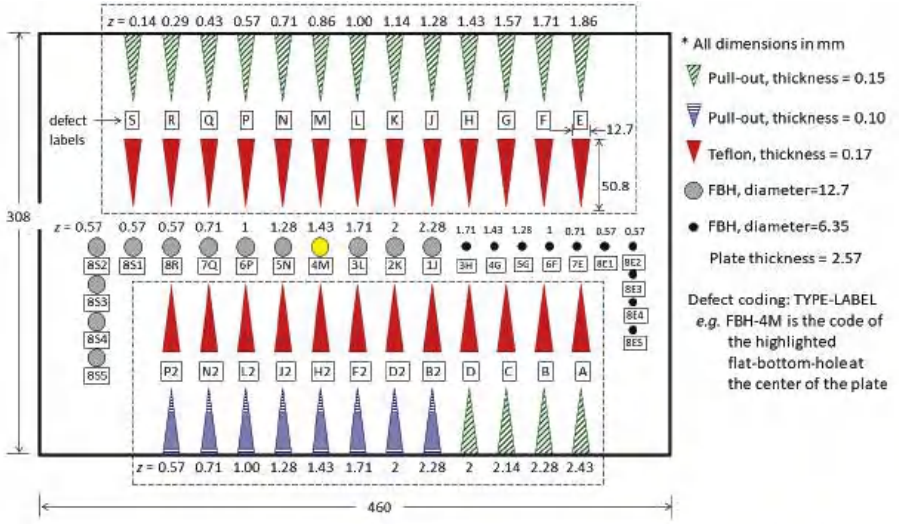
The experiments were carried out on an academic carbon-fibre-reinforced polymer (CFRP) plate (30.8 cm × 46 cm × 2.57 mm) with 73 defects of 3 different types, i.e., 23 round flat-bottom holes (FBH), 25 triangular Teflon inserts, and pullouts. In order to manufacture the pullout defect, a metallic sheet is removed after polymer curing. Therefore, the pullout can only be located at the edge of the part (Figure 1c). The Teflon insert is made of Teflon sheets inserted between plies (Figure 1b). In the case of FBH manufacturing, a hole is drilled to have a flat reflecting surface at the hole bottom at the backside of the sample (Figure 1a). One of the important defects in non-destructive inspection is delamination, which occurs between plies during manufacturing or by fatigue, bearing damage, impact, etc., during the life-cycle. The academic plate used in this study was prepared to investigate the differences in the thermal response of different artificial defect types. Strictly speaking, all artificial defects are at best an approximation of a real delamination. A pull-out seems to be closer to a real delamination (thermally speaking) but is difficult to produce anywhere other than on the borders of the specimen (which implies that the sample must have an open border). Teflon inserts are traditionally employed for other NDT techniques (e.g., ultrasounds) in thermography. However, Teflon behaves significantly different than a real delamination (air) does. Lastly, flat-bottom-holes are easier to produce, though they are open on the rear side of the specimen and possess a much larger volume than a real delamination. The surface of the specimen possesses a fairly good emissivity, so environmental reflections were negligible. Non-uniform heating had a greater impact on all techniques, as can be seen in Section 4.

(**a**) (**b**) (**c**)

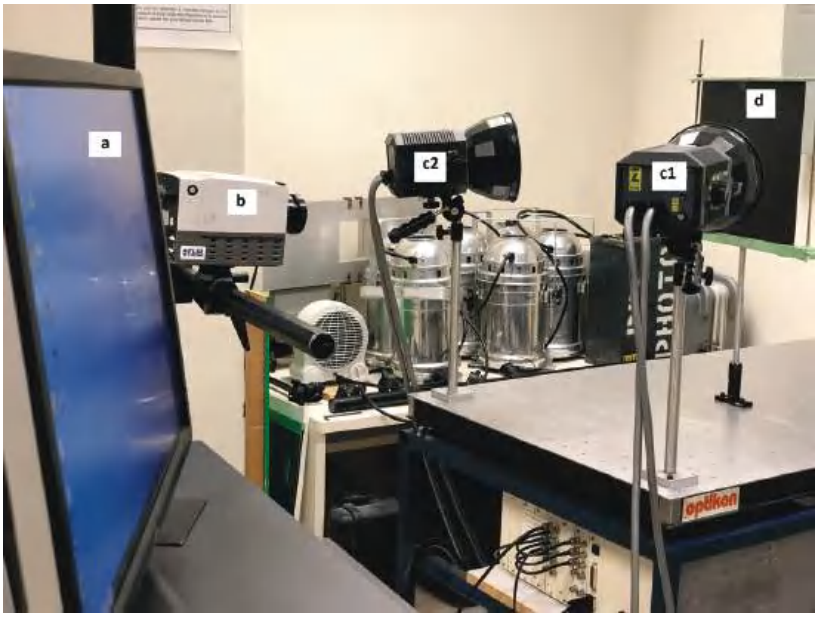
Figure 1. Schematic of a defects in the form of (**a**) flat bottom hole; (**b**) Teflon insert; and (**c**) pullouts.

The defects vary in size, depth and thickness and are presented in Table 1, and the schematic of the plate shows their respective locations in Figure 2a. The thermophysical properties of CFRP involved in the NDE are: k—thermal conductivity (W/m/K), ρ—density (kg/m^3) and c—specific heat capacity (J/kg/K). The other important thermal properties are: $\alpha = k/\rho/c$—thermal diffusivity and $e = \sqrt{k\rho c}$—thermal effusivity. The thermophysical information of the CFRP plate is shown in Table 2. The PT experimental setup, two flash lamps for 5 ms sent a thermal pulse (6.4KJ/flash (Balcar, France)) to

the specimen; a cooled infrared camera (FLIR Phoenix (FLIR Systems, Inc., Wilsonville, Oregon, USA), InSb, midwave, 3–5 mm, Stirling Cooling) with a frame rate of 180 Hz was used to record the temperature profile in the reflection mode (Figure 2b). The technical camera specifications of the thermal camera are presented in Table 3. The data processing was performed on a PC with 56 GB memory and an Intel(R) Core(TM) i7-4820K control processing unit. Infrared images were taken from a distance of 70 cm by the IR camera without pan nor tilt in a controlled environment.



(**a**)



(**b**)

Figure 2. (**a**) CTA CFRP plate, where Z is the defect depth, and labels are used to identify the location of each defect; (**b**) pulsed thermography setup. a, PC; b, IR camera; c1 and c2, left and right flashes; d, CFRP specimen.

Table 1. Defect specifications for the CFRP Plate, Z is the depth of the defect below the inspected surface. Thickness is the defect thickness or thickness of the holes in case of the FBH type of defect.

Defect Code	Z (mm)	Dimensions (mm)	Thickness (mm)	Defect Code	Z (mm)	Dimensions (mm)	Thickness (mm)	Defect Code	Z (mm)	Dimensions (mm)	Thickness (mm)
Teflon Inserts				**Pull-Outs**				**FlatBottom Holes**			
Tef-A	2.43	12.7 × 50.8	0.17	PO15-A	2.43	12.7 × 50.8	0.15	FBH-1J	2.28	12.70	0.29
Tef-B	2.28	12.7 × 50.8	0.17	PO15-B	2.28	12.7 × 50.8	0.15	FBH-2K	2.00	12.70	0.57
Tef-C	2.14	12.7 × 50.8	0.17	PO15-C	2.14	12.7 × 50.8	0.15	FBH-3L	1.71	12.70	0.86
Tef-D	2.00	12.7 × 50.8	0.17	PO15-D	2.00	12.7 × 50.8	0.15	FBH-4M	1.43	12.70	1.14
Tef-E	1.86	12.7 × 50.8	0.17	PO15-E	1.86	12.7 × 50.8	0.15	FBH-5N	1.28	12.70	1.29
Tef-F	1.71	12.7 × 50.8	0.17	PO15-F	1.71	12.7 × 50.8	0.15	FBH-6P	1.00	12.70	1.57
Tef-G	1.57	12.7 × 50.8	0.17	PO15-G	1.57	12.7 × 50.8	0.15	FBH-7Q	0.71	12.70	1.86
Tef-H	1.43	12.7 × 50.8	0.17	PO15-H	1.43	12.7 × 50.8	0.15	FBH-8R	0.57	12.70	2.00
Tef-J	1.28	12.7 × 50.8	0.17	PO15-J	1.28	12.7 × 50.8	0.15	FBH-8S1	0.57	12.70	2.00
Tef-K	1.14	12.7 × 50.8	0.17	PO15-K	1.14	12.7 × 50.8	0.15	FBH-8S2	0.57	12.70	2.00
Tef-L	1.00	12.7 × 50.8	0.17	PO15-L	1.00	12.7 × 50.8	0.15	FBH-8S3	0.57	12.70	2.00
Tef-M	0.86	12.7 × 50.8	0.17	PO15-M	0.86	12.7 × 50.8	0.15	FBH-8S4	0.57	12.70	2.00
Tef-N	0.71	12.7 × 50.8	0.17	PO15-N	0.71	12.7 × 50.8	0.15	FBH-8S5	0.57	12.70	2.00
Tef-P	0.57	12.7 × 50.8	0.17	PO15-P	0.57	12.7 × 50.8	0.15	FBH-3H	1.71	6.35	0.86
Tef-Q	0.43	12.7 × 50.8	0.17	PO15-Q	0.43	12.7 × 50.8	0.15	FBH-4G	1.43	6.35	1.14
Tef-R	0.29	12.7 × 50.8	0.17	PO15-R	0.29	12.7 × 50.8	0.15	FBH-5G	1.28	6.35	1.29
Tef-S	0.14	12.7 × 50.8	0.17	PO15-S	0.14	12.7 × 50.8	0.15	FBH-6F	1.00	6.35	1.57
Tef-B2	2.28	12.7 × 50.8	0.17	PO10-B2	2.28	12.7 × 50.8	0.10	FBH-7E	0.71	6.35	1.86
Tef-D2	2.00	12.7 × 50.8	0.17	PO10-D2	2.00	12.7 × 50.8	0.10	FBH-8E1	0.57	6.35	2.00
Tef-F2	1.71	12.7 × 50.8	0.17	PO10-F2	1.71	12.7 × 50.8	0.10	FBH-8E2	0.57	6.35	2.00
Tef-H2	1.43	12.7 × 50.8	0.17	PO10-H2	1.43	12.7 × 50.8	0.10	FBH-8E3	0.57	6.35	2.00
Tef-J2	1.28	12.7 × 50.8	0.17	PO10-J2	1.28	12.7 × 50.8	0.10	FBH-8E4	0.57	6.35	2.00
Tef-L2	1.00	12.7 × 50.8	0.17	PO10-L2	1.00	12.7 × 50.8	0.10	FBH-8E5	0.57	6.35	2.00
Tef-N2	0.71	12.7 × 50.8	0.17	PO10-N2	0.71	12.7 × 50.8	0.10				
Tef-P2	0.57	12.7 × 50.8	0.17	PO10-P2	0.57	12.7 × 50.8	0.10				

Table 2. Thermal properties of the CFRP.

Material	Density ρ (kg/m^3)	Specific Heat c (J/kg°K)	Conductivity k (W/(m°K))	Diffusivity α (m^2/s 10^{-7})	Effisivity e (W s$^{0.5}$/(m^2 °K))
CFRP ($\perp$)	1600	1200	0.8	4.167	1239.3

Table 3. Technical specification of Phoenix Thermal Camera from FLIR Systems.

Thermal Camera Specifications	
Parameters	**Values**
Detector	Indium Antimonide (InSb)
Spectral Range	1.5–5.0 microns
Cold Filter Bandpass	3.0–5.0 µm standard
Resolution	320 × 256 pixels
Detector size	30 × 30 µm
Well Capacity	18 M electrons
Integration Type	Snapshot

Table 3. *Cont.*

Thermal Camera Specifications	
Parameters	**Values**
Integration Time (Electronic shutter speed)	9 µs to full frame time
Sensor Assembly f/#	f/2.5 standard, f/4.1 optional
Sensor Cooling	Stirling closed cycle cooler; optional Liquid Nitrogen (LN2)
Lens Mount	Bayonet Twist-Lock
Spec Performance (Thermal resolution)	<25 milliKelvin
Dynamic Range	14 bits
Max Frame Rates with RTIE Electronics	320 × 256: 120 frames per sec in full frame; 13.6 kHZ in smallest window (2 × 64)
Max Frame Rates with DAS Electronics	320 × 256: 345 frames per sec in full frame; 38 kHZ in smallest window (2 × 128)

3.4. Metrics

In this section, we added two metrics—one to yield a thermal score indicating thermal anomalies, another to measure the segmentation potential.

3.4.1. Contrast-to-Noise Ratio (CNR)

The signal-to-noise ratio (SNR) is a metric that quantitatively assesses the desired signal quality by estimating the signal level with respect to the background noise. The contrast-to-noise ratio (CNR) is similar to SNR, but it measures the image quality based on the contrast between a defective area and its neighbourhood. Usamentiaga [38] proposed a definition of SNR, which is more robust against noise and image enhancement operations. Equation (11) shows this definition, which has been used in this study. For this purpose, two areas are considered: an area in the defect area (carea) and a region around the defect region as a reference region (narea).

$$CNR = \frac{\mid \mu_{carea} - \mu_{narea} \mid}{\sqrt{\frac{(\sigma_{carea}{}^2 + \sigma_{narea}{}^2)}{2}}} \tag{11}$$

where μ_{carea} and μ_{narea} are the average levels of contrast in carea and narea, respectively; σ_{carea} and σ_{narea} are the standard deviation of the contrast in carea and narea, respectively.

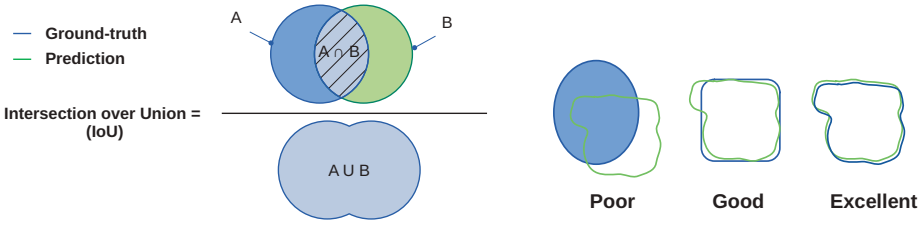
3.4.2. Jaccard Similarity Coefficient Score

The Jaccard similarity coefficient [39] (also known as Jaccard index or Intersection-Over-Union (IoU)) is a statistical method that emphasizes the similarity between two finite datasets (as illustrated in Figure 3):

This approach mathematically represents Equation (12) and is formally defined as the number of the shared members/pixels between two sets (intersection), divided by the total number of members in either set (union) and multiplied by 100. $J(\mathbf{A}, \mathbf{B})$ provides a value between 0 (no similarity) and 1 (identical sets). Hence, the higher the value of IoU, the higher the level of similarities between the two sets (Figure 3b).

$$J(\mathbf{A}, \mathbf{B}) = \frac{|\mathbf{A} \cap \mathbf{B}|}{|\mathbf{A} \cup \mathbf{B}|} = \frac{|\mathbf{A} \cap \mathbf{B}|}{|\mathbf{A}| + |\mathbf{B}| - |\mathbf{A} \cap \mathbf{B}|} \qquad 0 \leq J(\mathbf{A}, \mathbf{B}) \leq 1 \tag{12}$$

For the remainder of this article, we will refer to the low-rank matrix **A** as low-rank matrix (LRM).

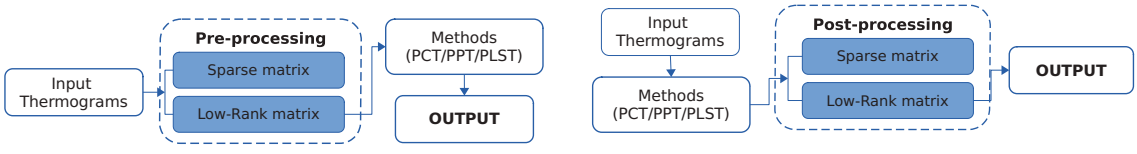


(**a**) (**b**)

Figure 3. (**a**) Jaccard index similarity definition; (**b**) similarity between the ground-truth and the detected area.

3.5. Analysis

The previous section recalls the RPCA we used in our experiments. As described in Figure 4a,b, we conducted two experiments. The main difference between our experiments is that: in the first experiment (Figure 4a), the LRM is computed directly from the raw data; while in the second (Figure 4b), the LRM is computed from the output of the processing methods. For the remainder of this article, we refer to the first experiment as a pre-processing experiment and to the second as a post-processing experiment.



(**a**) (**b**)

Figure 4. (**a**) Using the method for pre-processing; (**b**) Using the method for post-processing.

We chose to compare our approach with three state-of-the-art approaches, principal component thermography (PCT) [6,27], pulsed phase thermography (PPT) [9] and partial least-squares thermography (PLST) [11,12], due to the popularity and simplicity of these methods.

The metrics are computed using different protocols. The defective areas were labelled using LabelMe © [40]. From the border of the defective region, n pixels are considered as a transient region, and from the boundaries of this area, n pixels are automatically counted as a non-defective or sound area. Figure 5 illustrates the aforementioned regions so as to estimate the CNR score. According to Equation (11) and the labelled regions, the average and standard deviation values are obtained for all data.

Regarding the second metric, Figure 6 depicts the automatic segmentation approach and Jaccard index calculation. In our segmentation approach, after the image's contrast correction, a bilateral filter [41] smoothed the image. Then, after applying local thresholding, the small artifacts are removed from the image. The obtained mask from the segmentation step can be compared with the ground truth in order to compute the metric score.

Figure 5. Examples of reference and defect regions. The boundaries of the reference region are between the green and red lines, whilst the defective region is inside the blue line area.

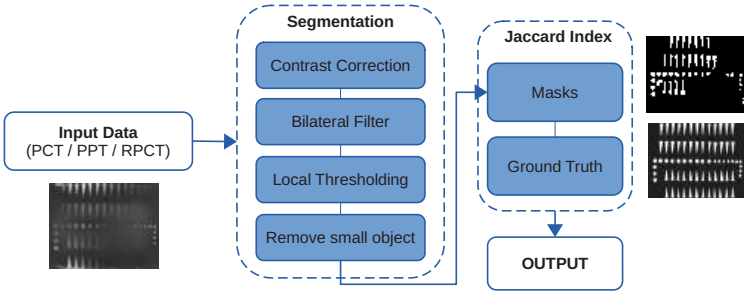


Figure 6. Segmentation and Jaccard index computation flow graph.

4. Results

The original data acquired by pulsed thermography (raw data) is used as pre- and post-processing for different processing approaches. Figure 7 shows some representative results (selected arbitrarily) of the different methods. The first column in Figure 7 results from different techniques on raw data, where the second column presents RPCA results as a pre-processing method, and the last column shows the RPCA approach used as a post-processing method.

Figures 8–10 present the thermal profile across the different lines in images where the defects are either detectable or non-detectable. The first and last lines in each image (green and blue) show the pullout defects profile, while the second and fourth lines (lime and teal) represent the FBHs, and the third line (olive) presents the Teflon inserts profile.

The detailed maximum CNR values of all methods for all defect types are presented in Tables 4–6. The maximum CNR values between different methods are in bold. Figures 11–14 present the maximum CNR value in full sequences for different methods. The CNR values of all defects and all processing techniques were calculated using the defects and reference areas, such as the ones shown in Figure 5.

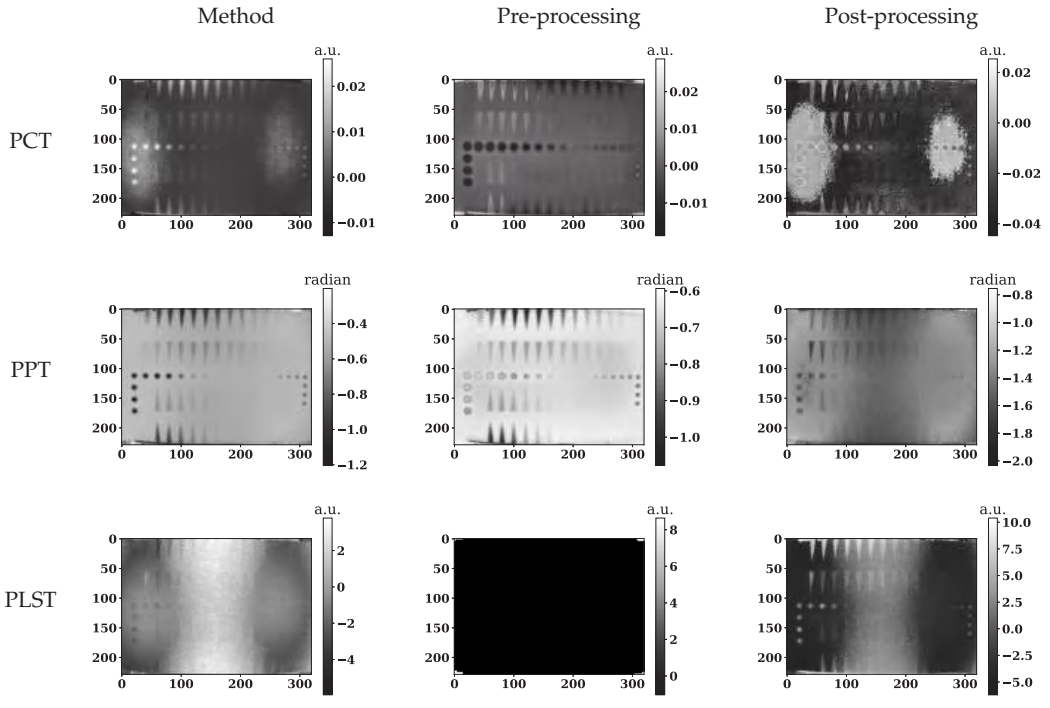


Figure 7. (*1st row*) These images present the 3rd component of PCT data on raw data after using a low-rank matrix for pre-processing and post-processing, respectively. (*2nd row*) These images present PPT data at 0.135 Hz on raw data after using a low-rank matrix for pre-processing and post-processing, respectively. (*3rd row*) These images present the 3rd component of PLST data on raw data after using a low-rank matrix for pre-processing and post-processing, respectively.

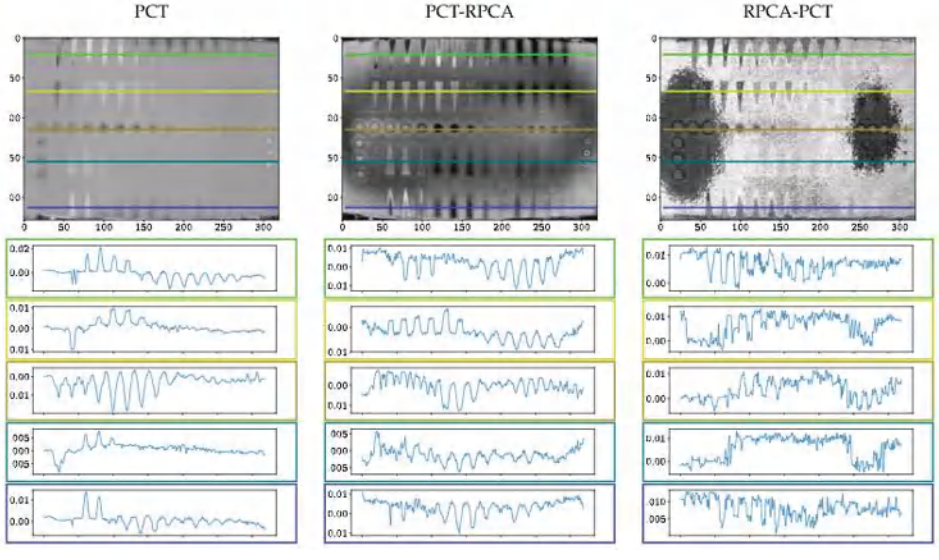


Figure 8. Profiles across the sample after using different processing techniques.

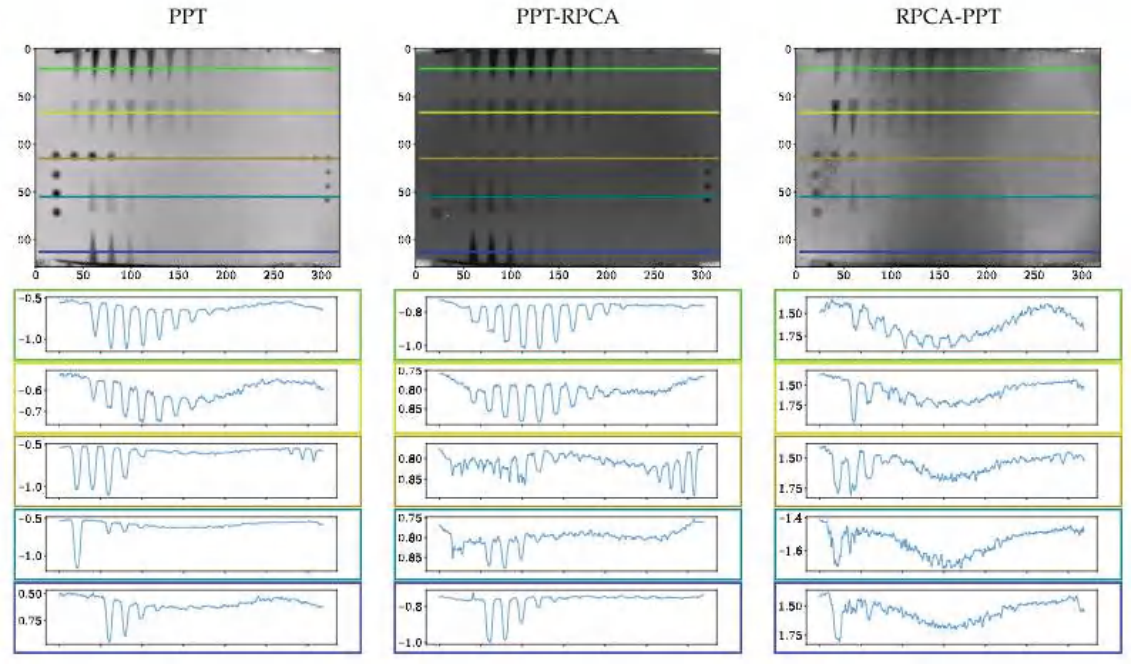


Figure 9. Profiles across the sample after using different processing techniques.

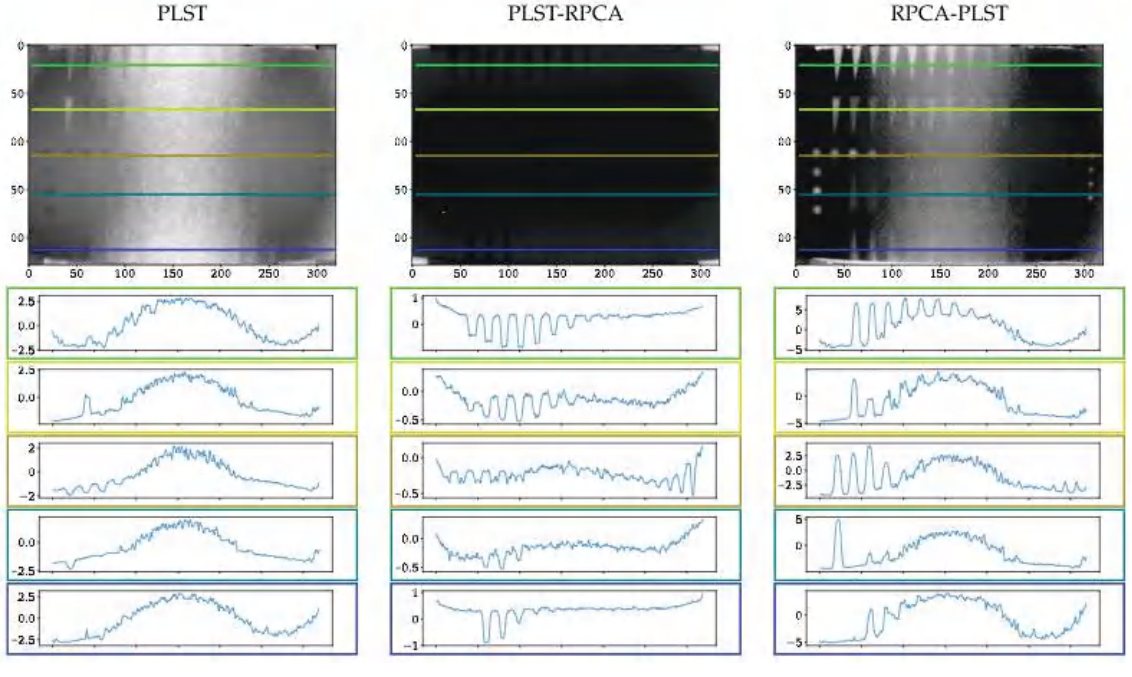


Figure 10. Profiles across the sample after using different processing techniques.

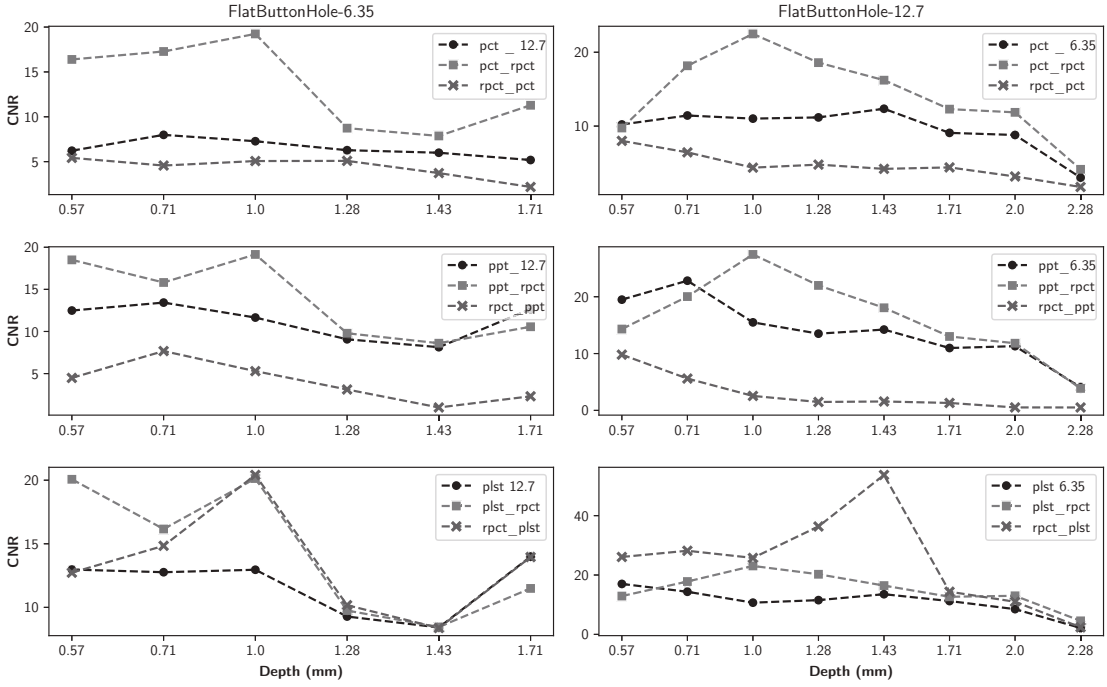


Figure 11. Maximum CNR by different FBHs as a function of defect depth for all data sequences.

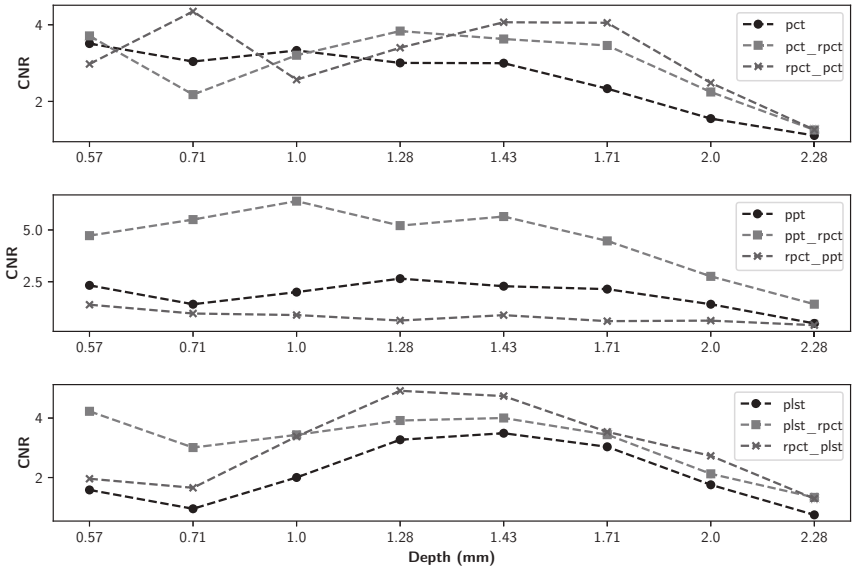


Figure 12. Maximum CNR for pullout-10 as a function of defect depth for all data sequences.

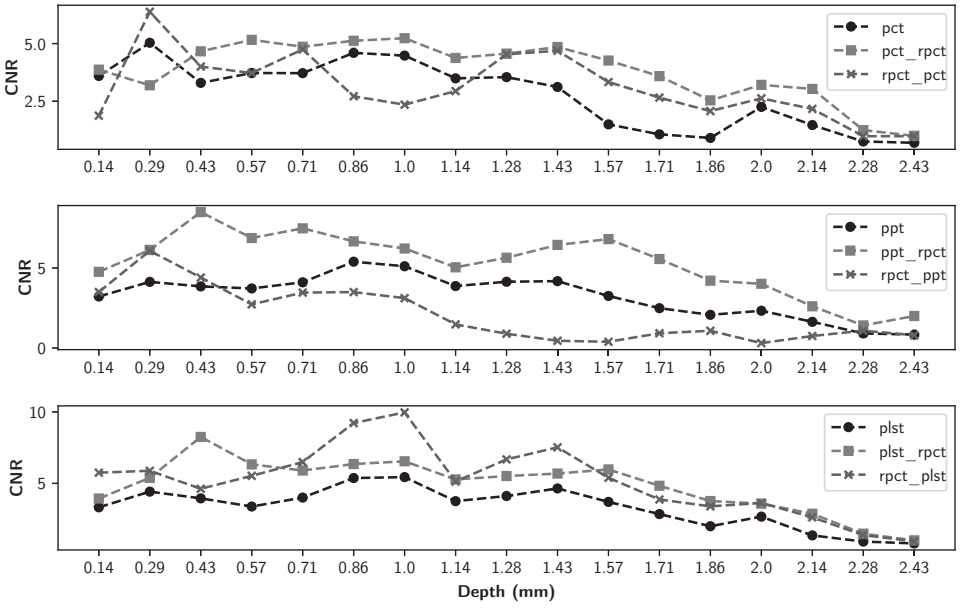


Figure 13. Maximum CNR for pullout-15 as a function of defect depth for all data sequences.

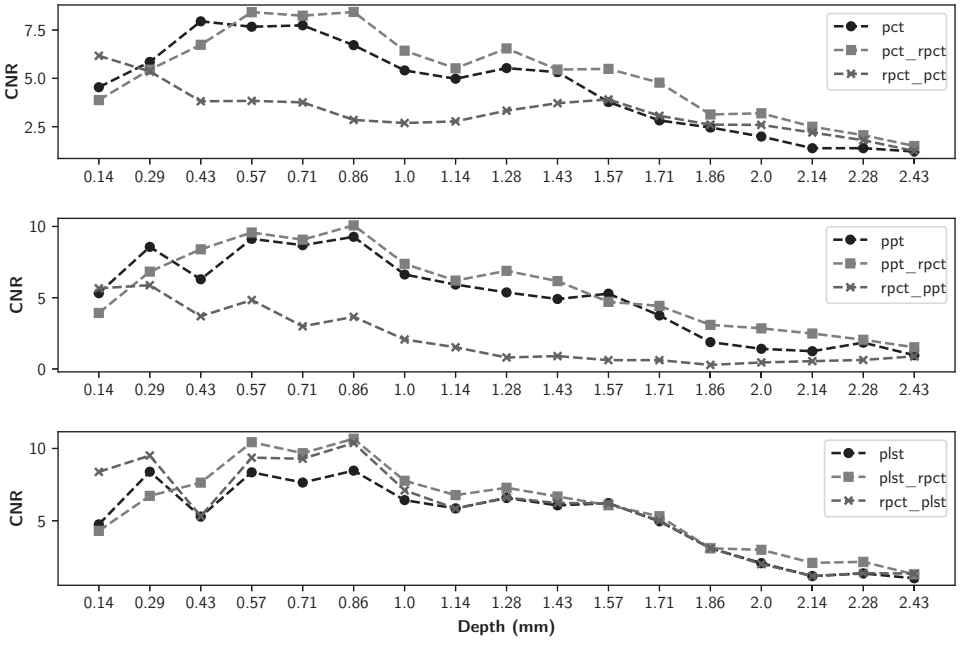


Figure 14. Maximum CNR for teflon insert as a function of defect depth for all data sequences.

Table 4. Maximum CNR values for all data regarding Flat bottom holes in different depths and diameters.

			PCT					PLST					PPT				
Defect	**Z**	**Dim.**	**On Raw data**	**Pre-P.**	**Post-P.**	**Pre-P vs. PCT**	**Post-P vs. PCT**	**On Raw Data**	**Pre-P.**	**Post-P.**	**Pre-P. vs. PLST**	**Post-P. vs. PLST**	**On Raw data**	**Pre-P.**	**Post-P.**	**Pre-P. vs. PPT**	**Post-P. vs. PPT**
FBH-8E1	0.57	6.35	6.22	**16.39**	5.42	163.56%	–12.87%	12.97	**20.07**	12.72	54.71%	–1.91%	12.48	**18.50**	4.48	48.18%	–64.08%
FBH-7E	0.71	6.35	7.98	**17.28**	4.57	116.38%	–42.82%	12.76	**16.16**	**14.84**	26.61%	16.27%	13.43	**15.82**	7.67	17.77%	–42.86%
FBH-6F	1	6.35	7.28	**19.24**	5.07	164.38%	–30.36%	12.96	**20.13**	**20.41**	55.36%	57.54%	11.65	**19.15**	5.29	64.32%	–54.6%
FBH-5G	1.28	6.35	6.28	**8.73**	5.08	39.14%	–18.99%	9.28	**9.75**	**10.17**	5.07%	9.62%	9.07	**9.79**	3.11	7.88%	–65.7%
FBH-4G	1.43	6.35	6	**7.86**	3.73	31.24%	–37.84%	8.42	**8.45**	8.38	0.43%	–0.49%	8.13	**8.61**	0.98	5.82%	–87.99%
FBH-3H	1.71	6.35	5.18	**11.28**	2.18	117.75%	–57.93%	13.99	11.49	13.95	–17.87%	–0.24%	12.65	10.56	2.30	–16.52%	–81.78%
FBH-8R	0.57	12.7	10.22	9.74	8	–4.67%	–21.74%	16.99	12.91	**26.08**	–24.06%	53.45%	19.49	14.31	9.81	–26.56%	–49.67%
FBH-7Q	0.71	12.7	11.43	**18.17**	6.44	58.91%	–43.64%	14.36	**17.80**	**28.17**	24%	96.18%	22.82	20.03	5.6	–12.25%	–75.47%
FBH-6P	1	12.7	11.01	**22.48**	4.37	104.14%	–60.29%	10.68	**23.06**	**25.76**	115.88%	141.17%	15.49	**27.44**	2.54	77.19%	–83.59%
FBH-5N	1.28	12.7	11.17	**18.59**	4.78	66.38%	–57.21%	11.53	**20.25**	**36.35**	75.59%	215.16%	13.5	**21.99**	1.48	62.89%	–89.05%
FBH-4M	1.43	12.7	12.35	**16.22**	4.21	31.35%	–65.92%	13.53	**16.44**	**53.71**	21.49%	296.9%	14.22	**18.05**	1.57	26.97%	–88.99%
FBH-3L	1.71	12.7	9.08	**12.29**	4.40	35.39%	–51.5%	11.22	**12.67**	**14.37**	12.86%	28.04%	10.97	**13.01**	1.31	18.51%	–88.1%
FBH-2K	2	12.7	8.79	**11.86**	3.18	34.85%	–63.88%	8.50	**12.99**	**10.99**	52.79%	29.22%	11.3	**11.82**	0.52	4.58%	–95.41%
FBH-1J	2.28	12.7	3.02	**4.14**	1.76	37.02%	–41.88%	2.15	**4.56**	**2.39**	112.19%	11.12%	4.06	3.87	0.51	-4.83%	–87.4%

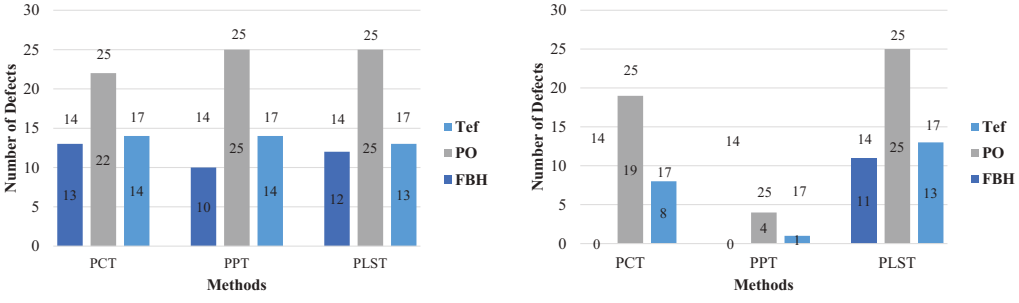
Table 5. Maximum CNR values for all data regarding Teflon inserts in different depths and diameters.

			PCT					PLST					PPT				
Defect	**Z**	**Dim.**	**On Raw data**	**Pre-P.**	**Post-P.**	**Pre-P vs. PCT**	**Post-P vs. PCT**	**On Raw data**	**Pre-P.**	**Post-P.**	**Pre-P. vs. PLST**	**Post-P. vs. PLST**	**On Raw data**	**Pre-P.**	**Post-P.**	**Pre-P. vs. PPT**	**Post-P. vs. PPT**
Tef-S	0.14	12.7 × 50.8	4.54	3.88	**6.17**	−14.66%	35.82%	4.75	4.3	**8.38**	−9.45%	76.38%	5.32	3.93	**5.66**	−26.07%	6.41%
Tef-R	0.29	12.7 × 50.8	5.85	5.45	5.36	−6.87%	−8.39%	8.39	6.72	**9.51**	−19.92%	13.33%	8.57	6.83	5.88	−20.29%	−31.39%
Tef-Q	0.43	12.7 × 50.8	7.95	6.74	3.81	−15.28%	−52.04%	5.29	**7.64**	**5.31**	44.45%	0.51%	6.29	**8.4**	3.7	33.43%	−41.24%
Tef-P	0.57	12.7 × 50.8	7.67	**8.43**	3.84	9.83%	−50.01%	8.35	**10.44**	**9.36**	24.97%	12.14%	9.13	**9.58**	4.84	4.89%	−47.02%
Tef-N	0.71	12.7 × 50.8	7.74	**8.24**	3.76	6.46%	−51.48%	7.64	**9.67**	**9.29**	26.58%	21.54%	8.69	**9.07**	2.99	4.43%	−65.56%
Tef-M	0.86	12.7 × 50.8	6.72	**8.44**	2.85	25.58%	−57.58%	8.47	**10.67**	**10.38**	26.07%	22.56%	9.27	**10.08**	3.66	8.72%	−60.55%
Tef-L	1	12.7 × 50.8	5.41	**6.43**	2.69	18.86%	−50.32%	6.43	**7.77**	**7.11**	20.78%	10.52%	6.63	**7.38**	2.07	11.26%	−68.83%
Tef-K	1.14	12.7 × 50.8	4.98	**5.52**	2.78	10.87%	−44.27%	5.85	**6.78**	5.85	15.75%	−0.15%	5.92	**6.2**	1.52	4.76%	−74.28%
Tef-J	1.28	12.7 × 50.8	5.53	**6.55**	3.33	18.4%	−39.83%	6.58	**7.28**	**6.61**	10.72%	0.46%	5.37	**6.89**	0.81	28.21%	-84.99%
Tef-H	1.43	12.7 × 50.8	5.32	**5.46**	3.72	2.5%	−30.19%	6.07	**6.68**	**6.22**	10.05%	2.47%	4.91	**6.17**	0.9	25.65%	−81.63%
Tef-G	1.57	12.7 × 50.8	3.78	**5.49**	**3.91**	45.37%	3.44%	6.21	6.07	6.21	−2.21%	−0.03%	5.28	4.71	0.62	−10.9%	−88.29%
Tef-F	1.71	12.7 × 50.8	2.82	**4.78**	**3.06**	69.26%	8.54%	4.96	**5.32**	**5**	7.16%	0.75%	3.75	**4.42**	0.62	17.85%	−83.54%
Tef-E	1.86	12.7 × 50.8	2.46	**3.13**	**2.6**	27.47%	5.98%	3.1	3.1	3.1	−0.26%	−0.1%	1.88	**3.09**	0.28	64.31%	−84.92%
Tef-D	2	12.7 × 50.8	1.99	**3.19**	**2.59**	60.61%	30.41%	2.07	**2.99**	2.01	44.54%	−2.95%	1.41	**2.85**	0.46	101.45%	−67.82%
Tef-C	2.14	12.7 × 50.8	1.39	**2.5**	**2.2**	80.36%	58.63%	1.18	**2.09**	1.18	77.64%	0%	1.25	**2.49**	0.55	99.52%	−56.02%
Tef-B	2.28	12.7 × 50.8	1.39	**2.06**	**1.8**	48.47%	29.82%	1.36	**2.16**	**1.38**	59.35%	1.88%	1.85	**2.05**	0.63	10.86%	-65.83%
Tef-A	2.43	12.7 × 50.8	1.21	**1.5**	**1.24**	23.56%	2.39%	1.02	**1.29**	**1.35**	26.42%	32%	0.97	**1.52**	0.88	57.56%	−8.49%

Table 6. Maximum CNR values for all data regarding Pullouts in different depths and diameters.

			PCT					PLST					PPT			
Defect	**Z**	**Dim.**	**On Raw data**	**Pre-P.**	**Post-P.**	**Pre-P vs. PCT**	**Post-P vs. PCT**	**On Raw data**	**Pre-P.**	**Post-P.**	**Pre-P. vs. PLST**	**Post-P. vs. PLST**	**On Raw data**	**Pre-P.**	**Post-P.**	**Pre-P. vs. PPT**	**Post-P. vs. PPT**
PO10-P2	0.57	12.7 × 50.8	3.51	**3.71**	2.98	5.7%	−15.21%	1.58	**4.23**	**1.96**	167.02%	23.75%	2.33	**4.73**	1.391	103.23%	−40.17%
PO10-N2	0.71	12.7 × 50.8	3.04	2.18	**4.35**	−28.38%	42.98%	0.95	**3.01**	**1.66**	216.63%	74.21%	1.41	**5.5**	0.966	288.97%	−31.68%
PO10-L2	1	12.7 × 50.8	3.33	3.21	2.57	−3.75%	−22.88%	2	**3.44**	**3.38**	71.58%	68.78%	2	**6.39**	0.892	219.71%	−55.38%
PO10-J2	1.28	12.7 × 50.8	3.01	**3.84**	**3.4**	27.65%	13.18%	3.27	**3.92**	**4.92**	19.79%	50.34%	2.65	**5.21**	0.628	96.57%	−76.3%
PO10-H2	1.43	12.7 × 50.8	3	**3.63**	**4.07**	21.05%	35.62%	3.49	**4**	**4.74**	14.67%	35.79%	2.28	**5.65**	0.887	147.24%	−61.16%
PO10-F2	1.71	12.7 × 50.8	2.33	**3.46**	**4.05**	48.24%	73.61%	3.03	**3.44**	**3.54**	13.48%	16.58%	2.14	**4.47**	0.598	108.63%	−72.1%
PO10-D2	2	12.7 × 50.8	1.55	**2.25**	**2.48**	44.72%	60.05%	1.76	**2.13**	**2.73**	21.01%	55.41%	1.41	**2.76**	0.623	95.19%	−55.94%
PO10-B2	2.28	12.7 × 50.8	1.11	**1.26**	**1.26**	13.05%	13.59%	0.75	**1.33**	**1.29**	78.79%	73.15%	0.49	**1.42**	0.403	189.57%	−17.59%
PO15-S	0.14	12.7 × 50.8	3.6	**3.87**	1.87	7.62%	−47.93%	3.33	**3.94**	**5.75**	18.48%	72.78%	3.22	**4.76**	**3.498**	47.84%	8.67%
PO15-R	0.29	12.7 × 50.8	5.04	3.19	**6.38**	−36.62%	26.69%	4.42	**5.39**	**5.88**	21.79%	32.96%	4.13	**6.12**	**6.073**	48.21%	46.97%
PO15-Q	0.43	12.7 × 50.8	3.3	**4.67**	**4**	41.54%	21.42%	3.95	**8.25**	**4.62**	108.57%	16.94%	3.85	**8.49**	**4.402**	120.52%	14.37%
PO15-P	0.57	12.7 × 50.8	3.72	**5.16**	**3.72**	38.69%	0.05%	3.38	**6.33**	**5.53**	87.11%	63.41%	3.72	**6.87**	2.728	84.62%	−26.67%
PO15-N	0.71	12.7 × 50.8	3.72	**4.86**	**4.75**	30.8%	27.68%	4.01	**5.9**	**6.49**	47.19%	61.94%	4.11	**7.47**	3.462	82.03%	−15.68%
PO15-M	0.86	12.7 × 50.8	4.6	**5.12**	2.72	11.44%	−40.94%	5.38	**6.35**	**9.22**	18.2%	71.53%	5.39	**6.66**	3.494	23.54%	−35.19%
PO15-L	1	12.7 × 50.8	4.48	**5.23**	2.35	16.83%	−47.53%	5.44	**6.55**	**9.96**	20.38%	83.21%	5.11	**6.22**	3.117	21.72%	−38.97%
PO15-K	1.14	12.7 × 50.8	3.49	**4.38**	2.94	25.31%	−15.83%	3.76	**5.28**	**5.14**	40.36%	36.63%	3.87	**5.04**	1.476	30.23%	−61.86%
PO15-J	1.28	12.7 × 50.8	3.55	**4.56**	**4.54**	28.72%	27.93%	4.12	**5.51**	**6.68**	33.88%	62.17%	4.14	**5.64**	0.895	36.18%	−78.38%
PO15-H	1.43	12.7 × 50.8	3.12	**4.85**	**4.69**	55.61%	50.38%	4.65	**5.69**	**7.52**	22.38%	61.81%	4.18	**6.44**	0.453	54.13%	−89.15%
PO15-G	1.57	12.7 × 50.8	1.5	**4.26**	**3.33**	184.84%	122.65%	3.71	**5.98**	**5.39**	61.21%	45.45%	3.25	**6.8**	0.392	109.1%	−87.95%
PO15-F	1.71	12.7 × 50.8	1.06	**3.59**	**2.66**	237.22%	149.62%	2.86	**4.83**	**3.88**	69.32%	35.73%	2.49	**5.56**	0.925	123.3%	−62.84%
PO15-E	1.86	12.7 × 50.8	0.91	**2.54**	**2.07**	179.52%	128.19%	2.01	**3.77**	**3.4**	87.79%	69.29%	2.07	**4.21**	1.079	102.84	−47.97
PO15-D	2	12.7 × 50.8	2.25	**3.21**	**2.63**	42.41%	16.46%	2.68	**3.59**	**3.63**	34.03%	35.67%	2.33	**4.01**	0.309	72.26	−86.73
PO15-C	2.14	12.7 × 50.8	1.47	**3.03**	**2.17**	106.75%	47.78%	1.36	**2.88**	**2.63**	111%	92.6%	1.64	**2.61**	0.752	59.22	−54.09
PO15-B	2.28	12.7 × 50.8	0.75	**1.25**	**0.98**	65.25%	30.24%	0.93	**1.48**	**1.38**	58.89%	48.18%	0.91	**1.41**	**1.085**	55.04	18.97
PO15-A	2.43	12.7 × 50.8	0.7	**1**	**0.98**	43.19%	41.03%	0.81	**1.01**	**0.95**	25.84%	17.39%	0.84	**2**	0.802	137.81	−4.64

Figure 15a,b illustrate the numbers of enhanced defects using pre- and post-processing, respectively. The numbers inside the columns represent the enhanced defects when using different techniques, and the number above the columns are the total number of defects in each case.



(**a**) (**b**)

Figure 15. Number of defects that are enhanced for each experiment. (**a**) Results of the pre-processing experiments. (**b**) Results of the post-processing experiments.

The best Jaccard index for all data sequences for different methods is shown in Table 7.

Table 7. Jaccard index values for different methods on segmented data.

Method	On Raw Data	Pre_Processing	Post_Processing
PCT	60.43	64.08	53.94
PPT	61.19	62.82	55
PLST	50.66	55.36	55.35

Figure 7 illustrates selected results from different methods. In this figure, the first image from each row presents the selected technique on raw data (PCT, PPT or PLST); the second and third images show the effect of using the LRM as a pre- and post-processing method.

Our segmentation approach was evaluated by the Jaccard index presented in Table 7.

5. Discussion

Figure 7 implies that although pre-processing can reduce the non-uniform heating impact, post-processing accentuates this effect. Thermal profiles of different methods across the different lines are shown in Figures 8–10. As depicted in the graphs, the flat thermal profiles show the non-defective or sound area, and when the amplitude is increased or decreased, the available discontinuities can be guaranteed. The application of pre-processing before PCT and PPT approaches improved the defect detection; also, in the case of PLST, both pre- and post-processing can increase the detection of anomalies. In addition, the graphs show similar results with quantitative metrics, which will be explained later. From Tables 4–6 and Figures 11–14, one can note that the results from the pre-processing experiments are noticeably better than those obtained from the post-processing experiment. Note that these results are compared with results obtained without using low-rank matrices for both experiments. For the PCT method, one can note:

- The pre-processing experiments have led to a clear improvement of the results, regardless of the defect type. For 13 of the 14 FBH defects, one can observe an increase in the CNR score. The ratio of this improvement varies from 31.24% to 163.56%. The CNR scores obtained for the PO defects show a higher score in 22 of the 25 defects,

with a ratio that varies from 0.43% to 115.88%. Similarly, the CNR scores obtained for the Teflon inserts also show a CNR score increase for 14 of the 17 defects. The ratio of this improvement varies from 2.5% to 80.36%.

- The results of the post-processing experiments do not show any improvement for the FBH defects. Nevertheless, for the PO defects, one can note that there is a higher CNR score for 19 of 25 defects. The ratio of this improvement varies from 0.05% to 149.62%. For Teflon defects, 8 of the 17 defects have a higher CNR score, with a ratio between 2.39% and 58.63%.

From the PPT method results, one can observe:

- As already observed with the PCT, the results of the pre-processing experiments offer an improvement for every type of defect. For 10 of the 14 FBH defects, one can observe that their CNR score increases, with a ratio between 4.58% and 77.19%. The PO defects show an increase in the CNR score for all of the defects. The ratio of improvement varies from 21.72% to 288.97%. For Teflon inserts, the number of defects with a higher CNR is similar to what was observed for the previous method, with 14 of the 17 defects with an improved CNR value. The ratio of improvement varies from 4.43% to 101.45%.
- The results obtained for the post-processing experiment show very little improvement. No improvement at all was recorded for the FBH. For the PO defects, 4 of the 25 defects had an increased CNR value, with a ratio between 8.67% and 46.97%. Only one Teflon defect of the 17 defects had its CNR increased by a ratio of 6.41%.

Finally, from the PLST method results, one can note:

- The pre-processing experiments shows a similar trend as the trend observed for the two other methods. For 12 of the 14 FBH defects, the CNR score increased, with a ratio from 0.43% to 115.88%. All of the PO defects have their CNR score increased, with a ratio between 13.48% and 216.63%. Finally, for the Teflon insert, 13 defects of the 17 obtained an increased CNR score, with a ratio between 7.16% and 77.64%.
- For the post-processing approach, one can note that the results are quite similar to those obtained during the pre-processing experiments. For 11 of the 17 FBH defects, an increase in the CNR value was observed, with a ratio from 9.62% to 296.9%. All of the PO defects show an improvement of their CNR score, ranging from 16.98% to 92.6%. For 13 of the 17 Teflon defects, the CNR score has improved, with a ratio from 0.46% to 76.38%.

Moreover, as indicated in Figures 11–14, regarding the relative depths, in all cases (FBHs, POs and TEFs), the deeper the defect, the lower the CNR value (as expected). Comparing the two experiments, one can observe that the pre-processing experiment leads to a larger number of defective regions for PCT and PPT methods than the post-processing experiments. Nevertheless, this observation is not valid for the PLST method, where the results are pretty similar in both experiments. For the PO defect, the increase in terms of CNR score is higher in the pre-processing experiments; the mean ratio of improvement is 2.6 times higher than it is for the post-processing experiments. Similarly, the mean ratio of improvement for the Teflon defects is 1.7 times higher in the pre-processing experiment than in the post-processing experiments. Nonetheless, the mean improvement ratio is 2.5 times higher in the post-processing experiment than in the pre-processing experiment. To conclude, our results show that computing an LRM from the raw data before applying any state-of-the-art method significantly improves the results of the method. In the particular case of FBH defects, one can consider computing an LRM before and after the method.

As one can note in Table 7 and see in Figure 15b, using the LRM, prior to the state-of-the-art processing method, leads to better Jaccard index scores and therefore segmentation in all cases. One can also note that the Jaccard index score for the PLST method does not change much between the pre-processing and post-processing experiments. The Jaccard index score for the PCT and PPT methods decreases noticeably for the segmentation of the

post-processing experiment results compared with the segmentation of the raw data. This indicates that the results of the segmentation worsen.

6. Conclusions

The present study investigates the benefits of the low-rank matrices for pulsed thermography. The investigation conducted for this study focuses on enhancing defective regions located within a reference sample of CFRP. The sample we used had three types of defects. Two experiments were conducted: during the first experiment, the low-rank matrix was computed from the raw data before applying any processing. During the second experiment, the low-rank matrix is computed from the output of a method, after it was applied on raw data. For both experiments, we used PPT [9], PCT [6,27] and PLST [11,12]. Two figures of merit, the contrast-to-noise ratio (CNR) and the Jaccard similarity coefficient, were used to evaluate the results quantitatively.

Our results conclude that using a low-rank matrix, when used as a pre-processing method, noticeably improves the results of all of the techniques. The low-rank matrix reconstruction effectively reduces the noise and non-uniform heating. When used as a post-processing method, the results vary from one method to another. The results indicate that pre-processing can improve 67.12% of PCT results more than post-processing, especially regarding FBHs (the detectability of FBHs, pullouts and Teflon inserts was increased to 92.86%, 88% and 82.35%, respectively). Furthermore, pre-processing has a better effect on PPT results (67.12% of the defects were detected) than post-processing. For FBHs, pullouts and Teflon inserts, the detectability of defects reached 71.43%, 100% and 82.35%. The detectability of pullouts and Teflon insert defects in both pre- and post-processing has improved, reaching 100% and 76.47%, respectively; however, the detectability is better after using pre-processing in the PLST method. In addition, when used on the output of PLST, the low-rank matrix reconstruction still shows better results than the PLST alone. Nonetheless, this conclusion is not shared for both PPT and PCT. The Jaccard index proved that pre-processing can improve the segmentation potential in all aforementioned methods. In the case of PLST, improvements were made for both pre-processing and post-processing.

This study presents very promising results regarding the improvement of anomaly detection in pulsed thermography in CFRPs. To make the proposed approach more practical in NDT techniques, future research will be directed towards the application of pre- and post-processing on a wider range of materials.

Author Contributions: Conceptualization and methodology, S.E., J.R.F., L.-D.T., M.K., C.I.-C. and X.P.V.M.; data analysis and processing, J.R.F. and S.E., experimental data acquisition, M.K., L.-D.T. and C.I.-C.; resources, L.-D.T. and X.P.V.M.; writing—original draft preparation, S.E. and J.R.F.; writing—review and editing, M.K., L.-D.T., C.I.-C. and X.P.V.M.; and supervision, X.P.V.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research is funded by NSERC CREATE « oN DuTy!» Initiative, NSERC DG program, and the Canada Research Chair in Multipolar Infrared Vision (MIVIM). Part of the funding also comes from LDCOMP collaborative R&D proposal jointly funded by the Ministère de l'Économie et de l'Innovation - Québec (MEI) (File number: 2018-PI-1-SQA) and SKYWIN (Wallonie, Belgium, Convention n° 8188). The authors wish to thank also the following sponsors: XIe Commission mixte permanente Wallonie-Bruxelles-Québec 2019-2021 (project 11.812).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing not applicable.

Conflicts of Interest: The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

ASM	Active Shape Model
ALM	Augmented Lagrangian Multiplier
APG	Accelerated Proximal Gradient
a.u	arbitrary units
CFRP	Carbon Fiber Reinforced Plastic
CIS	Cold Image Subtraction
CNN	Convolutional neural network
CNR	Contrast to Noise Ratio
DFT	DiscreteFourier Transform
DRPCA	Double Robust Principal Component Analysis
EALM	Exact Augmented Lagrange Multiplier
ECT	Eddy Current Thermography
ECPT	Eddy Current Pulsed Thermography
EOF	Empirical Orthogonal Functions
ESPCA	Edge-Group Sparse Principal Component Analysis
ESPCT	Edge-Group Sparse Principal Component Thermography
FBH	Flat Bottom Holes
GPGPU	General-purpose computing on graphics processing units
IALM	Inexact Augmented Lagrange Multiplier
ICA	Independent Component Analysis
IoU	Intersection over Union
IRNDT	Infrared Non-Destructive Testing
IRT	InfraRed Thermography
LADMAP	Linearized Alternating Direction Method with Adaptive Penalty
LatLRRT	Latent Low-Rank Representation Thermography
LN	Liquid Nitrogen
LRM	Low-Rank Matrix
MWIR	Mid-Wave InfraRed
NDT	Non Destructive Testing
NMF	Non-negative Matrix Factorization
NP	Non-Deterministic Polynomial
OIALM	Orthogonal Inexact Augmented Lagrange Multiplier
PCA	Principal Component Analysis
PCP	Principal Component Pursuit
PCT	Principal Component Thermography
PLS	Partial Least Square
PLSR	Partial Least Square Regression
PLST	Partial Least Square Thermography
PO	pullouts
PPT	Pulsed Phase Thermography
PT	Pulsed Thermography
RMSE	Root Mean Square Error
ROI	region of interest
RPCA	Robust Principal Component Analysis
RPCT	Robust Principal Component Thermography
SNR	Signal to Noise Ratio
SPCA	Sparse Principal Component Analysis
SPCT	Sparse Principal Component Thermography
SVM	Support Vector Machine

Tef	Teflon Inserts
TSR	Thermographic Signal Reconstruction
TRPCA	Tensor RPCA
UT	Ultrasound Testing
WIALM	Weighted contraction IALM

References

1. Vo Dong, P.A.; Azzaro-Pantel, C.; Cadene, A.L. Economic and environmental assessment of recovery and disposal pathways for CFRP waste management. *Resour. Conserv. Recycl.* **2018**, *133*, 63–75. [CrossRef]
2. Abrate, S. Impact on laminated composite materials. *Appl. Mech. Rev.* **1994**, *44*, 155–190. [CrossRef]
3. Fleuret, J.; Ibarra-Castanedo, C.; Ebrahimi, S.; Maldague, X. Latent Low Rank Representation Applied to Thermography. In Proceedings of the 2020 International Conference on Quantitative InfraRed Thermography, Porto, Portugal, 21 September–3 October 2020; pp. 21–30.
4. Khodayar, F.; Lopez, F.; Ibarra-Castanedo, C.; Maldague, X. Optimization of the inspection of large composite materials using robotized line scan thermography. *J. Nondestruct. Eval.* **2017**, *36*, 32. [CrossRef]
5. Shepard, S.M. Advances in pulsed thermography. In *Thermosense XXIII*; Rozlosnik, A.E., Dinwiddie, R.B., Eds.; International Society for Optics and Photonics; SPIE: Bellingham, WA, USA, 2001; Volume 4360, pp. 511–515.
6. Rajic, N. Principal Component thermography for flaw contrast enhancement and flaw depth characterization in composite structures. *Compos. Struct.* **2002**, *58*, 521–528. [CrossRef]
7. Wang, Q.; Hu, Q.; Qiu, J.; Pei, C.; Li, X.; Zhou, H.; Xia, R.; Liu, J. Image enhancement method for laser infrared thermography defect detection in aviation composites. *Opt. Eng.* **2019**, *58*, 103104. [CrossRef]
8. Alard, C.; Lupton, R.H. A Method for Optimal Image Subtraction. *Astrophys. J.* **1998**, *503*, 325–331. [CrossRef]
9. Maldague, X.; Marinetti, S. Pulse phase infrared thermography. *J. Appl. Phys.* **1996**, *79*, 2694–2698. [CrossRef]
10. Ebrahimi, S.; Fleuret, J.; Klein, M.; Théroux, L.D.; Georges, M.; Ibarra-Castanedo, C.; Maldague, X. Robust Principal Component Thermography for Defect Detection in Composites. *Sensors* **2021**, *21*, 2682. [CrossRef]
11. Lopez, F.; Nicolau, V.; Maldague, X.; Ibarra-Castanedo, C. Multivariate infrared signal processing by partial least-squares thermography. In Proceedings of the 16th International Symposium on Applied Electromagnetics and Mechanics, Québec, QC, Canada, 31 July–3 August 2013.
12. Lopez, F.; Ibarra-Castanedo, C.; de Paulo Nicolau, V.; Maldague, X. Optimization of pulsed thermography inspection by partial least-squares regression. *Ndt Int.* **2014**, *66*, 128–138. [CrossRef]
13. Bouwmans, T.; Zahzah, E.H. Robust PCA via principal component pursuit: A review for a comparative evaluation in video surveillance. *Comput. Vis. Image Underst.* **2014**, *122*, 22–34. [CrossRef]
14. Fan, J.; Gao, Y.; Wu, Z.; Li, L. Infrared Dim Small Target Detection Technology Based on RPCA. In *DEStech Transactions on Computer Science and Engineering*; DEStech Publications, Inc.: Lancaster, PA, USA, 2017.
15. Wan, M.; Gu, G.; Qian, W.; Ren, K.; Chen, Q.; Zhang, H.; Maldague, X. Total Variation Regularization Term-Based Low-Rank and Sparse Matrix Representation Model for Infrared Moving Target Tracking. *Remote Sens.* **2018**, *10*, 510. [CrossRef]
16. Xu, Y.; Wu, Z.; Chanussot, J.; Wei, Z. Joint Reconstruction and Anomaly Detection From Compressive Hyperspectral Images Using Mahalanobis Distance-Regularized Tensor RPCA. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 2919–2930. [CrossRef]
17. Zhu, P.; Cheng, Y.; Banerjee, P.; Tamburrino, A.; Deng, Y. A novel machine learning model for eddy current testing with uncertainty. *Ndt Int.* **2019**, *101*, 104–112. [CrossRef]
18. Draganov, I.R.; Mironov, R.P.; Neshov, N.N.; Manolova, A.H. Wild animals population estimation from Thermograph-IC videos using tensor decomposition. In Proceedings of the 14th International Conference On Communications, Electromagnetics and Medical Applications 2019 (CEMA'19), Sofia, Bulgaria, 17–19 October 2019.
19. Draganov, I.; Mironov, R. Tracking of Domestic Animals in Thermal Videos by Tensor Decompositions. In Proceedings of the New Approaches for Multidimensional Signal Processing: Proceedings of International Workshop, NAMSP 2020, Sofia, Bulgaria, 9–11 July 2020.
20. Liang, Y.; Bai, L.; Shao, J.; Cheng, Y. Application of Tensor Decomposition Methods In Eddy Current Pulsed Thermography Sequences Processing. In Proceedings of the 2020 International Conference on Sensing, Measurement & Data Analytics in the Era of Artificial Intelligence (ICSMD), Xi'an, China, 15–17 October 2020.
21. Li, G.; Zheng, Z.; Shao, Y.; Shen, J.; Zhang, Y. Automated Tire Visual Inspection Based on Low Rank Matrix Recovery. Available online: https://www.researchgate.net/publication/347083889_Automated_Tire_Visual_Inspection_Based_on_Low_Rank_Matrix_Recovery/fulltext/5fdd1aaf299bf14088228f8a/Automated-Tire-Visual-Inspection-Based-on-Low-Rank-Matrix-Recovery.pdf (accessed on 21 October 2021).
22. Wu, T.; Gao, B.; Woo, W.L. Hierarchical low-rank and sparse tensor micro defects decomposition by electromagnetic thermography imaging system. *Philos. Trans. R. Soc.* **2020**, *378*, 20190584. [CrossRef]
23. Cao, J.; Yang, G.; Yang, X.; Li, J. A Visual Surface Defect Detection Method Based on Low Rank and Sparse Representation. Available online: http://www.ijicic.org/ijicic-160104.pdf (accessed on 21 October 2021).
24. Wang, Q.; Paynabar, K.; Pacella, M. Online automatic anomaly detection for photovoltaic systems using thermography imaging and low rank matrix decomposition. *J. Qual. Technol.* **2021**, 1–14. [CrossRef]

25. Kaur, K.; Mulaveesala, R. Statistical Post-processing Approaches for Active Infrared Thermography: A Comparative Study. In Proceedings of the 2021 IEEE 11th Annual Computing and Communication Workshop and Conference (CCWC), Las-Vegas, NV, USA, 7–30 January 2021.
26. Sun, S.; Ren, H.; Dan, T.; Wei, W. 3D segmentation of lungs with juxta-pleural tumor using the improved active shape model approach. *Technol. Health Care* **2021**, *29*, 385–398. [CrossRef]
27. Rajic, N. *Principal Component Thermography*; Technical report; Defence Science and Technology Organisation: Victoria, Australia, 2002.
28. Hermosilla-Lara, S.; Joubert, P.Y.; Placko, D.; Lepoutre, F.; Piriou, M. Enhancement of open-cracks detection using a principal component analysis/wavelet technique in photothermal nondestructive testing. In Proceedings of the 6th International Conference on Quantitative InfraRed Thermography, Dubrovnik, Croatia, 24–27 September 2002; pp. 41–46.
29. Bertsekas, D.P. Enlarging the region of convergence of Newton's method for constrained optimization. *J. Optim. Theory Appl.* **1982**, *36*, 221–252. [CrossRef]
30. Candès, E.J.; Li, X.; Ma, Y.; Wright, J. Robust Principal Component Analysis? *J. ACM* **2011**, *58*, 1–37. [CrossRef]
31. Lin, Z.; Chen, M.; Ma, Y. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv* **2010**, arXiv:1009.5055.
32. Candès, E.; Recht, B. Exact Matrix Completion via Convex Optimization. *Found. Comput. Math.* **2008**, *9*, 717. [CrossRef]
33. Maldague, X. *Theory and Practice of Infrared Technology for Nondestructive Testing*; Wiley: Hoboken, NJ, USA, 2001.
34. Maldague, X.P.V.; Moore, P.O. *Nondestructive Testing Handbook: Infrared and Thermal Testing*, 3rd ed.; American Society for Nondestructive Testing: Columbus, OH, USA, 2001.
35. Ibarra-Castanedo, C.; Genest, M.; Piau, J.M.; Guibert, S.; Bendada, A.; Maldague, X.P. Active infrared thermography techniques for the nondestructive testing of materials. In *Ultrasonic and Advanced Methods for Nondestructive Testing and Material Characterization*; World Scientific: Singapore, 2007; pp. 325–348. [CrossRef]
36. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37–52. [CrossRef]
37. Busse, G. Nondestructive evaluation of polymer materials. *Ndt Int.* **1994**, *27*, 253–262. [CrossRef]
38. Usamentiaga, R.; Ibarra-Castanedo, C.; Maldague, X. More than Fifty Shades of Grey: Quantitative Characterization of Defects and Interpretation Using SNR and CNR. *J. Nondestruct. Eval.* **2018**, *37*, 25. [CrossRef]
39. Jaccard, P. Lois de Distribution Florale dans la Zone Alpine. Bulletin de la Société vaudoise des sciences naturelles. Available online: https://www.e-periodica.ch/digbib/view?pid=bsv-002:1902:38::503#110 (accessed on 21 October 2021).
40. Wada, K. labelme: Image Polygonal Annotation with Python. Available online: https://github.com/wkentaro/labelme (accessed on 21 October 2021) .
41. Tomasi, C.; Manduchi, R. Bilateral filtering for gray and color images. In Proceedings of the Sixth International Conference on Computer Vision (ICCV), Bombay, India, 7 January 1998; pp. 839–846.

sensors

MDPI

Article

Corona Discharge Characteristics under Variable Frequency and Pressure Environments

Pau Bas-Calopa [1], Jordi-Roger Riba [1,*] and Manuel Moreno-Eguilaz [2]

[1] Electrical Engineering Department, Universitat Politècnica de Catalunya, 08222 Terrassa, Spain; pau.bas@upc.edu
[2] Electronics Engineering Department, Universitat Politècnica de Catalunya, 08222 Terrassa, Spain; manuel.moreno.eguilaz@upc.edu
* Correspondence: jordi.riba-ruiz@upc.edu; Tel.: +34-937-398-365

Abstract: More electric aircrafts (MEAs) are paving the path to all electric aircrafts (AEAs), which make a much more intensive use of electrical power than conventional aircrafts. Due to the strict weight requirements, both MEA and AEA systems require to increase the distribution voltage in order to limit the required electrical current. Under this paradigm new issues arise, in part due to the voltage rise and in part because of the harsh environments found in aircrafts systems, especially those related to low pressure and high-electric frequency operation. Increased voltage levels, high-operating frequencies, low-pressure environments and reduced distances between wires pose insulation systems at risk, so partial discharges (PDs) and electrical breakdown are more likely to occur. This paper performs an experimental analysis of the effect of low-pressure environments and high-operating frequencies on the visual corona voltage, since corona discharges occurrence is directly related to arc tracking and insulation degradation in wiring systems. To this end, a rod-to-plane electrode configuration is tested in the 20–100 kPa and 50–1000 Hz ranges, these ranges cover most aircraft applications, so that the corona extinction voltage is experimentally determined by using a low-cost high-resolution CMOS imaging sensor which is sensitive to the visible and near ultraviolet (UV) spectra. The imaging sensor locates the discharge points and the intensity of the discharge, offering simplicity and low-cost measurements with high sensitivity. Moreover, to assess the performance of such sensor, the discharges are also acquired by analyzing the leakage current using an inexpensive resistor and a fast oscilloscope. The experimental data presented in this paper can be useful in designing insulation systems for MEA and AEA applications.

Keywords: more electric aircraft; electrical discharges; visual corona; corona extinction voltage; variable frequency; low pressure

Citation: Bas-Calopa, P.; Riba, J.-R.; Moreno-Eguilaz, M. Corona Discharge Characteristics under Variable Frequency and Pressure Environments. *Sensors* **2021**, *21*, 6676. https://doi.org/10.3390/s21196676

Academic Editor: Manuel José Cabral dos Santos Reis

Received: 15 September 2021
Accepted: 5 October 2021
Published: 8 October 2021

1. Introduction

More electric aircrafts (MEAs) allow for reducing weight [1], fuel consumption, greenhouse gas emissions, operation and maintenance costs and boosting overall system efficiency when compared with conventional aircrafts [2]. However, engineers are facing important challenges due to the increased voltage levels MEAs require, the increase in the power density and the dv/dt, or the reduction in distances between electrical wires, thus increase the likelihood of electric arc occurrence [3,4] with the consequent safety risks.

Jet aircrafts typically fly at altitudes between 33,000 and 42,000 feet (10,000 m to 12,800 m) [5], thus operating under harsh environmental conditions. Some electric and electronic aircraft systems operate in unpressurized zones [6], so electric and electronic aircraft systems must be designed to operate under a broad range of pressures, in the range 1 atm to 0.15 atm [7].

The development of MEA and AEA systems is accompanied by a rise of the distribution voltage levels, since for a given power, the lower the current, the higher the voltage, and vice versa. However, according to Paschen's law, when operating at higher

voltage levels and reduced pressure, there is risk of partial discharges (PDs) and electric breakdown [8], the inception voltages of such discharges being below the ones found at sea level [9,10].

Direct current (dc) distribution systems of current aircrafts are operated at 28 V, 270 V (±135 V) or 540 V (±270 V), whereas alternating current (ac) distribution systems are operated at 230 V or 115 V phase voltage with variable or wide frequency (typically 320–800 Hz) [11], or 230 V or 115 V phase voltage with constant frequency (400 Hz) [12,13]. In AEA, voltage levels in the range 2 to 3 kV seem advantageous [13]. It is believed that ac distribution systems in the voltage range between 0.6 and 2 kV lead to wiring systems with less weight, reduced power losses and higher efficiency. However, above 2 kV, additional insulation requirements add extra mass to the system, thus needing careful analysis [13]. Because of the need of more electrical power, next generations MEA aircrafts will probably raise the distribution voltage above 1 kV [14,15]. According to NASA, future aerospace systems can operate at voltages up to 20 kV (designed for 40 kV), with high-frequency operation (400 Hz to 4000 Hz) [16]. The combination of low pressure, high voltage and high-operating frequencies stresses insulation systems [13], with the consequent degradation risk due to partial discharge and arc tracking occurrence [17,18] because electrical discharge inception voltages can be much lower than those at sea level [1].

Wiring issues in aircrafts due to electrical discharges and arc tracking leading to insulation degradation have caused catastrophic accidents [13]. Different insulation materials have been proposed to combat insulation degradation [19,20]. This is of paramount importance because MEA and AEA make an increasing use of electric and electronic apparatus and devices, so polymer insulation materials are inevitably exposed to harsh and varying environments. Thus, care must be taken in selecting appropriate insulating materials since reliability is an issue [21]. Before electric breakdown occurrence, partial discharges (PDs) appear, PDs being discharges that do not entirely channel the insulation between two electrodes [22]. They are roughly classified as internal discharges, external discharges and corona discharges. Although short duration PDs are usually harmless, when they persist over time, they tend to generate important insulation damage in polymeric materials because PDs can produce a partially conductive path or track on the insulation outer surface, thus favoring the flow of an electric current and ultimately arc tracking activity or even complete electrical breakdown [23]. Arc tracking occurring in organic (polymeric) insulation systems, damages the polymer material, which shifts from insulating to conductor because of the tremendous thermal shocks due to the electron bombardment generated by the electrical discharge [24]. This effect also breaks the polymeric chains and degrades the insulation, generating conducting carbon tracks, which reduce the insulating properties of the polymer surface and promote electrical breakdown [25], fire hazard [26] and explosions [27], even at very low voltage [28]. Atmospheric pressure, applied voltage, supply frequency and geometry are dominant variables to determine corona discharge inception and extinction levels.

It is worth noting that reliability and safety are key points in aircraft systems. To design reliable aircraft insulation systems, it is necessary to have a deep knowledge of the conditions leading to a corona [6] as a function of environmental pressure and supply frequency, because if these conditions are not controlled, they can lead to damaging effects, including arc tracking and electrical breakdown [2]. To better understand the effect of low pressure and supply frequency on the development of electrical discharges, it is imperative to run extensive test plans. Due to the difficulty to operate under low-pressure environments and using high-voltage generators with adjustable frequency, there is a lack of experimental data obtained under conditions compatible with aeronautic environments. This paper aims to contribute in this field. In addition, some of the studies are focused to analyze the disruptive spark breakdown [29,30], but non-uniform gaps can lead to corona inception and extinction voltages much lower than those required to ignite disruptive or breakdown discharges.

To analyze the effect of low-pressure environments jointly with the effect of the supply frequency, a rod-to-plane electrode configuration is tested in the 20–100 kPa and 50–1000 Hz intervals, these ranges account for the wide range of pressure and frequencies found in aircraft applications.

The detection of partial discharges and arcing activity in aircrafts in the very early stage is a problem that remains unsolved, so there is an imperious need to develop sensor systems to solve this important safety problem. Although there are several sensors that potentially can be applied to detect electrical discharges such as PD detection, antennas to detect electromagnetic noise and radio interference voltage, or acoustic sensors, they are too complex or are severely affected by the noise found in aeronautic applications. In addition, these methods do not directly allow to locate the discharge points. Therefore, this paper focus on the visible-UV light emitted by the electrical discharges because this method offers immunity to noise, while allowing to locate the discharge points.

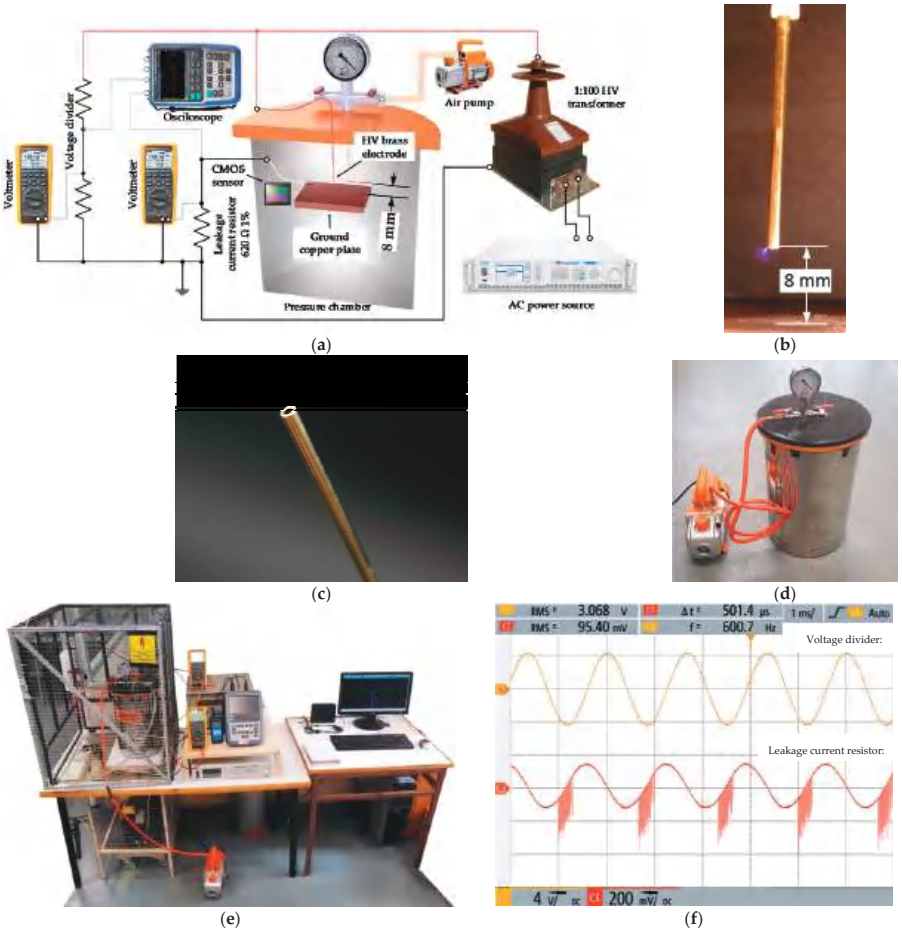
It is known that the corona effect generates visible (mainly blue) and ultraviolet (UV) light [31]. Thus, by using optical sensors sensitive to these spectral regions, it is possible to detect the corona discharges in the early stage [2]. A corona can also be detected by means of other methods, which are usually more complex, such as optical spectrophotometers [32], audible noise meters [33], PD and radio interference voltage (RIV) detectors [34] or UHF sensors [35]. However, the simpler and straightforward way to locate the discharge point is by using visible-UV imaging sensors. Therefore, to determine the conditions leading to a corona, the corona extinction voltage is determined by using a low-cost high-resolution CMOS imaging sensor. This sensor is sensitive to the visible and near ultraviolet spectral ranges, and the discharge points are identified from the images generated by the CMOS sensor, as well as the intensity of the discharge, thus offering high sensitivity, simplicity low-cost measurements and immunity to electromagnetic noise. Results attained with the imaging sensor are compared with those obtained by analyzing the leakage current. Experimental data presented in this paper can be useful to design insulation systems for future MEA and AEA applications, thus ensuring the reliability of aircraft insulation systems for electrical and electronic circuits.

Specific objectives of this research work include determining the combined effect of pressure and frequency on visual corona and specifically on corona extinction voltage (CEV) for aeronautics applications using a low-cost CMOS imaging sensor, and to compare the sensitivity of such sensor with that of a leakage current sensor.

The paper is organized as follows. Section 2 details the experimental setup to generate a variable frequency high voltage and the instrumentation used, as well as the sensors used to detect the corona extinction voltage. The experimental results are presented in Section 3, whereas Section 4 discusses the results attained. Finally, the conclusions of the paper are developed in Section 5.

2. Experimental Setup

Corona experiments were performed inside a pressurized chamber that allows reducing the pressure from 100% to 20% of the pressure at sea level, i.e., from 100 kPa to 20 kPa approximately, covering the altitude/pressure interval of commercial jet liners. The low-pressure chamber is composed of a stainless-steel cylindrical container (diameter = 130 mm, height = 375 mm) with a sealed methacrylate lid to allow the wireless imaging sensor to transmit the long-exposure photographs to a computer placed outside the low-pressure chamber, as displayed in Figure 1c. The pressure is regulated using a vacuum pump (1/4 HP, 0.085 m^3/min, Bacoeng BA-1, Bacoeng, Suzhou, China) and a manometer (76 mmHg, $\pm$2.5%, Bacoeng, Suzhou, China). Experiments were conducted at a constant room temperature of 25 °C. The humidity effect was not studied but limited to below 25% during the experiments.



(**a**) (**b**) (**c**) (**d**) (**e**) (**f**)

Figure 1. Experimental setup. (**a**) Sketch of the experimental setup and the instrumentation used in the high-voltage tests at variable pressure, frequency and voltage; (**b**) photograph of the rod-to-plane electrode used in the experiments; (**c**) detail of the tip of the electrode; (**d**) low-pressure chamber; (**e**) photograph of the experimental setup including the low-pressure chamber; (**f**) snapshot of the oscilloscope used to detect corona activity connected to the terminals of the leakage current resistor.

The applied voltage and supply frequency were regulated by means of a SP300VAC600W programmable ac source (600 W, 0–300 V, ±0.1 V, 15–1000 Hz, APM Technologies, Dongguan, China) following the IEC61000-4-14 standard. A high-voltage instrument transformer (single-phase, turns ratio 1:100, maximum voltage 36 kV, 600 VA, VKPE-36, Laboratorio Electrotécnico, Cornellà de Llobregat, Spain) was connected to the output of the SP300VAC600W programmable ac source to step up the output voltage provided by this source.

A voltage divider with a voltage ratio of 1000:1 was used to measure the high voltage at the output of the high-voltage transformer, so that the load voltage was measured with a calibrated true-RMS voltmeter (0–1000 V_{RMS}, 0.4%, 0–10 A, Fluke 289, Fluke, Everett, WA, USA).

The rod-to-plane gap is composed of a MBT5M brass tube (Albion Alloys, Poole, UK) with outer and inner diameters of Ø = 1.5 mm and Ø = 0.8 mm, respectively. The tip of the electrode was placed at a height of 8 mm above a grounded flat copper plane. A rod-to-plane arrangement was used in this work because this geometry is among the reference gaps used in high-voltage applications [36], thus allowing the generation of

PDs. The tip was cut with a hacksaw for metals and polished with a metal grinding wheel (fine grain 220 g, 2800 rev/min). This geometry was chosen in order to generate a corona before arc appearance under the conditions analyzed in this work (20–100 kPa, 50–1000 Hz, 25 °C, humidity < 25%, <6 kV) being compatible with the dimensions of the low-pressure chamber.

The experimental corona extinction voltage (CEV) values shown and analyzed in Section 3 are measured by the means of two detection methods. The first method is based on visual corona tests, a corona representing a pre-arc condition in its very early stage before obvious damage in the insulation can be appreciated. To detect the visual corona phenomenon and locate the discharge area, a high-resolution low-cost back-illuminated CMOS imaging sensor (sensor size 8.0 mm, cell size 0.8 μm × 0.8 μm, 8000 × 6000 pixels, 48 Mpixels, 30 frames/second, lens focal 17.9 mm, quad Bayer filter array, images in raw format, IMX586, Sony, Tokyo, Japan) was used, because back-illuminated CMOS sensors are sensitive to visible and UV light [37]. To increase the sensitivity of the measurements, long-exposure pictures were taken for 32s in manual focus mode, selecting an ISO of 400. This is a low-cost sensor that allows for locating corona discharge regions, as well as quantifying the intensity of the discharges, thus easing maintenance tasks. This sensor also enables reducing the costs and complexity of the instrumentation while offering excellent measurement sensitivity and accuracy. Due to a special arrangement of the photodiodes, back-illuminated CMOS imaging sensors allow capturing more light compared with conventional CMOS sensors, thus performing better under low-light conditions, particularly in the UV spectrum [37,38].

To determine the existence of a corona in the images taken by the CMOS sensor, they were first converted to grayscale (rgb2gray function in Matlab®). Next, the mean value of the pixels of a selected window centered near the corona focus was calculated and compared with the mean value of the pixels from the rest of the image. If the first value is greater than the second by 5%, it is assumed that there is a corona. This simple processing approach is quite immune to the effect of external light (partial darkness).

The second method is based on measuring the leakage current. In this case the sensing system consists of a 620 Ω ± 1% low-inductance resistor connected in series between the ground copper plate and the laboratory electrical ground. The leakage current from the discharges produces a voltage drop across the resistor that was monitored and registered with a fast digital insulated oscilloscope (5GSa/s, 0–1000 V, 0.5% + 0.05% voltage range, RTH1004, Rohde & Schwarz, Munich, Germany) equipped with two RT-ZI10 passive voltage probes (500MHz, 1kV, 10:1, R&S ®, R&S, Munich, Germany).

Corona appearances in the leakage current is seen as peaks superimposed in the current waveform. Therefore, by using a peak detection algorithm (based on the findpeaks function of Matlab®) it is easy to differentiate between corona and no corona conditions.

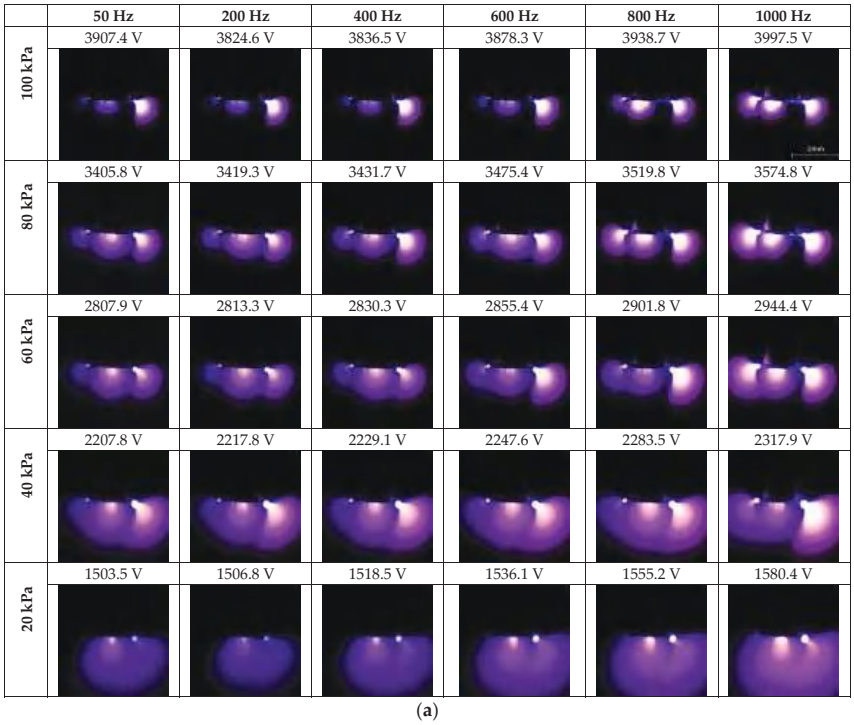
3. Experimental Results

This section details the experimental results obtained by using the setup and instrumentation detailed in Section 2.

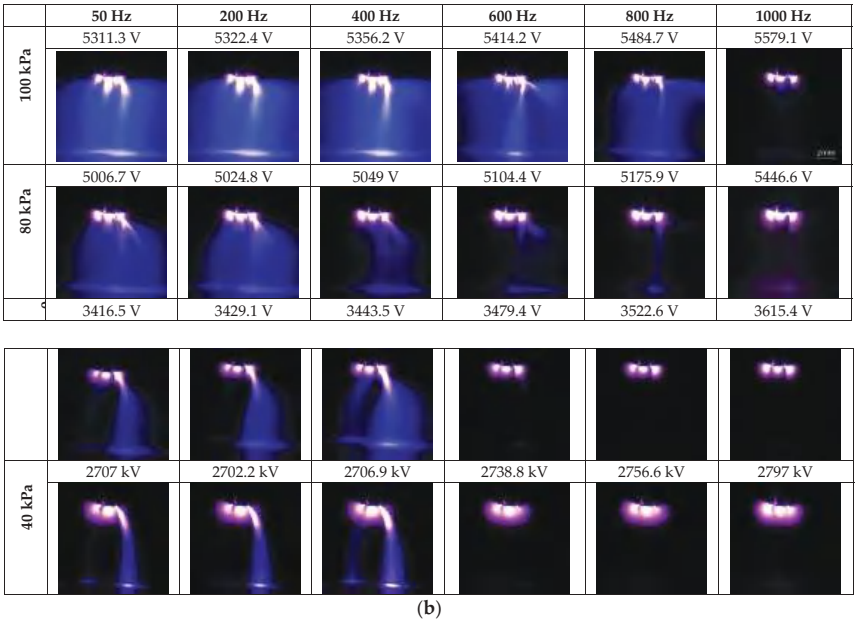
3.1. Visual Corona Photographs Taken with the Back-Illuminated CMOS Sensor

In order to describe the effects of frequency and pressure on corona discharges, long-exposure photographs (32 s exposition time, RGB mode, ISO 400, manual focus, automatic white balance) were taken using the setup detailed in Section 2. The discharges were performed in the 20–100 kPa range in increments of 20 kPa and for different frequencies (50 Hz, 200 Hz, 400 Hz, 600 Hz, 800 Hz and 1000 Hz). Some of the long-exposure photographs are shown in Figure 2, which show the effects of pressure and frequency on the visual corona discharges. It is noted that the voltage levels corresponding to the photographs in Figure 2 are higher than the CEV values to facilitate a good description of the discharge patterns. It is noted that at low pressure, specifically around 20 kPa, care must be taken

when increasing the voltage level, because there is very little difference between the CEV value and the voltage level at which complete breakdown occurs.



(**a**)



(**b**)

Figure 2. Long-exposure images taken with the back-illuminated CMOS sensor. (**a**) Negative ac corona discharges. (**b**) Positive ac streamer corona discharges.

Figure 2a shows negative corona discharges before streamers of positive corona appear. The major visual effect is due to the pressure change. At high pressures, corona discharges appear in several spots or "beads" and the active region of ionization is relatively small and well-defined. As pressure decreases, the active region slightly expands and becomes more diffuse, while the number of corona spots reduces. Figure 2a also shows that the supply frequency has very little visual effect on the distribution of the corona discharges.

Figure 2b shows positive corona discharges superimposed with negative discharges, the last ones appearing at lower voltages. According to the images included in this figure, the streamers become more localized and ultimately develop into fewer beams of light (650 μm ± 100 μm in diameter, measured from the images) as pressure decreases. Figure 2b also shows that the density of positive streamers also reduces when the supply frequency increases. It can also be observed that in some cases, for a given pressure, there is a maximum frequency from which streamers are not formed, and a further voltage increase may be followed by electrical breakdown.

3.2. Obtained Corona Extinction Voltages (CEV)

This section describes the experimental CEV results attained when analyzing rod-to-plane gap geometry, as described in Figure 1b. To obtain the CEV value, the voltage is progressively increased from 0 kV until identifying corona activity, this point corresponds to the corona inception voltage (CIV). Next, the voltage is increased by about 10% and slowly reduced until the corona effect extinguishes. The last point is where a corona manifest corresponds to the corona extinction voltage (CEV), i.e., the minimum voltage value where corona activity can be found.

Figure 3 summarizes the process to determine the CEV value. This process was repeated three times for each measurement, and these values were annotated.

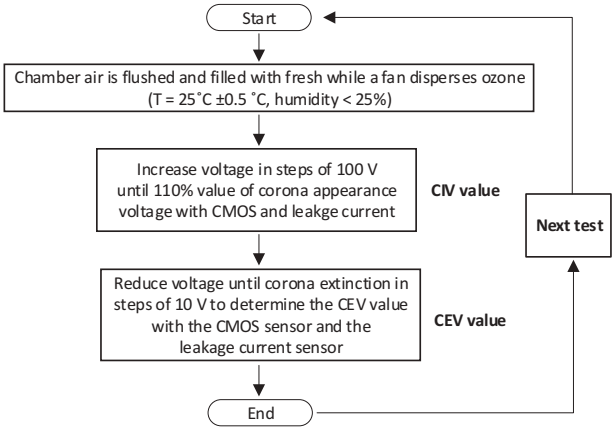


Figure 3. Procedure to determine the value of the corona extinction voltage (CEV).

To avoid the CEV value to be influenced by ozone formation, the atmospheric air in the low-pressure chamber was completely replaced in each test.

Table 1 summarizes the tests performed for each condition (pressure range 20–100 kPa, frequency range 50–1000 Hz).

Table 1. Tests performed (three consecutive repetitions each test).

Pressures	Frequencies
100 kPa	50, 200, 400, 600, 800, 1000 Hz
80 kPa	50, 200, 400, 600, 800, 1000 Hz
60 kPa	50, 200, 400, 600, 800, 1000 Hz
40 kPa	50, 200, 400, 600, 800, 1000 Hz
20 kPa	50, 200, 400, 600, 800, 1000 Hz

The voltage amplitude was increased with discrete steps of 100 V consisting of a ramp with a standard 1 V/ms rate. To determine the CEV value, the voltage was decreased with steps of 10 V at a rate of −1 V/ms.

Table 2 summarizes the CEV values obtained by using the CMOS imaging sensor according to the experimental setup shown in Figure 1b when analyzing different frequencies (50 Hz, 200 Hz, 400 Hz, 600 Hz, 800 Hz and 1000 Hz) and different pressures (100 kPa, 80 kPa, 60 kPa, 40 kPa and 20 kPa).

Table 2. Experimental results corresponding to the rod-to-plane electrode geometry. CEV versus environmental pressure and supply frequency.

Pressure	Test	Sensor	50 Hz	200 Hz	400 Hz	600 Hz	800 Hz	1000 Hz
100 kPa	Test 1	Camera	3761.5	3817.1	3784.0	3792.0	3659.9	3579.1
		Leakage current	3723.2	3788.1	3754.5	3792.0	3659.9	3579.1
	Test 2	Camera	3802.3	3781.0	3826.3	3899.8	3798.5	3732.3
		Leakage current	3774.4	3781.0	3826.3	3835.9	3798.5	3732.3
	Test 3	Camera	3800.7	3819.6	3807.1	3795.8	3843.2	3779.7
		Leakage current	3763.9	3800.5	3786.8	3774.7	3843.2	3779.7
	Average		3788.2	3805.9	3805.8	3829.2	3767.2	3697.0
80 kPa	Test 1	Camera	3382.9	3414.8	3410.6	3441.2	3341.8	3146.4
		Leakage current	3326.5	3377.3	3410.6	3430.5	3341.8	3146.4
	Test 2	Camera	3367.0	3398.1	3412.6	3424.6	3358.8	3272.4
		Leakage current	3348.2	3379.1	3373.0	3404.0	3358.8	3272.4
	Test 3	Camera	3385.4	3377.7	3412.9	3415.1	3358.1	3271.9
		Leakage current	3329.1	3359.3	3393.4	3404.2	3358.1	3271.9
	Average		3378.4	3396.9	3412.1	3427.0	3352.9	3230.2
60 kPa	Test 1	Camera	2757.9	2774.2	2761.2	2763.5	2791.7	2689.5
		Leakage current	2719.8	2736.0	2742.6	2742.2	2781.3	2688.1
	Test 2	Camera	2798.5	2842.9	2861.9	2868.7	2873.7	2813.1
		Leakage current	2778.4	2823.8	2842.3	2837.2	2851.0	2813.1
	Test 3	Camera	2853.7	2918.9	2901.8	2909.8	2916.8	2837.4
		Leakage current	2815.5	2851.7	2862.4	2889.6	2916.8	2837.4
	Average		2803.4	2845.3	2841.7	2847.3	2860.7	2780.0
40 kPa	Test 1	Camera	2255.6	2267.0	2270.0	2270.7	2274.6	2254.3
		Leakage current	2198.5	2248.1	2270.0	2249.5	2274.6	2254.3
	Test 2	Camera	2266.1	2315.4	2330.7	2333.9	2366.7	2354.2
		Leakage current	2227.7	2296.3	2330.7	2333.9	2366.7	2354.2
	Test 3	Camera	2246.4	2316.4	2321.5	2313.3	2322.5	2308.0
		Leakage current	2246.4	2297.3	2321.5	2313.3	2322.5	2308.0
	Average		2256.0	2299.6	2307.4	2306.0	2321.2	2305.5
20 kPa	Test 1	Camera	1468.8	1520.2	1522.6	1529.0	1526.7	1529.8
		Leakage current	1449.3	1520.2	1522.6	1529.0	1526.7	1529.8
	Test 2	Camera	1374.7	1443.1	1444.3	1426.4	1439.3	1458.0
		Leakage current	1355.2	1443.1	1444.3	1426.4	1439.3	1458.0
	Test 3	Camera	1450.2	1474.0	1463.6	1488.5	1552.0	1629.1
		Leakage current	1450.2	1474.0	1463.6	1488.5	1552.0	1629.1
	Average		1431.2	1479.1	1476.8	1481.3	1506.0	1539.0

Electrode diameter = 1.5 mm, tip angle = 90°, electrode tip to plane distance = 8 mm.

For a better analysis, the results presented in Table 2 are potted in Figure 4.

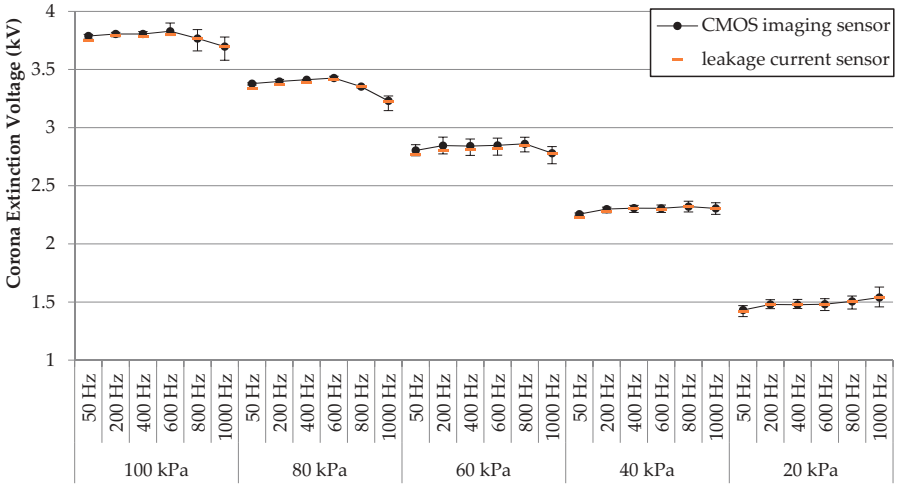


Figure 4. Experimental results of the rod-to-plane electrode geometry. CEV values (kV) versus pressure (kPa) and supply frequency (Hz). Results from the imaging sensor and the leakage current sensor (resistor) were plotted together.

Results in Figure 4 show that when analyzing the rod-to-plane electrode geometry, the effect of frequency in the range 50–1000 Hz is much less than the effect of pressure in the 100–20 kPa interval in the CEV values. Although high frequencies tend to reduce the CEV values in the 100–60 kPa range, this effect disappears at lower pressures.

To further analyze the effect of pressure, Figure 5 shows the CEV versus pressure error plots at each analyzed frequency. Such error plots show that the CEV reduces with pressure almost linearly. The parameters of the linear fits are summarized in Table 3.

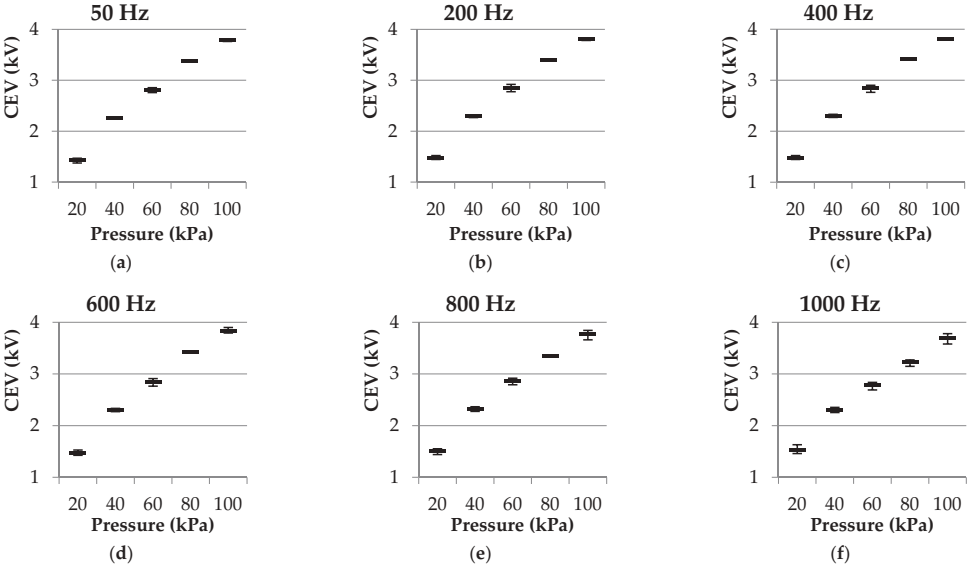


Figure 5. Experimental results of the rod-to-plane electrode geometry obtained with the CMOS imaging sensor. CEV values (kV) dispersion plot of the three measurements at each point versus pressure (kPa). (**a**) 50 Hz; (**b**) 200 Hz; (**c**) 400 Hz; (**d**) 600 Hz; (**e**) 800 Hz; (**f**) and 1000 Hz.

Table 3. Linear fit parameters $CEV = CEV_0 + m \cdot P$, where CEV_0 is the CEV at zero pressure in Volt, m is the slope in Volt/kPa and P is the pressure in kPa.

Frequency	CEV_0		m		R^2	
	Imaging Sensor	Leakage Current	Imaging Sensor	Leakage Current	Imaging Sensor	Leakage Current
50 Hz	980.5	965.9	29.182	28.908	0.9853	0.9867
200 Hz	1040.1	1031.2	28.755	28.565	0.9845	0.9871
400 Hz	1040.0	1043.4	28.813	28.548	0.9835	0.9839
600 Hz	1033.1	1037.5	29.084	28.765	0.9849	0.9847
800 Hz	1095.4	1093.2	27.771	27.771	0.9820	0.9827
1000 Hz	1138.1	1138.0	26.204	26.204	0.9863	0.9864

R^2 is the coefficient of determination of linear regression, indicating how well data fits.

Results summarized in Table 3 show a quasi-linear relationship of the CEV versus P plots measured at different frequencies in the 50–1000 Hz range, according to the high values of the determination coefficient R^2. These results also show similar values of the CEV_0 and m parameters for the different frequencies, thus corroborating the low effect of the frequency in the CEV value.

Table 4 compares the CEV values obtained with both sensors.

Table 4. Average difference between the CEV values attained with the CMOS image and the leakage current sensors for each frequency.

Frequency	Difference
50 Hz	1.153%
200 Hz	0.695%
400 Hz	0.388%
600 Hz	0.466%
800 Hz	0.078%
1000 Hz	0.003%

From the values shown in Table 4 it can be observed that both methods have similar sensitivity, whereas the difference between the results attained with the imaging sensor and those with the leakage current sensor decreases with frequency.

4. Discussion

The results presented in Figure 4 clearly show that CEV values are mainly affected by ambient pressure. The results plotted in Figure 4 are in accordance with previous studies analyzing gas discharges for specific supply frequencies [7]. This effect is due to the fact that the mean free path between ion collisions is inversely proportional to air density, and thus, a larger number of successful secondary ionizations are produced at a lower pressure, so that partial discharges can occur at lower voltages than the ones required at atmospheric pressure [39].

Regarding the effect of frequency, Linder and Steele [40] studied the effect of frequency on breakdown, proving that breakdown voltage decreases as the operating frequency increases. There are other studies describing that CIV values usually decrease when increasing the frequency, although this effect reduces at lower pressures [10]. This same effect was observed in the experimental results presented in this paper. However, as can be seen in Figure 4, the CEV values at 20 kPa slightly rise when increasing the frequency up to 1000 Hz.

The formation of a larger number of negative corona spots and brighter negative discharges at atmospheric pressure in contrast to what was observed at low pressure as shown in Figure 2a, can be attributed to the fact that at atmospheric pressure a higher voltage is needed to produce a corona; therefore, more spots are suitable for ionization and more molecules are ionized in the process, thus increasing the brightness of the

discharge [41]. The shape change observed in Figure 2a from a localized and defined to a more diffuse and homogenous corona when lowering pressure, may be due to the fact that at a low pressure, the free path of ionization is larger, so that ionized particles can travel further, thus increasing the active area of ionization. This visual effect of pressure on a corona has also been described in previous studies [42].

The results in Figure 2b show that when pressure reduces, the number of streamers also reduces, becoming less diffuse and more localized. This effect was described in [43] using a high-speed photographic camera. In the images presented in this paper, atmospheric-pressure streamers appear as a diffuse bluish glow within the gap due to the 32 s long-exposure effect.

To the best of our knowledge, there is a scarcity of publications analyzing in detail the combined effect of variable frequency and variable pressure on visual corona. However, it has been shown that although frequency has no significant visual effect on negative corona (see Figure 2a), there is a slight effect on the streamers of positive corona (see Figure 2b).

The sensibility to detect corona discharges of the image sensor has been tested and compared with that of the leakage current sensor, obtaining very close results, as shown in Figure 4 and in Table 4, where the percentage differences are calculated, which are very low. A similar comparison was performed in [7] where it was also concluded that the imaging method with a CMOS camera has almost the same sensitivity as other sensitive methods for corona detection.

It is noted that a drawback of the detection method based on the CMOS sensor is that it requires partial darkness to operate. However, partial darkness is often found in aeronautics applications since wires and harnesses are often inside troughs, ducts, or conduits whose interior is usually dark. The authors are aware of this drawback, so they are working in the integration of solar-blind imaging sensors, which can also operate under usual sunlight conditions.

5. Conclusions

This paper conducted an experimental study to determine the effect of pressure and frequency on visual corona using a CMOS imaging sensor and by measuring the leakage current, proving that both sensing systems present very similar sensitivity, although the imaging sensor allows locating the points where the electrical discharges occur. The study was conducted by analyzing a rod-to-plane air gap in the 20–100 kPa and 50–1000 Hz intervals, covering most aeronautic applications. The results show that pressure and frequency both have an effect on corona extinction voltage (CEV). CEV increases remarkably with air pressure, but the effect of frequency is lower, causing the CEV to decrease with frequency in the 100–60 kPa pressure range, this effect diminishes with pressure. In addition, a visual description of the effects of pressure and frequency on a corona was performed. The results presented show that the CMOS image sensor has enough sensitivity to be used as a corona detector in low-pressure environments and under a wide range of electrical frequencies. In addition, it was shown that although the difference between the CEV values found with the CMOS imaging sensor and by analyzing the leakage current is very low, this difference tends to reduce at higher frequencies.

Author Contributions: Conceptualization, J.-R.R. and P.B.-C.; methodology, J.-R.R.; validation, J.-R.R., P.B.-C. and M.M.-E.; formal analysis, M.M.-E.; investigation, J.-R.R. and P.B.-C.; writing—original draft preparation, J.-R.R.; writing—review and editing, M.M.-E. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by Ministerio de Ciencia e Innovación de España, grant number PID2020-114240RB-I00 and by the Generalitat de Catalunya, grant number 2017 SGR 967.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Rui, R.; Cotton, I. Impact of low pressure aerospace environment on machine winding insulation. In Proceedings of the 2010 IEEE International Symposium on Electrical Insulation, San Diego, CA, USA, 6–9 June 2010; pp. 1–5. [CrossRef]
2. Riba, J.-R.; Gómez-Pau, Á.; Moreno-Eguilaz, M. Experimental Study of Visual Corona under Aeronautic Pressure Conditions Using Low-Cost Imaging Sensors. *Sensors* **2020**, *20*, 411. [CrossRef]
3. Belijar, G.; Chanaud, G.; Hermette, L.; Risacher, A. *Study of Electric Arc Ignition, Behavior and Extinction in Aeronautical Environment, in Presence of FOD*; Saint-Exupéry: Toulouse, France, 2017; pp. 1–8.
4. Andrea, J.; Buffo, M.; Guillard, E.; Landfried, R.; Boukadoum, R.; Teste, P. Arcing fault in aircraft distribution network. In Proceedings of the Annual Holm Conference on Electrical Contacts, Denver, CO, USA, 10–13 September 2017; pp. 317–324. [CrossRef]
5. Hao, Z.; Wang, X.; Cao, X. Harmonic Control for Variable-Frequency Aviation Power System Based on Three-Level NPC Converter. *IEEE Access* **2020**, *8*, 132775–132785. [CrossRef]
6. Karady, G.G.; Sirkis, M.D.; Liu, L. Investigation of corona initiation voltage at reduced pressures. *IEEE Trans. Aerosp. Electron. Syst.* **1994**, *30*, 144–150. [CrossRef]
7. Riba, J.-R.; Gomez-Pau, A.; Moreno-Eguilaz, M. Sensor Comparison for Corona Discharge Detection Under Low Pressure Conditions. *IEEE Sens. J.* **2020**, *20*, 11698–11706. [CrossRef]
8. Capineri, L.; Dainelli, G.; Materassi, M.; Dunn, B.D. Partial discharge testing of solder fillets on PCBs in a partial vacuum: New experimental results. *IEEE Trans. Electron. Packag. Manuf.* **2003**, *26*, 294–304. [CrossRef]
9. Clean Sky. *9th Call for Proposals (CFP09)—List and Full Description of Topics*; Clean Sky: Brussels, Belgium, 2018; pp. 1–354.
10. Esfahani, A.N.; Shahabi, S.; Stone, G.; Kordi, B. Investigation of Corona Partial Discharge Characteristics Under Variable Frequency and Air Pressure. In Proceedings of the 2018 IEEE Electrical Insulation Conference (EIC), San Antonio, TX, USA, 17–20 June 2018; pp. 31–34. [CrossRef]
11. Kang, T.; Ryu, J. Determination of Aircraft Cruise Altitude with Minimum Fuel Consumption and Time-to-Climb: An Approach with Terminal Residual Analysis. *Mathematics* **2021**, *9*, 147. [CrossRef]
12. Wheeler, P.; Bozhko, S. The more electric aircraft: Technology and challenges. *IEEE Electrif. Mag.* **2014**, *2*, 6–12. [CrossRef]
13. Borghei, M.; Ghassemi, M. Insulation Materials and Systems for More and All-Electric Aircraft: A Review Identifying Challenges and Future Research Needs. *IEEE Trans. Transp. Electrif.* **2021**, *7*, 1930–1953. [CrossRef]
14. Mermigkas, A.C.; Clark, D.; Haddad, A.M. Investigation of High Altitude/Tropospheric Correction Factors for Electric Aircraft Applications. In *Lecture Notes in Electrical Engineering*; Springer: Cham, Switzerland, 2020; Volume 598, pp. 308–315. [CrossRef]
15. el Bayda, H.; Valensi, F.; Masquere, M.; Gleizes, A. Energy losses from an arc tracking in aeronautic cables in DC circuits. *IEEE Trans. Dielectr. Electr. Insul.* **2013**, *20*, 19–27. [CrossRef]
16. Woodworth, A.A.; Shin, E.E.; Lizcano, M. *High Voltage Insulation for Electrified Aircraft*; NASA: Washington, DC, USA, 2018; p. 11.
17. Shahsavarian, T.; Li, C.; Baferani, M.A.; Cao, Y. Surface discharge studies of insulation materials in aviation power system under DC voltage. In Proceedings of the 2020 IEEE Conference on Electrical Insulation and Dielectric Phenomena (CEIDP), East Rutherford, NJ, USA, 18–30 October 2020; pp. 271–274. [CrossRef]
18. Borghei, M.; Ghassemi, M. Characterization of Partial Discharge Activities in WBG Power Converters under Low-Pressure Condition. *Energies* **2021**, *14*, 5394. [CrossRef]
19. Dricot, F.; Reher, H.J. Survey of Arc Tracking on Aerospace Cables and Wires. *IEEE Trans. Dielectr. Electr. Insul.* **1994**, *1*, 896–903. [CrossRef]
20. Fabiani, D.; Montanari, G.C.; Cavallini, A.; Saccani, A.; Toselli, M. Nanostructured-coated XLPE showing improved electrical properties: Partial discharge resistance and space charge accumulation. In Proceedings of the 2011 International Symposium on Electrical Insulating Materials, Kyoto, Japan, 6–10 September 2011; pp. 16–19. [CrossRef]
21. Du, B.; Liu, H. Effects of atmospheric pressure on tracking failure of gamma-ray irradiated polymer insulating materials. *IEEE Trans. Dielectr. Electr. Insul.* **2010**, *17*, 541–547. [CrossRef]
22. IEEE. *The Authoritative Dictionary of IEEE Standards Terms*, 7th ed.; IEEE Std 100-2000; IEEE Press: Piscataway Township, NJ, USA, 2000; pp. 1–1362. [CrossRef]
23. Cella, B. *On-line Partial Discharges Detection in Conversion Systems Used in Aeronautics*; Université de Toulouse: Toulouse, France, 2015.
24. Du, B.X.; Liu, Y.; Liu, H.J. Effects of low pressure on tracking failure of printed circuit boards. *IEEE Trans. Dielectr. Electr. Insul.* **2008**, *15*, 1379–1384. [CrossRef]
25. Douar, M.A.; Beroual, A.; Souche, X. Assessment of the resistance to tracking of polymers in clean and salt fogs due to flashover arcs and partial discharges degrading conditions on one insulator model. *IET Gener. Transm. Distrib.* **2016**, *10*, 986–994. [CrossRef]
26. Meng, D.; Zhang, B.Y.; Chen, J.; Lee, S.C.; Lim, J.Y. Tracking and erosion properties evaluation of polymeric insulating materials. In Proceedings of the ICHVE 2016—2016 IEEE International Conference on High Voltage Engineering and Application, Chengdu, China, 19–22 September 2016. [CrossRef]
27. Degardin, V.; Kone, L.; Valensi, F.; Laly, P.; Lienard, M.; Degauque, P. Characterization of the High-Frequency Conducted Electromagnetic Noise Generated by an Arc Tracking between DC wires. *IEEE Trans. Electromagn. Compat.* **2016**, *58*, 1228–1235. [CrossRef]
28. Babrauskas, V. Research on Electrical Fires: The State of the Art. *Fire Saf. Sci.* **2008**, *9*, 3–18. [CrossRef]

29. Yu, B.; Kang, X.L.; Zhao, Q. An algorithm for gas breakdown voltage prediction in low pressure gap. In Proceedings of the 3rd International Conference on Computer Science and Application Engineering, Sanya, China, 22–24 October 2019. [CrossRef]
30. Owaid, A.; Owaid, A.Y. The Effect of Axial Magnetic Field on The Breakdown Voltage of Air at Low Pressure. *Iraqi J. Sci.* **2020**, *61*, 3228–3234. [CrossRef]
31. Riba, J.-R.; Abomailek, C.; Casals-Torrens, P.; Capelli, F. Simplification and cost reduction of visual corona tests. *IET Gener. Transm. Distrib.* **2018**, *12*, 834–841. [CrossRef]
32. Kozioł, M.; Nagi, Ł.; Kunicki, M.; Urbaniec, I. Radiation in the Optical and UHF Range Emitted by Partial Discharges. *Energies* **2019**, *12*, 4334. [CrossRef]
33. Chen, L.; MacAlpine, J.M.K.; Bian, X.; Wang, L.; Guan, Z. Comparison of methods for determining corona inception voltages of transmission line conductors. *J. Electrostat.* **2013**, *71*, 269–275. [CrossRef]
34. Souza, A.L.; Lopes, I.J.S. Experimental investigation of corona onset in contaminated polymer surfaces. *IEEE Trans. Dielectr. Electr. Insul.* **2015**, *22*, 1321–1331. [CrossRef]
35. Chai, H.; Phung, B.T.; Mitchell, S. Application of UHF Sensors in Power System Equipment for Partial Discharge Detection: A Review. *Sensors* **2019**, *19*, 1029. [CrossRef]
36. He, Z.; Zhu, J.; Zhu, J.; Bian, X.; Shen, B. Experiments and analysis of corona inception voltage under combined AC-DC voltages at various air pressure and humidity in rod to plane electrodes. *CSEE J. Power Energy Syst.* **2021**, *7*, 875–888. [CrossRef]
37. Turner, J.; Igoe, D.; Parisi, A.V.; McGonigle, A.J.; Amar, A.; Wainwright, L. A review on the ability of smartphones to detect ultraviolet (UV) radiation and their potential to be used in UV research and for public education purposes. *Sci. Total Environ.* **2020**, *706*, 135873. [CrossRef]
38. Wilkes, T.; McGonigle, A.J.; Pering, T.D.; Taggart, A.J.; White, B.S.; Bryant, R.G.; Willmott, J.R. Ultraviolet Imaging with Low Cost Smartphone Sensors: Development and Application of a Raspberry Pi-Based UV Camera. *Sensors* **2016**, *16*, 1649. [CrossRef] [PubMed]
39. Naidu, M.S.; Kamaraju, V. *High-Voltage Engineering*, 6th ed.; McGraw-Hill Education: New York, NY, USA, 2020.
40. Linder, W.; Steele, H. Estimating voltage breakdown performance of high-altitude antennas. In Proceedings of the WESCON/59 Conference Record, San Francisco, CA, USA, 18–21 August 1959; Volume 3, pp. 9–16. [CrossRef]
41. Riba, J.-R.; Gómez-Pau, Á.; Moreno-Eguilaz, M. Insulation Failure Quantification Based on the Energy of Digital Images Using Low-Cost Imaging Sensors. *Sensors* **2020**, *24*, 7219. [CrossRef] [PubMed]
42. Lewis, T.G.; Karady, G.G.; Sirkis, M.D. *An Analysis of the Frequency Characteristics of Corona Discharge at Low Pressure*; Phillips Laboratory, Kirtland Air ForceBase: Albuquerque, NM, USA, 1991.
43. Briels, T.M.P.; van Veldhuizen, E.M.; Ebert, U. Positive streamers in air and nitrogen of varying density: Experiments on similarity laws. *J. Phys. D Appl. Phys.* **2008**, *41*, 234008. [CrossRef]

sensors

MDPI

Article

Effective Connectivity for Decoding Electroencephalographic Motor Imagery Using a Probabilistic Neural Network

Muhammad Ahsan Awais [1,*], Mohd Zuki Yusoff [1], Danish M. Khan [1,2], Norashikin Yahya [1], Nidal Kamel [1] and Mansoor Ebrahim [3]

[1] Centre for Intelligent Signal & Imaging Research (CISIR), Electrical & Electronic Engineering Department, Universiti Teknologi PETRONAS, Seri Iskandar 32610, Perak, Malaysia; mzuki_yusoff@utp.edu.my (M.Z.Y.); danish_mkhan@yahoo.com (D.M.K.); norashikin_yahya@utp.edu.my (N.Y.); nidalkamel2@hotmail.com (N.K.)

[2] Department of Telecommunications Engineering, NED University of Engineering and Technology, Karachi 75270, Pakistan

[3] Faculty of Engineering, Sciences, and Technology, Iqra University, Karachi 75500, Pakistan; mebrahim@iqra.edu.pk

* Correspondence: Muhammad_18001588@utp.edu.my

Abstract: Motor imagery (MI)-based brain–computer interfaces have gained much attention in the last few years. They provide the ability to control external devices, such as prosthetic arms and wheelchairs, by using brain activities. Several researchers have reported the inter-communication of multiple brain regions during motor tasks, thus making it difficult to isolate one or two brain regions in which motor activities take place. Therefore, a deeper understanding of the brain's neural patterns is important for BCI in order to provide more useful and insightful features. Thus, brain connectivity provides a promising approach to solving the stated shortcomings by considering inter-channel/region relationships during motor imagination. This study used effective connectivity in the brain in terms of the partial directed coherence (PDC) and directed transfer function (DTF) as intensively unconventional feature sets for motor imagery (MI) classification. MANOVA-based analysis was performed to identify statistically significant connectivity pairs. Furthermore, the study sought to predict MI patterns by using four classification algorithms—an SVM, KNN, decision tree, and probabilistic neural network. The study provides a comparative analysis of all of the classification methods using two-class MI data extracted from the PhysioNet EEG database. The proposed techniques based on a probabilistic neural network (PNN) as a classifier and PDC as a feature set outperformed the other classification and feature extraction techniques with a superior classification accuracy and a lower error rate. The research findings indicate that when the PDC was used as a feature set, the PNN attained the greatest overall average accuracy of 98.65%, whereas the same classifier was used to attain the greatest accuracy of 82.81% with the DTF. This study validates the activation of multiple brain regions during a motor task by achieving better classification outcomes through brain connectivity as compared to conventional features. Since the PDC outperformed the DTF as a feature set with its superior classification accuracy and low error rate, it has great potential for application in MI-based brain–computer interfaces.

Keywords: brain–computer interface; brain effective connectivity; PDC; DTF; PhysioNet motor imagery; probabilistic neural network; SVM; KNN; decision tree

Citation: Awais, M.A.; Yusoff, M.Z.; Khan, D.M.; Yahya, N.; Kamel, N.; Ebrahim, M. Effective Connectivity for Decoding Electroencephalographic Motor Imagery Using a Probabilistic Neural Network. *Sensors* **2021**, *21*, 6570. https://doi.org/10.3390/s21196570

Academic Editor: Manuel José Cabral dos Santos Reis

Received: 28 August 2021
Accepted: 24 September 2021
Published: 30 September 2021

1. Introduction

Many unfortunate people with severe motor disabilities are not able to communicate well with the outside world. Their disabilities become obstacles between them and their social lives. Millions of people around the globe are affected by these types of disabilities, which are caused by several medical conditions, such as trauma, stroke, and different neurodegenerative diseases, including Alzheimer's disease (AD), Parkinson's disease

(PD), motor neuron diseases (MND), etc. Such people are afraid to be neglected as a significant part of society. Research communities around the world are working on the development of brain–computer interfaces based on different medical applications, such as brain-controlled wheelchairs, thus helping patients who have lost their abilities to communicate and providing them with mobility.

The brain–computer interface (BCI) has been one of the most rapidly growing technologies in recent years. It provides a control system that is capable of transforming a user's intentions into special commands to be used as a communication bridge between the brain and the outside world [1,2]. BCIs are a combination of software and hardware technologies that allow the brain to control external devices, such as prosthetic arms/legs or wheelchairs, by decoding different brain patterns. A general representation of a BCI system is given in Figure 1.

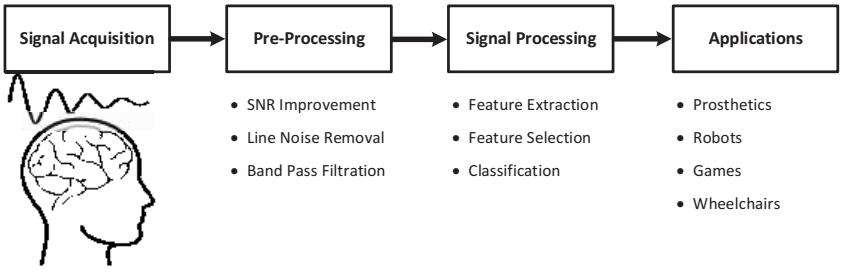


Figure 1. General overview of brain–computer interface (BCI) systems.

Electroencephalography (EEG) is the conventional method of monitoring the electrical activities of the brain by placing special sensors called electrodes on the surface of the scalp [3]. Electrical signals are generated by the inter-communication of cells within the brain. EEG-based BCIs identify explicit frequency patterns in the brain by sensing slight variations in the voltages that the brain emits while the person thinks in any way.

Motor imagery (MI) a traditional and active BCI paradigm that uses electroencephalography (EEG) to directly reflect a user's intention. Motor imagery can be expressed as the process of performing the imagination of motor tasks (i.e., the movement of body parts) without actually executing them physically [4].

Although the literature shows positive outcomes and accomplishments by using conventional MI-based BCI systems, including different state-of-the-art feature extraction and classification techniques, there are still many barriers and hurdles in using the technology efficiently and effectively. The major drawback of the existing MI-based BCIs is that they are based on traditional feature extraction and classification algorithms. Traditional feature extraction methods use MI-responsive frequency bands that do not have inter-subject or intra-subject consistency, which creates instability in BCI systems [5]. ERS/ERD analysis has proven to be complex due to its occurrence in different parts of the brain, during different time intervals, and at different frequencies, thus making it difficult to obtain significant features for classification [6]. Considering the low amplitude and noisy nature of EEG data, pattern inconsistency among multiple subjects and even altered patterns within a session with the same subjects can be expected. Various EEG studies confirmed the occurrence of MI-actuated signals in primary sensorimotor areas [7,8], whereas other researchers also reported the inter-communication of multiple brain parts during cognitive tasks [9,10], thus making it difficult to isolate one or two regions where the activity takes place. Furthermore, conventional MI-based BCI systems utilize temporal–spectral features from individual channels to recognize motor imagery patterns, which may not provide sufficient information. Therefore, a deeper understanding of the behavior of the brain's neural patterns is important in order to provide more useful and insightful features for BCIs, since the execution of motor or cognitive tasks results in the exchange of information

of multiple mutually interconnected brain regions. Thus, awareness of brain connectivity has become a key aspect of neuroscience and of understanding the behaviors of different regions. Different MI tasks are expected to have associations with particular brain connectivity patterns among the brain regions. Therefore, brain connectivity provides a promising approach to solving the stated shortcomings by considering inter-channel/region relationships during motor imagination. EEG recordings can be used to identify these connectivity patterns and offer unique features to infer a subject's intentions.

Several conventional classification algorithms have been widely used in brain-connectivity-based BCIs. In a study by Mehdi et al. [11], MVAR (multivariate autoregressive)-based source localization was used with an SVM for the classification of MI tasks, whereas Liang et al. [12] used the combination of the PDC and MEMD with an SVM to classify two-class MI. In another work [13], Panche et al. used a linear discriminant analysis (LDA) classifier for the prediction of MI tasks using the transfer-entropy-based effective connectivity. Lingyun et al. [14] used brain network analysis for the classification of lower-limb motor imagery; the authors used a sparse multinomial logistic regression (SMLR)-based SVM for the prediction of the MI. In the work of Rahman et al. [15], an fNIRS-based BCI was proposed by using effective temporal window estimation. The extracted features were used with three classifiers—LDA, SVM, and KNN—for the prediction of two-class MI.

Human brain mapping has primarily been used to construct maps that indicate regions of the brain that are activated by certain tasks. The term brain network or brain connectivity refers to sets of interconnected brain regions among which information is transferred. However, there has been insufficient discussion about the use of brain connectivity for motor-imagery-based pattern recognition. Our research is based on the analysis of effective connectivity, which explains the effects of neurons on each other, thus representing the causal connections between activated brain regions. The technique proposed in this study is based on brain connectivity, which is a unique direction for research on building a more accurate framework. The accuracy and performance improvements of the developed system in this work will be a step forward for the effective implementation of BCI applications, such as brain-controlled wheelchairs.

The rest of this paper is structured as follows: Section 2 describes the use of brain connectivity in the field of motor-imagery-based brain–computer interfaces. Subsequently, Section 3 covers the detailed description of an MI-based EEG dataset. Section 4 discusses the methods used in this work, including the preprocessing, feature extraction, and classification techniques. It further describes the detailed estimation of the PDC and DTF, along with the description of the evaluation measures. Section 5 presents the results and discussion. Finally, Section 6 concludes the paper.

2. Related Work

The study of brain connectivity is based on three distinct but related types of connectivity, including anatomical connectivity (AC), functional connectivity (FC), and effective connectivity (EC) [16,17]. Connectivity patterns are created by structural connections, such as synapses or fiber pathways, or they exemplify statistical or causal relationships, which are measured as cross-correlations, coherence, or information flow [18,19]. Among the different types of feature representations for motor-imagery-based EEG decoding, the connectivity models of multi-channel signals may produce more discriminating features for significant classification [20,21]. Several approaches to analyzing motor-imagery-based BCI systems that were established using brain connectivity have been proposed in the last few years.

Billinger et al. [22] proposed a technique for obtaining single-trial directed transfer functions (DTFs) by using vector-autoregressive (VAR) independent variable models for MI-based BCI classification, and the classification findings were identical to the band-power (BP) characteristics. Ming et al. [23] researched EEG characteristics associated with the movement of the left and right fingers. The event-related desynchronization (ERD) and movement-related cortical potential (MRCP) features were recovered using common

spatial patterns (CSP) and discriminative canonical pattern matching (DCPM). Since pre-movements have supportive MRCP and ERD characteristics, the proposed DCPM and CSP combination approach may be able to recognize them effectively. Mehdi et al. [11] proposed a method in which they used an MVAR model with the source localization algorithm (sLORETA) to extract active sources. After incorporating ANOVA for the reduction of the feature set, the authors used an SVM for the classification of the motor imagery tasks. In another study [24], a new time- and frequency-based causality was proposed by using a time-invariant BVAR model to investigate the flow of causality in the central region of the brain. As a result, improved performance with the new causality (NC) was reported as compared to the Granger causality.

In Rathee et al.'s study [25], the time-domain partial Granger causality (PGC) in terms of the connectivity feature set was used in an MI-based BCI environment. This resulted in the improved discriminability of MI tasks by using a single-trial effective connectivity distribution. However, Yang et al. [20] proposed a method in which time- and frequency-conditional Granger causality (CGC) was determined using a regularized orthogonal forward regression (ROFR) algorithm. The extracted features were classified using a boosted convolutional network, resulting in enhanced classification accuracy. Ahmad et al. [26] proposed an effective connectivity analysis for MI by using several variants, including DTF, direct DTF, and generalized PDC. A hierarchical feature selection technique was adopted to select the most important connectivity features, which resulted in successful discriminations of mental arithmetic tasks. Short-term DTF was used to investigate brain activities by implementing it to evaluate motor imagery experiments in three channels—C3, C4, and Cz [27,28]. Liang et al. [12] explored the effective connectivity in the motor cortex by using a combination of the PDC and multivariate empirical mode decomposition (MEMD). The results demonstrated the existence of significant effective connectivity in the bilateral hemisphere during the MI tasks.

However, in a study by Chung et al. [29], the temporal patterns of connectivity among EEG channels were evaluated according to the time-varying patterns of the channel-to-channel correlation coefficients and the average correlation coefficients per channel for left- and right-hand motor imagery. In [30], Lee et al. predicted the MI performance by using dynamic causal modeling to study the connectivity of a rest-state network, which affected the performance of the MI. As a result, a significant difference was observed in the network strength from the motor cortex to the right prefrontal cortex between the high- and low-MI-performance groups. Independent-source-based causal brain connectivity was introduced in [31] for the classification of left- and right-hand motor imagery. Chen et al. [32] used Granger causality analysis to analyze the brain connectivity between the motor, contralateral premotor, and sensorimotor areas. The results also revealed the significant difference in the G-causality trial numbers of left- and right-finger motor imagery. Li et al. [33] investigated effective connectivity in order to analyze and compare the rest state with right-hand motor imagination. In another study [13], Panche et al. proposed Renyi-based transfer entropy for measuring effective connectivity, resulting in significant robustness against varying amounts of data and noise levels.

3. Dataset Description

3.1. Ethical Approval

The dataset used in this work was entirely de-identified; therefore, no ethical review board (ERB) permission was required. The publicly accessible Physionet EEG motor imagery dataset used in this research work is available online [34], and it can be used without any further authorization.

3.2. Dataset

The developers of the BCI2000 instrumentation system created the dataset used in this work. The database includes more than 1500 one- and two-minute recordings from 109

volunteer subjects. The overall dataset contains the recordings of both the actual and the imagined motor tasks from all of the participants [35].

Each subject underwent 14 runs, where they performed several tasks, including the actual movement of the right/left fist, the imagination of the movement of the right/left fist, the actual movement of both fists/both feet, the imagination of the movement of both fists/both feet, and the opening/closing the eyes. EEG recordings were obtained as per the international 10–10 system from 64 electrode channels (excluding Nz, F9, F10, FT9, FT10, A1, A2, TP9, TP10, P9, and P10) that were sampled at the rate of 160 samples per second.

For the purpose of analysis, 14 out of the 64 EEG electrodes from all of the major regions of the brain (left region, right region, and central region) were used in this study. Among these 14 carefully chosen channels, two electrodes (i.e., Cz and Fz) are in the central region of the brain, six electrodes (i.e., T7, P7, C3, P3, FC3, and CP3) are located in the left-brain region, and the remaining six electrodes (i.e., T8, P8, C4, P4, FC4, and CP4) are in the right region of the brain. These selective electrodes are capable of analyzing the effective network during the MI-demonstrated brain networks underlying the mental behavior in several brain parts, including the primary motor cortex, primary and secondary somatosensory cortices, lateral premotor cortex, supplementary and presupplementary motor areas, frontal cortex, temporal cortex, and parietal cortex [36,37]. These are the most suitable electrodes for the feasible implementation of MI-based BCI using multiple brain regions with a reduced number of electrode channels. Moreover, these channels have been endorsed by several researchers in their work [38–42]. The channels chosen for the BCI application are shown in Figure 2.

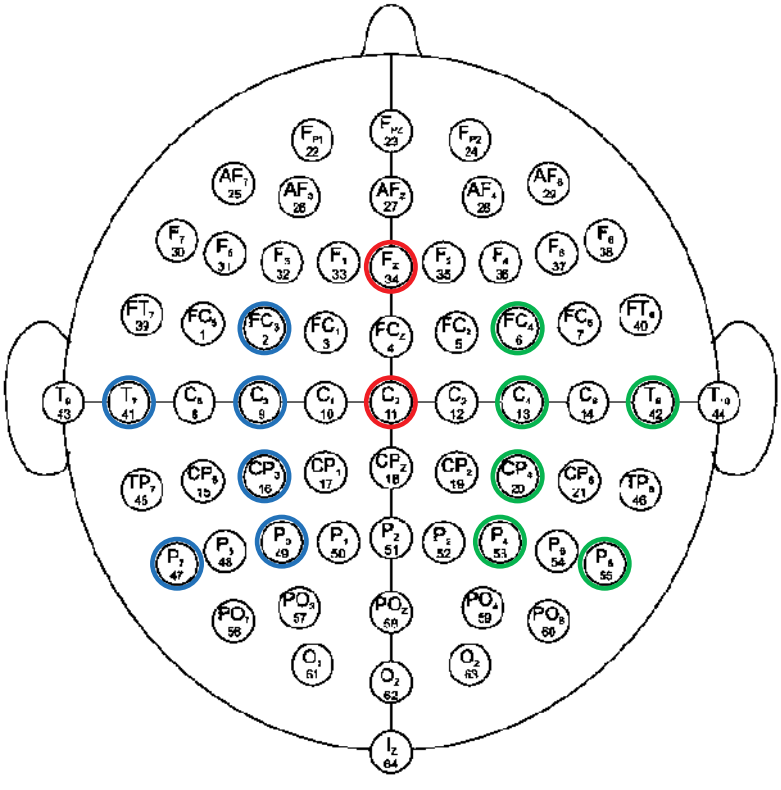


Figure 2. Standard 10–10 EEG electrode configuration. Selected electrodes from left, right and central region are encircled with blue, green and red color, respectively.

In this work, we excluded the data from 18 subjects—namely, S29, S30, S34, S37, S41, S51, S64, S72, S73, S74, S76, S88, S89, S92, S100, S102, S104, and S106—since they had contaminated EEG recordings or an insufficient number of samples in the available dataset. Thus, the EEG data from 91 out of the 109 subjects were used in this study. Among the several tasks available in the stated database, two-class motor imagery (imagination of opening/closing the left fist and imagination of opening/closing the right fist) was analyzed.

3.3. MI Paradigm

The subjects were asked to sit on a comfortable armchair in front of a screen in order to guide them through the experimental procedure. They were instructed not to move any parts of the body during the recording of the data. For the experiment, participants were instructed to perform the imagination of motor tasks as a target appeared on either the left or the right side of the screen. The subjects imagined opening and closing the corresponding fist until the target disappeared. Then, the subject relaxed. Every subject recorded three sessions for each type of MI task. However, a single session comprised seven to eight random trials of each class, i.e., left or right movement imagery. Each trial was carried out for four seconds, followed by a rest period of 4 s $\pm$ 5%, as shown in Figure 3.

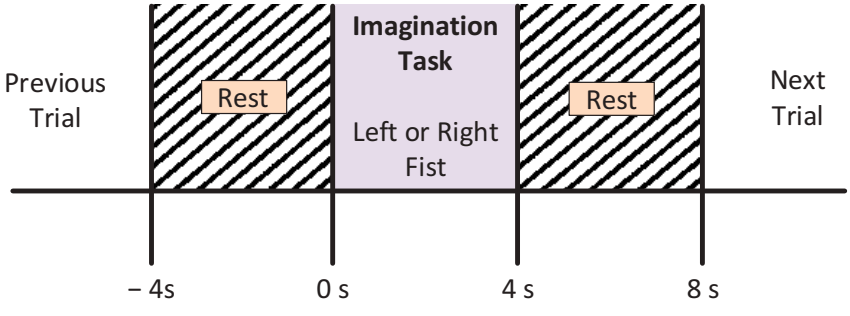


Figure 3. Motor imagery time scheme during the recording of the EEG signals.

4. Methodology

The proposed methodology aimed to utilize the estimation of brain connectivity (effective connectivity) for the classification of motor-imagery-based electroencephalographic signals using different classifiers. To the best of the authors' knowledge, a probabilistic neural network has never been used for the task of MI prediction using brain connectivity. This study provides a comparative analysis of all of the traditional and unique classifiers for two-class MI classification by using connectivity features.

Figure 4 represents the flow of the proposed methodology, in which raw EEG data from the 91 healthy subjects were preprocessed with different techniques. Based on the studies from the literature, 14 significant channels were selected out of a total 64 channels, and their effective connectivity in terms of the partial directed coherence (PDC) and directed transfer function (DTF) was used as a key feature. Both the PDC and DTF were analyzed separately by using the classifiers in order to recognize the 2-class MI-based patterns.

The proposed work was implemented on Intel® Xeon(R) CPU E3-1226 v3 at 3.30 GHz (installed memory: 16 GB). MATLAB 2020a was used as the programming platform for our proposed work.

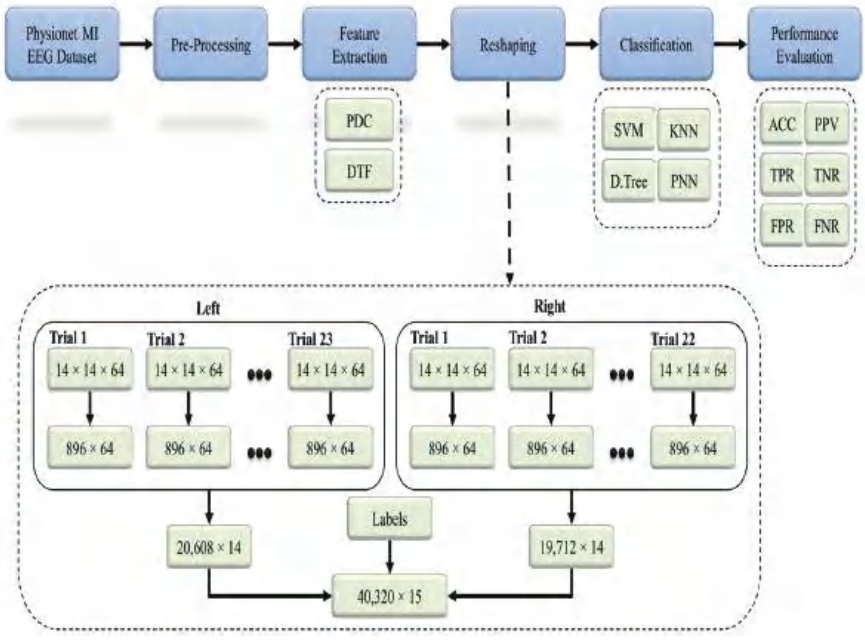


Figure 4. General workflow diagram for the classification of the two-class MI EEG using the effective connectivity matrices with the PDC and DTF.

4.1. Preprocessing

The enhancement of the raw signal is the prerequisite and fundamental step of EEG signal processing. The raw data hold both significant information and artifacts/undesired components (i.e., eye blinks, eyeball movement, jaw clenches, heavy breaths, etc.). The exclusion of unnecessary components from the signal is very important in order to improve the signal-to-noise ratio (SNR).

The data from the 91 subjects were preprocessed using Brainstorm in MATLAB. Brainstorm is an open-source platform dedicated to the analysis of several types of brain recordings, such as fNIRS, MEG, ECoG, and EEG. A high-pass filter was applied at 0.1 Hz for the purpose of DC offset correction, while a notch filter was used at 60 Hz to eliminate the electrical interference. The raw signal was bandpass filtered between 7 and 32 Hz to exclude all of the frequency components other than mu and beta, as the studies [43–45] revealed the occurrence of MI patterns in the stated frequency range. However, the artifacts were removed using the EEGLAB-based artifact removal algorithm called artifact subspace reconstruction (ASR). Major artifacts identified by the ASR technique, including eye blinks, muscle noise, and sensor motions, were removed from the data. Furthermore, excessive preprocessing was avoided in this research, since the creator of the DTF/PDC suggested that unnecessary preprocessing may impact the causality information and should, thus, be avoided [46].

Although the data were preprocessed to remove the artifacts and unnecessary signals, there was still redundant information available, which was also eliminated. These redundant data were actually the duration during the trial when the participants rested and did not perform any type of MI task (see Figure 3). After discarding the undesired part of the MI trials, 4095 trials (4 sec each) were available for further signal processing. Since the sampling frequency was 160 Hz, the total number of samples for each trial was equal to $4 \times 160 = 640$. In this work, we used 14 significant channels, as discussed in the dataset description.

4.2. Feature Extraction

The next critical step of the signal processing of the BCI system after preprocessing was feature extraction. This process was intended to extract specific characteristics of the signals that encoded the messages or commands elicited in the user's brain by either evoked or spontaneous inputs. In this work, effective connectivity was estimated using the partial directed coherence (PDC) and directed transfer function (DTF) for the MI-based EEG classification.

The basic code for the calculation of the PDC and DTF is available at [47]. The PDC and DTF were calculated by adjusting the maximum frequency to 32 Hz, whereas the number of bins was set to 64. Both the PDC and DTF were calculated for every 4 s of EEG data, which referred to a single MI trial (see Figure 3). Every subject recorded 23 trials for the left direction and 22 trials for the right direction. The feature sets calculated for each trial were in the form of 3D matrices (i.e., $14 \times 14 \times 64$), which were converted into 2D (i.e., 896×14) by performing matrix reshaping in order to execute the classification process. For the left class, each of the 23 trials (896×14) was concatenated to get a ($20{,}608 \times 14$) matrix, whereas for the right class, each of the 22 trials of (896×14) was concatenated to get a ($19{,}712 \times 14$) matrix. To create the final feature set, both the left and right feature sets were combined along with an extra column (i.e., labels for both classes) to get a ($40{,}320 \times 15$) matrix. This final feature set for each subject set was used as an input for different classification algorithms for the 2-class MI prediction (see Figure 4).

4.2.1. Effective Connectivity

Effective connectivity can be interpreted as the indirect or direct influence of one neural system on another at either a synaptic level or a cortical level [48]. According to [49], the EC should be recognized as the time-dependent and simplest possible circuit diagram that replicates the timing relationships between the recorded neurons. There are several brain connectivity estimators, including non-linear estimators, linear estimators, bivariate estimators, and multivariate connectivity estimators. The difference between bivariate and multivariate estimators is presented in [50], which states that the presence of multiple channels (i.e., more than two) in the case of an interrelated system of channels in bivariate connectivity estimators provides erroneous information due to the electrode channels positioned at different distances causing notable delays in the recorded signal.

This fact has led the research community to consider multivariate estimation for effective connectivity. These multivariate estimators based on Granger causality (GC) [51,52] allow precise measurement of the directed connectivity by eliminating the problem caused by multiple channels in the bivariate method. The Granger causality (GC) was established to determine the causal connection between two signals. If past information of signal $X(t)$ is given with the past information of signal $Y(t)$, then the causality between $X(t)$ and $Y(t)$ can be measured by decreasing the error (prediction error) of signal $Y(t)$.

The partial directed coherence [53] and directed transfer function [54] are among the most widely used connectivity estimators based on the multivariate autoregressive model (MVAR) under the umbrella of Granger causality (GC) to evaluate the directional influences of any given pair of channels in a dataset. The PDC and DTF, which are based on the MVAR model, can detect causal interactions between the signals and identify the directional propagation of the EEG activity in terms of the frequency function. The frequency dependency of estimators is an essential aspect, since various EEG rhythms play different roles in the processing of information. The PDC and DTF are insensitive to volume conduction and are very tolerant towards noise, as they are based on the phase differences between channels of multivariate data. The physiological information provided by means of the parametric and multivariate autoregressive (MVAR)-based methods has revealed their effectiveness in brain research.

4.2.2. Partial Directed Coherence

Baccala and Sameshima [53] developed an analysis method called partial directed coherence (PDC) as an extension of Granger's conditional causality in the frequency domain. The analysis aims to set up the connectivity links among different brain regions for the frequency range selected from an electroencephalographic signal. In this work, we selected two frequency ranges, commonly known as alpha (7–13 Hz) and beta (13–32 Hz).

The multi-channel EEG data obtained by using multivariate autoregressive (MVAR) model can be defined as follows:

$$Y(t) = \sum_{r=1}^{l} A(r)Y(t-r) + E(t), \tag{1}$$

where $Y(t)$ denotes the 14-channel EEG time-series data, l is the model order, $A(r)$ is the coefficient matrix with lag r, and $E(t)$ is the error (prediction error) of the multivariate autoregressive model. Equation (1) can be rearranged to determine the prediction error as follows:

$$E(t) = \sum_{r=0}^{l} \hat{A}(r)Y(t-r), \tag{2}$$

where the following value belongs to $\hat{A}(r)$:

$$\hat{A}(r) = \begin{cases} 1 - A(r), & r = 0. \\ -A(r), & r > 0. \end{cases} \tag{3}$$

The error function in Equation (2) can be expressed in the frequency domain:

$$E(f) = A(f)Y(f), \tag{4}$$

where f represents the frequency. So,

$$A(f) = \sum_{r=0}^{l} \hat{A}(r) \exp^{-j2\pi fr}. \tag{5}$$

The PDC in the frequency domain can be calculated as

$$P_{ij}(f) = \frac{A_{ij(f)}}{\sqrt{\sum_{k=1}^{x} |a_{kj}(f)^2|^2}}. \tag{6}$$

In Equation (6), the number of analyzed channels (except for the current channel j) is denoted by x, while P_{ij} represents the PDC's correlation indicators from Y_j to Y_i at a specific frequency f. The capital A represents the whole coefficient matrix; however, the small a refers to the matrix elements. The sum of all causal inference estimations P_{ij} at certain frequencies while reviewing the influence on all channels x by channel j is 1. This proves that higher values of P_{ij} result in a greater influence of channel j on channel i.

4.2.3. Directed Transfer Function

Kaminski and Blinowska [54] presented an analysis method based on a multivariate model called the directed transfer function (DTF). This work generalized Granger's work to some extent and claimed to have significant superiority in brain connectivity estimation. The PDC and DTF differ in such a way that the PDC detects active direct directional coupling, while the directed transfer function illustrates the presence of both direct (i.e., the immediate causal influence path) and indirect (the signal traveling through intermediate structures rather than an instant direct causal influence path) directional signal propagation [55].

The only different step in the DTF is that of taking the inverse of $A(f)$ from Equation (5) and then performing the normalization.

$$H(f) = A(f)^{-1} \tag{7}$$

where $H(f)$ is the frequency-domain representation of the transfer function of the system and can be obtained as follows:

$$H(f) = \frac{1}{1 + \sum_{r=0}^{l} \hat{A}(r) \exp^{-j2\pi fr}}. \tag{8}$$

Then, the squared DTF from channel j to i can be given as

$$D_{i \leftarrow j}^2 = \frac{|H_{ij}(f)|^2}{\sum_{m=1}^{y} |H_{im}(f)|^2} \tag{9}$$

$$D_{i \leftarrow j} = \frac{|H_{ij}(f)|}{\sqrt{\sum_{m=1}^{y} |H_{im}(f)|^2}} \tag{10}$$

Here, $D_{i \leftarrow j}$ represents the normalized version of causality from channel j by channel i at some specific frequency f, while the transfer matrix of the multivariate autoregressive model is denoted by H_{ij}.

4.2.4. Connectivity Estimation

There were several steps (i.e., dataset adjustment, model order calculation, MVAR coefficient determination, and PDC/ DTF estimation) involved in the estimation of effective connectivity.

1. The first step in the estimation of connectivity was to adjust the MI EEG dataset by selecting the significant electrode channels from the primary dataset. The selection of 14 channels was already discussed in the preprocessing section.
2. After selecting the number of channels for the connectivity estimation, the data were divided into several trials, and the connectivity was computed separately for each trial.
3. Next critical step was the calculation of the model order l, which defined how many previous samples were needed for the prediction of the current samples. This was an automatic process that required a minimum and maximum range of order (i.e., 1–20 in our case) and an optimizing algorithm (i.e., Schwarz's Bayesian Information Criterion in our case) to select the order with the minimum error. However, the model order l was calculated by using the ARFIT toolbox with the parameters suggested by several researchers [56–58].
4. After estimating the optimized model order l, the next step incorporated the estimation of the MVAR coefficients (see Equation (1)).
5. The next step was to define the sampling frequency (i.e., 160 Hz) and the number of frequency bins among which the total frequency range (i.e., 7–32 Hz) would be divided for the connectivity analysis. In this work, we set the number of frequency bins to 64 so that the connectivity estimation process would be repeated 64 times for each bin of the frequencies.
6. The next step after the assignment of the above parameters was to find the difference $\hat{A}$ by subtracting the MVAR coefficient matrix A from the identity matrix I, as in Equation (3).
7. After calculating the difference from the identity matrix, a Fourier transform was performed to convert the time-series MVAR matrix $\hat{A}$ into the frequency domain $A(f)$ (see Equation (5)).
8. The estimation of both the PDC and DTF followed all of the above-mentioned steps; however, for the DTF, the only different step was to find the inverse of the frequency

domain matrix (i.e., $H(f) = A(f)^{-1}$), where H is called the transfer matrix of the system (see Equation (7)).

9. The final step in the estimation of the connectivity was the normalization of $A(f)$ and $H(f)$ for the PDC and DTF, respectively (see Equations (9) and (10)). The normalized outputs P and D were then called the PDC and DTF, respectively.
10. The 14-channel data were used while incorporating 64 frequency bins; therefore, the estimation of the PDC and DTF resulted in a $14 \times 14 \times 64$ matrix for each trial. Since the estimated connectivity matrix was in 3D, matrix reshaping was carried out to convert the 3D matrix into a 2D matrix for the purpose of classification.

4.3. Classification

Classification is a mechanism by which target variables or classes are predicted from given information. An extensive motor-imagery-based EEG dataset covering 91 healthy subjects was used for the estimation of brain connectivity with measures of the effective connectivity (i.e., PDC and DTF). The extracted features in terms of the DTF and PDC were classified by using four classifiers to predict 2 classes of left/right EEG MI. The k-fold cross-validation (CV) technique was used to explore the performance of the proposed method, and k = 5 and 10 were used for all experiments, as they were found to be appropriate.

K-nearest neighbors (KNN) belongs to the category of supervised learning, and it can be used for both classification and regression problems. In KNN, the item is identified by a majority vote from its neighbors, with the item being allocated to the most common class of its nearest k neighbors [59]. In this work, standardized data were used in KNN with the Euclidean distance function while using equal-distance weights, and the number of neighbors was set to 3.

Support vector machine (SVM) is a supervised-learning-based classification model that is explicitly described by a separating hyperplane [60]. This work aims to use a 2-class SVM with a Gaussian kernel function for the classification of MI-based EEG data. Standardized data were used in the stated variant of the SVM with a kernel scale of 0.9399 and a box constraint of 1.

The decision tree is a non-parametric predictive modeling approach to supervised machine learning that covers both classification and regression problems. As the name indicates, it utilizes a tree-like structure of decisions. In the tree model, class labels or final outcomes are represented by the leaves, while the decision nodes contain the split data [61]. Decision trees are significantly easy to interpret and simple to implement.

Probabilistic Neural Networks (PNNs)

A probabilistic neural networks (PNN) is a supervised approach that excels in decision making and classification. The network is trained with objects from known classes by using a collection of known instances, and thereafter, it can distinguish new objects according to the categories specified in the training set [62].

A PNN is a feedforward neural network that is commonly used in classification problems, and it does not require the extensive forward and backward calculations used in traditional neural networks. A PNN is intimately associated with the estimation of the Parzen window probability distribution function (PDF). This comprises several networks, where each sub-network is a Parzen window PDF estimator for each class [63]. A PNN is a feedforward multilayered network with four layers—an input layer, pattern layer, summation layer, and output layer—as shown in Figure 5.

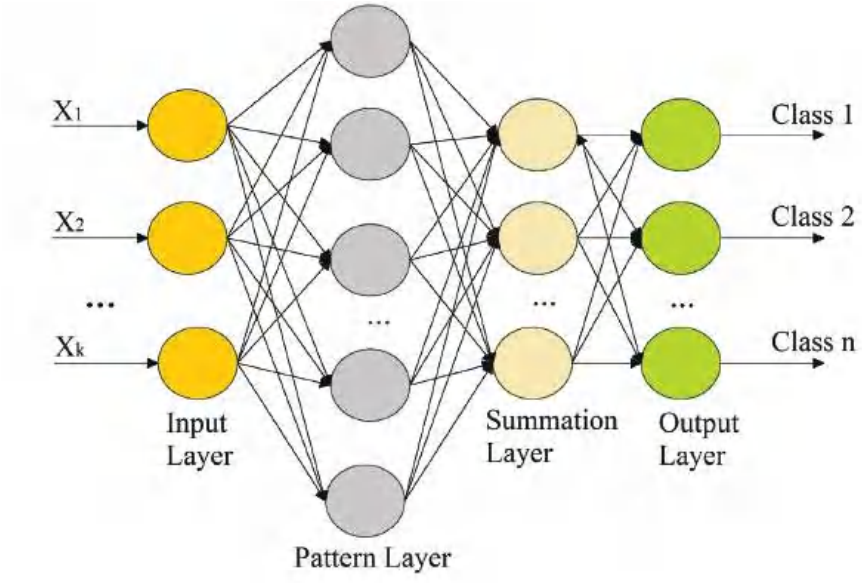


Figure 5. Illustration of a probabilistic neural network.

The 1st layer performs the distribution of the input values from the input layer to the neurons of the 2nd layer, i.e., the pattern layer. The pattern layer incorporates pattern units that are equal to the number of samples. Each pattern unit performs a two-step process on the input, including the computation of the Euclidean distance and the implementation of the kernel function, which computes the output when the pattern x is received from the input layer. The calculations from the pattern layer are passed to the 3rd layer (i.e., summation layer). The number of neurons in the 3rd layer is equal to the number of available classes, where each neuron is linked to all of the neurons in the pattern layer associated with the class represented by the particular unit. Each neuron represents the probability of the pattern x that is classified. The summation layer also determines the false detection of any given class. Weighted votes or values for each class from the 3rd layer are presented to the decision layer, where the majority voting is carried out in order to compare the values for each class, and the highest value predicts the target class.

The training process is perhaps the most time-intensive element of the preparation of a network for usage, since it depends on iterative algorithms that define the weight values. There are three stages in which the weight definition process occurs. In the 1st stage, weights are defined between the input and pattern layer. The definition of weights is instantaneous. As a result, training is accomplished by showing the training set examples all at once (after normalization) and sharing the node values for the layer of examples. In the 2nd stage, weights are predetermined between the pattern and summation layers and are equal to 1. In the 3rd stage, the weights are preset between the summation and output layers; however, their values may be given based on certain factors according to the training samples [62]. In this work, the spread of 0.04 was used as a hyperparameter for the proposed PNN model. This parameter was rigorously tuned to achieve consistent results.

4.4. Evaluation Parameters

The performance of the proposed work was determined by using several evaluation parameters. The evaluation parameters incorporated in this work are given below:

1. Classification accuracy (CA):

$$CA = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

2. Sensitivity or true positive rate (TPR):

$$TPR = \frac{TP}{TP + FN} \tag{12}$$

3. Specificity or true negative rate (TNR):

$$TNR = \frac{TN}{TN + FP} \tag{13}$$

4. Precision or positive predictive value (PPV):

$$PPV = \frac{TP}{TP + FP} \tag{14}$$

5. False positive rate (FPR):

$$FPR = \frac{FP}{FP + TN} \tag{15}$$

6. False negative rate (FNR):

$$FNR = \frac{FN}{FN + TP} \tag{16}$$

Here, TP, FP, TN, and FN represent true positives, false positives, true negatives, and false negatives, respectively.

4.5. Statistical Investigation

A multivariate analysis of variance (MANOVA) pairwise comparison was performed with a significance threshold of 0.05 to determine the statistical significance in the features of the left and right motor imagery EEG. For this, the 2 classes (left and right) were treated as fixed factors, while connections from all of the subjects were used as dependent variables. The mean difference between the 2 classes for each connection was tested with a 95% confidence interval, and the connection was marked as significant if the p-value was less than 0.05. As an adjustment for multiple comparisons, Bonferroni correction was implemented.

5. Results and Discussions

In this work, motor-imagery-based pattern recognition was experimented upon by using effective connectivity features, the PDC, and the DTF. The effective connectivity among three different brain portions—the central, left, and right regions—was measured using 14 electrode channels (i.e., 196 connectivity pairs) for which statistical analysis was carried out in order to determine the significant connections. The performance of the proposed work was evaluated separately for each subject; however, the results are presented as the averages of all 91 subjects.

5.1. Statistical Analysis

Figure 6 presents the connectivity pairs with significant differences that were obtained through MANOVA. The green color highlights the significant connections, whereas the white boxes indicate the insignificant pairs.

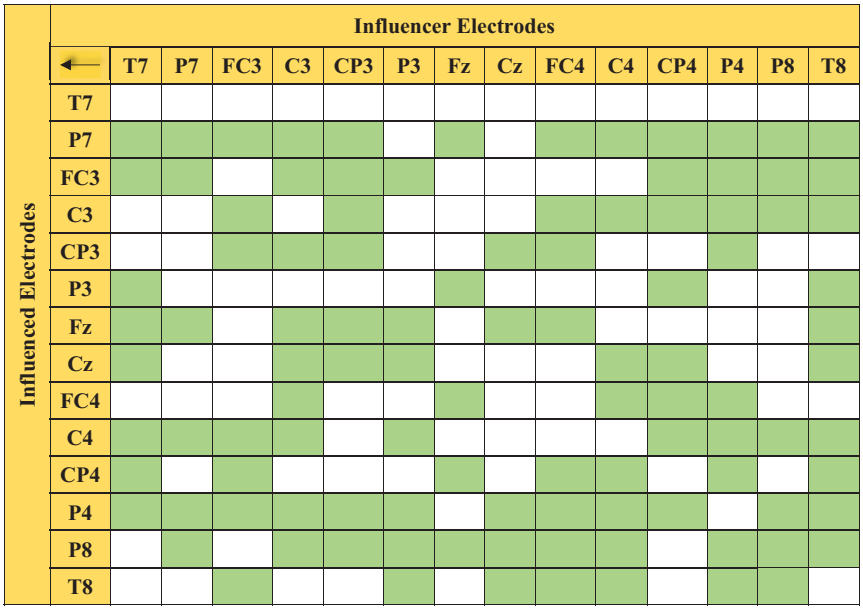


Figure 6. Matrix-form illustration of 105 significant connections (in green) listed in Table 1.

The statistical analysis showed that 105 out of the 196 connections were identified as significant with a 95% confidence level. With more than half (55.6%) of the connections at the 95% confidence level, the classification of the left/right MI EEG based on the 14 electrodes is expected to give good performance. Table 1 provides the *p*-value of the significant pairs. As described in Section 4.5, a *p*-value that is less than 0.05 basically indicates that the difference between the classes of the said connections is significant. This implies that the significant connections are dominant PDC features that largely contribute to the high classification accuracy of the proposed method.

Table 1. List of *p*-values of 105 significant connection pairs.

Pair	*p*-Value	Pair	*p*-Value	Pair	*p*-Value
T7←P7	0.000	CP3←Cz	0.025	C4←T8	0.022
T7←FC3	0.000	CP3←P4	0.000	CP4←P7	0.000
T7←P3	0.001	CP3←P8	0.000	CP4←FC3	0.000
T7←Fz	0.000	P3←FC3	0.000	CP4←C3	0.000
T7←Cz	0.011	P3←Fz	0.000	CP4←P3	0.003
T7←C4	0.000	P3←Cz	0.000	CP4←Cz	0.000
T7←CP4	0.008	P3←C4	0.024	CP4←FC4	0.005
T7←P4	0.000	P3←P4	0.019	CP4←C4	0.002
P7←P7	0.005	P3←P8	0.002	CP4←P4	0.000
P7←FC3	0.000	P3←T8	0.000	P4←P7	0.000
P7←Fz	0.000	Fz←P7	0.002	P4←FC3	0.000
P7←C4	0.017	Fz←P3	0.011	P4←C3	0.000
P7←P4	0.000	Fz←FC4	0.000	P4←CP3	0.000
P7←P8	0.000	Fz←CP4	0.010	P4←FC4	0.002

Table 1. *Cont.*

Pair	*p*-Value	Pair	*p*-Value	Pair	*p*-Value
FC3←P7	0.000	Fz←P8	0.000	P4←C4	0.027
FC3←C3	0.003	Cz←CP3	0.005	P4←CP4	0.000
FC3←CP3	0.005	Cz←Fz	0.000	P4←P8	0.000
FC3←Cz	0.000	Cz←P4	0.000	P4←T8	0.003
FC3←CP4	0.001	Cz←P8	0.021	P8←P7	0.000
FC3←P4	0.047	Cz←T8	0.021	P8←FC3	0.000
FC3←T8	0.000	FC4←P7	0.000	P8←C3	0.000
C3←P7	0.000	FC4←C3	0.000	P8←C4	0.000
C3←FC3	0.020	FC4←CP3	0.000	P8←P4	0.000
C3←CP3	0.000	FC4←Fz	0.000	P8←P8	0.040
C3←Fz	0.023	FC4←CP4	0.000	P8←T8	0.001
C3←Cz	0.000	FC4←P4	0.000	T8←P7	0.000
C3←FC4	0.000	FC4←P8	0.002	T8←FC3	0.000
C3←C4	0.000	FC4←T8	0.000	T8←C3	0.029
C3←P4	0.000	C4←P7	0.000	T8←P3	0.000
C3←P8	0.022	C4←C3	0.000	T8←Fz	0.003
CP3←P7	0.000	C4←Cz	0.000	T8←Cz	0.000
CP3←FC3	0.000	C4←FC4	0.000	T8←C4	0.000
CP3←C3	0.000	C4←CP4	0.000	T8←CP4	0.000
CP3←CP3	0.024	C4←P4	0.013	T8←P4	0.000
CP3←Fz	0.036	C4←P8	0.000	T8←P8	0.000

5.2. Classification of the MI EEG Using the EC

In this research work, two cases (with respect to the feature set and the classifier) were examined for the classification of the two-class motor imagery EEG recordings. A description of each case is given below.

- **Case 1:** The partial directed coherence (PDC) was used as a feature set with four classifiers: SVM, decision tree, KNN, and PNN.
- **Case 2:** The directed transfer function (DTF) was used as a feature set with the four classifiers stated in Case 1.

In Case 1, the PDC with SVM provided average classification accuracies of 96.30% and 97.45% for 91 subjects when using 5- and 10-fold cross-validation (CV), respectively. The PDC with KNN resulted in mean CAs of 97.85% for 5-fold CV and 98.63% for 10-fold CV. The PDC with the decision tree resulted in average CAs of 63.85% and 64.92%, whereas the average accuracies for the PDC with the PNN were 97.87% and 98.65% for 5- and 10-fold cross-validation, respectively. The classification accuracies and other evaluation parameters for Case 1 are presented in Table 2.

Table 2. Performance in left/right MI EEG classification when using the PDC.

EC	k-Fold	Classifier	CA (%)	TPR (%)	TNR (%)	PPV (%)	FPR (%)	FNR (%)
PDC	5-Fold CV	SVM	96.30	95.49	97.19	97.37	2.81	4.51
		KNN	97.85	97.90	97.81	97.90	2.19	2.10
		D. Tree	63.85	64.57	63.10	64.89	36.90	35.43
		PNN	97.87	97.93	97.82	97.92	2.18	2.07
	10-Fold CV	SVM	97.45	96.98	97.96	98.08	2.04	3.02
		KNN	98.63	98.68	98.60	98.60	1.40	1.32
		D. Tree	64.92	65.58	64.21	66.00	37.79	34.42
		PNN	98.65	98.68	98.63	98.69	1.37	1.32

In Case 2, the DTF with SVM provided average classification accuracies of 81.83% and 82.69% for 91 subjects when using 5- and 10-fold cross-validation (CV), respectively. The DTF with the KNN resulted in mean CAs of 82.04% for 5-fold CV and 82.67% for 10-fold CV. The DTF with the decision tree resulted in average CAs of 61.42% and 61.95%, whereas the average accuracies for the DTF with the PNN were 82.16% and 82.81% for 5- and 10-fold CV, respectively. The classification accuracies and other evaluation parameters for Case 2 are presented in Table 3.

Table 3. Performance in left/right MI EEG classification when using the DTF.

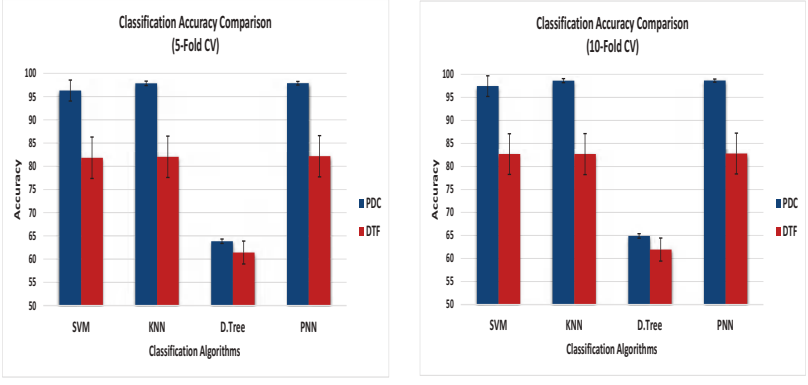
EC	k-Fold	Classifier	CA (%)	TPR (%)	TNR (%)	PPV (%)	FPR (%)	FNR (%)
DTF	5-Fold CV	SVM	81.83	77.43	88.84	91.50	11.16	22.57
		KNN	82.04	82.38	81.68	82.51	18.32	17.62
		D. Tree	61.42	62.19	60.62	62.58	39.38	37.81
		PNN	82.16	82.64	81.65	82.49	18.35	17.36
	10-Fold CV	SVM	82.69	78.55	89.04	91.46	10.96	21.45
		KNN	82.67	83.02	82.34	83.14	17.66	16.98
		D. Tree	61.95	62.72	61.15	63.03	38.85	37.28
		PNN	82.81	83.27	82.33	83.13	17.67	16.73

From Tables 2 and 3, it can be seen that, independent of the classifier, the classification based on PDC was better than that on the DTF. This was due to the fact that the PDC could eliminate the indirect effects of sources in the system [64]. Suppose that three sources—X1, X2, and X3—in a system are communicating such that X1 drives X2 while X2 drives X3. Thus, the connectivity between them should be X2 $\leftarrow$ X1 and X3 $\leftarrow$ X2, which will be correctly identified by the PDC. On the other hand, due to indirect effects [65], the DTF also suggests that X1 drives X3 (X3 $\leftarrow$ X1), which is incorrect. As the number of sources increases in a system, these indirect connections also increase. Since there is no way to differentiate between direct and indirect connections in the DTF, the classification accuracy based on the DTF is less than that with the PDC.

The evaluation parameters for the different classification algorithms stated in Cases 1 and 2 were calculated in terms of the TPR (true positive rate), TNR (true negative rate), PPV (positive predictive value), FPR (false positive rate), and FNR (false negative rate). The probabilistic neural network achieved the maximum values of the TPR, TNR, and PPV when used with the PDC as well as the DTF for both 5- and 10-fold cross-validation. The maximum values for the FPR and FNR and the minimum values for the TPR, TNR, and

PPV were recorded for the decision tree, which provided the lowest classification accuracy with both the PDC and DTF using 5- and 10-fold CV.

Figure 7 demonstrates the comparison of classification accuracies along with the errors of the classifiers using 5- and 10-fold CV, as described in Cases 1 and 2. Using the 5-fold cross-validation, the PNN outperformed all other classification algorithms by achieving average classification accuracies of 97.87% and 82.16% for the PDC and DTF, respectively, whereas the 10-fold CV resulted in enhanced maximum accuracies of 98.65% for the PDC and 82.81% for the DTF when using the PNN as a classification algorithm. On the other hand, the decision tree gave the lowest accuracy using the PDC and DTF for both 5- and 10-fold cross-validation. The error rate for each classifier using the DTF was greater than that of the classifiers when using the PDC. The minimum error was recorded for the PNN when using 10-fold CV, whereas the maximum was presented by the SVM when using 5-fold cross-validation.



(**a**) (**b**)

Figure 7. Classification accuracy of the four classifiers—SVM, KNN, decision trees, and PNN—based on the PDC and DTF as the feature sets when using (**a**) 5-fold CV and (**b**) 10-fold CV.

As discussed earlier, the MI classification was performed for 91 healthy subjects, and the results presented above are the averages of all of the subjects. Among all 91 subjects, the PNN's accuracy with the PDC varied between 94.01% and 98.28% for 5-fold CV and between 95.06% and 99.00% for 10-fold CV. In contrast, the PNN's accuracy for the DTF varied between 70.66% and 92.95% for 5-fold CV and between 71.08% and 93.33% for 10-fold CV.

However, the standard deviation of the classification accuracy among the 91 subjects for each case provided information about the stability of the given classifiers. The standard deviation of the classification accuracy in each case is given in Table 4.

Table 4. Standard deviation (SD) of the accuracy for the classification of MI EEG based on the PDC and DTF when using 5- and 10-fold cross-validation.

EC	Classifier	SD (%)		EC	Classifier	SD (%)	
		5-Fold	10-Fold			5-Fold	10-Fold
PDC	SVM	2.26	2.24	DTF	SVM	4.47	4.45
	KNN	0.48	0.44		KNN	4.46	4.46
	D.Tree	0.47	0.48		D.Tree	4.44	4.41
	PNN	0.39	0.34		PNN	2.47	2.50

From Table 4, it can be seen that the PNN classifier had the lowest standard deviation for both Case 1 and Case 2, whereas the SVM classifier had the maximum standard deviation for both cases. Therefore, the PNN was proven to be the most stable classifier in both cases.

Although the PNN outperformed the other classification algorithms, there was a small difference between the prediction accuracies of the PNN and KNN. One of the major disadvantages of kNN is that the technique is not precise when calculating class probabilities with low values of k [66]. However, the PNN is an exclusive classifier, since a number of classifications can map every input pattern. The PNN's main benefits include, an intrinsically parallel training process, no issues with local minima, a fast training procedure, and the assurance that the training structure converges on an optimum classifier as the size of the training set increases. Once the training samples have been added or deleted, no significant retraining is required. As a result, a PNN learns faster than neural networks and has already been shown to be successful in a range of tasks. A PNN is a supervised neural network that may be used for system categorization and pattern recognition based on these facts and benefits [67]. Another benefit of the PNN method is that probabilities for classification results may be immediately determined from a structural analysis. It is unlike other classification techniques, such as the SVM, which performs the process for calculating the probabilities associated with the classification results as a separate step after the model is created [68]. The computational cost of the proposed methodology is based on the k-fold cross-validation technique. It takes around 9.5 min on average to cross-validate each fold. However, the testing takes just a few seconds after the training procedure.

5.3. Comparison of the Proposed EC-Based MI EEG Classification Methods with Conventional Methods and Related Published Papers

We tested the prediction of the two-class MI EEG using the same 14-channel EEG dataset with several traditional feature extraction techniques, including the average power, root mean square, standard deviation, variance, entropy, discrete wavelet transform (DWT), and power spectral density (PSD). A comparison of the proposed and traditional methods is given in Table 5, which justifies the significance of the connectivity features compared to the traditional feature extraction techniques. Thus, it proves the hypothesis of achieving better results with the connectivity features.

Table 5. Comparison of the proposed method with the conventional feature extraction techniques in terms of classification accuracy (%).

Classifier	Proposed Features		Conventional Feature						
	PDC	**DTF**	**F1**	**F2**	**F3**	**F4**	**F5**	**F6**	**F7**
SVM	97.45	82.67	75.72	74.17	79.82	80.55	62.25	77.28	74.12
KNN	98.63	82.69	77.03	78.53	81.31	81.97	62.54	78.19	74.86
D.Tree	64.92	61.95	74.87	73.64	76.33	74.81	60.92	76.94	71.47
PNN	98.65	82.81	78.26	78.91	82.28	82.62	64.76	80.47	75.98

F1—Average power, F2—Root mean square, F3—Standard deviation F4—Variance, F5—Entropy, F6—DWT, and F7—PSD.

In contrast with the proposed work in which brain connectivity analysis was utilized in the prediction of two-class motor imagery with the Physionet MI database, Sagee et al. [69] used wavelet decomposition for mu and beta rhythms with Naive Bayes and ANN classifiers to achieve the accuracies of 86.31% and 93.05%, respectively. Kim et al. [70] used the strong uncorrelating transform complex common spatial patterns (SUTCCSP) algorithm to extract features after obtaining the mu and beta bands through multivariate empirical mode decomposition (MEMD). The authors achieved an accuracy of 77.70% with a random forest classifier by using the extracted features. Dose et al. [71] extended the use of deep neural networks by incorporating subject-specific adaptation with transfer learning to

get 86.49% accuracy for two-class MI prediction. Lun et al. [72] used a novel deep learning framework based on graph convolutional neural networks (GCNs) that learned the generalized features, and this achieved a classification accuracy of 88.57%. Qiu et al. [73] calculated the symbolic transfer entropy (STE) between electrode channels and constructed the brain networks of various cognitive behaviors of each participant by using the directed minimum spanning tree (DMST) algorithm. Finally, the spectral distribution set scoring (SDSS) method was used to recognize 69.35% of the labels. Carlos et al. [74] used a functional-connectivity-based graph method to acquire features and used PSD-based feature selection techniques to obtain 90% accuracy by using a linear discriminant analysis (LDA) classifier. Funda et al. [75] proposed a two-stage channel selection method and local transformation-based feature extraction for the classification of motor imagery/movement tasks and achieved a significant prediction of 95.95% by using KNN (see Table 6).

Table 6. Comparison of the classification accuracy with related papers that used the same dataset (the Physionet EEG motor imagery dataset).

Work	Year	Channels	Features	Classification Method	Accuracy (%)
Y. Kim et al. [70]	2016	14	Strong uncorrelating transform complex common spatial patterns (SUT-CCSP)	Random Forest	77.70
GS. Sagee et al. [69]	2017	64	Mu and beta rhythms	ANN	93.05
C. Filho et al. [74]	2017	64	FC-based graph method	LDA	90.00
H. Dose et al. [71]	2018	64	Raw EEG data	1D CNN	86.49
FK. Onay et al. [75]	2019	22	1D local transformation-based features	KNN	95.95
X. Lun et al. [72]	2020	64	Time-resolved EEG data	Graph CNN (GCNs)	88.57
L. Qiu et al. [73]	2020	64	symbolic transfer entropy (STE)	Directed minimum spanning tree (DMST)	69.35
Proposed Work	2021	14	Partial directed coherence (PDC)	**PNN**	**98.65**
			Directed transfer function (DTF)		**82.81**

6. Conclusions

This study aimed to use the effective connectivity of the brain by considering inter-channel/region relationships during the imagination of left-/right-hand movements, which were determined by using the partially directed coherence (PDC) and directed transfer function (DTF). The PDC and DTF were then used as feature inputs for four types of machine learning (ML) algorithms for the classification of left and right motor imagery classes. A probabilistic neural network (PNN) based on the PDC features outperformed other PDC- and DTF-based ML algorithms. The PDC manifested its prediction superiority over DTF due to its ability to eliminate the indirect effects of sources in the system. The proposed framework solved a major disadvantage of conventional techniques by integrating a better knowledge of the brain's neural patterns to improve the consistency and complexity, as well as by using multiple brain areas instead of just sensorimotor regions. The high classification accuracy of the left and right motor imagery based on the effective connectivity of the brain strongly supports the use of the PDC in BCI motor imagery applications. The use of graph theory for the identification of specific motor imagery patterns can be incorporated into the proposed technique to help people with motor disabilities by providing them with some reliable assistive technology, such as a brain-controlled wheelchair. However, there is a need for further testing on multiple classes (more than two), along with the optimization of the process in order to reduce the computational cost before the proposed technique can find its way in real-time BCI applications.

Author Contributions: Conceptualization, M.Z.Y. and M.A.A.; methodology, M.A.A. and D.M.K.; software, M.A.A., D.M.K., and N.K.; analysis and validation , M.A.A. and D.M.K., N.Y.; writing—original draft preparation, M.A.A.; writing—review and editing, M.A.A., M.Z.Y., N.Y., and D.M.K.; supervision, M.Z.Y. and N.Y.; funding acquisition, M.Z.Y. and M.E. All authors have read and agreed to the submitted version of the manuscript.

Funding: This research is supported in part by the Ministry of Education Malaysia under the Higher Institutional Centre of Excellence (HICoE) Scheme awarded to the Centre for Intelligent Signal and Imaging Research (CISIR), in part by the Graduate Assistantship Scheme of Universiti Teknologi PETRONAS, and in part by Iqra University (Pakistan) Fund under Grant 015ME0-224.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The databases used in this study are public and can be found at https://www.physionet.org/content/eegmmidb/1.0.0/ (accessed on 20 August 2021).

Acknowledgments: The authors would like to acknowledge the financial support provided by Ministry of Education Malaysia under the Higher Institutional Centre of Excellence (HICoE) Scheme, Iqra University, Pakistan under Grant 015ME0-224 and the GA scheme by the Center for Graduate Studies (CGS) at Universiti Teknologi PETRONAS (UTP).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Abiri, R.; Borhani, S.; Sellers, E.W.; Jiang, Y.; Zhao, X. A comprehensive review of EEG-based brain–computer interface paradigms. *J. Neural Eng.* **2019**, *16*, 011001. [CrossRef] [PubMed]
2. Jin, J.; Liu, C.; Daly, I.; Miao, Y.; Li, S.; Wang, X.; Cichocki, A. Bispectrum-based channel selection for motor imagery based brain-computer interfacing. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2020**, *28*, 2153–2163. [CrossRef] [PubMed]
3. Biasiucci, A.; Franceschiello, B.; Murray, M.M. Electroencephalography. *Curr. Biol.* **2019**, *29*, R80–R85. [CrossRef] [PubMed]
4. Padfield, N.; Zabalza, J.; Zhao, H.; Masero, V.; Ren, J. EEG-based brain-computer interfaces using motor-imagery: Techniques and challenges. *Sensors* **2019**, *19*, 1423. [CrossRef]
5. Kam, T.E.; Suk, H.I.; Lee, S.W. Non-homogeneous spatial filter optimization for ElectroEncephaloGram (EEG)-based motor imagery classification. *Neurocomputing* **2013**, *108*, 58–68. [CrossRef]
6. Asensio-Cubero, J.; Gan, J.; Palaniappan, R. Multiresolution analysis over simple graphs for brain computer interfaces. *J. Neural Eng.* **2013**, *10*, 046014. [CrossRef]
7. Tam, W.k.; Wu, T.; Zhao, Q.; Keefer, E.; Yang, Z. Human motor decoding from neural signals: A review. *BMC Biomed. Eng.* **2019**, *1*, 22. [CrossRef]
8. Wasaka, T.; Kida, T.; Kakigi, R. Facilitation of information processing in the primary somatosensory area in the ball rotation task. *Sci. Rep.* **2017**, *7*, 1–9. [CrossRef]
9. Mišić, B.; Sporns, O. From regions to connections and networks: New bridges between brain and behavior. *Curr. Opin. Neurobiol.* **2016**, *40*, 1–7. [CrossRef]
10. McEvoy, L.K.; Smith, M.E.; Gevins, A. Dynamic cortical networks of verbal and spatial working memory: Effects of memory load and task practice. *Cereb. Cortex* **1998**, *8*, 563–574. [CrossRef]
11. Rajabioun, M. Motor imagery classification by active source dynamics. *Biomed. Signal Process. Control* **2020**, *61*, 102028. [CrossRef]
12. Liang, S.; Choi, K.S.; Qin, J.; Wang, Q.; Pang, W.M.; Heng, P.A. Discrimination of motor imagery tasks via information flow pattern of brain connectivity. *Technol. Health Care* **2016**, *24*, S795–S801. [CrossRef]
13. Panche, I.D.L.P.; Alvarez-Meza, A.M.; Orozco-Gutierrez, A. A data-driven measure of effective connectivity based on Renyi's α-entropy. *Front. Neurosci.* **2019**, *13*, 1277. [CrossRef]
14. Gu, L.; Yu, Z.; Ma, T.; Wang, H.; Li, Z.; Fan, H. EEG-based classification of lower limb motor imagery with brain network analysis. *Neuroscience* **2020**, *436*, 93–109. [CrossRef]
15. Rahman, M.A.; Khanam, F.; Ahmad, M. Detection of effective temporal window for classification of motor imagery events from prefrontal hemodynamics. In Proceedings of the 2019 International Conference on Electrical, Computer and Communication Engineering (ECCE), Cox's Bazar, Bangladesh, 7–9 February 2019; pp. 1–6.
16. Lang, E.W.; Tomé, A.M.; Keck, I.R.; Górriz-Sáez, J.; Puntonet, C.G. Brain connectivity analysis: A short survey. *Comput. Intell. Neurosci.* **2012**, 412512. [CrossRef]
17. Khan, D.M.; Yahya, N.; Kamel, N.; Faye, I. Automated Diagnosis of Major Depressive Disorder Using Brain Effective Connectivity and 3D Convolutional Neural Network. *IEEE Access* **2021**, *9*, 8835–8846. [CrossRef]
18. Sporns, O. Brain connectivity. *Scholarpedia* **2007**, *2*, 4695. [CrossRef]
19. Khan, D.; Kamel, N.; Muzaimi, M.; Hill, T. Effective Connectivity for Default Mode Network Analysis of Alcoholism. *Brain Connect.* **2021**, *11*, 12–29. [CrossRef]

20. Li, Y.; Lei, M.; Zhang, X.; Cui, W.; Guo, Y.; Huang, T.W.; Wei, H.L. Boosted Convolutional Neural Networks for Motor Imagery EEG Decoding with Multiwavelet-based Time-Frequency Conditional Granger Causality Analysis. *arXiv* **2018**, arXiv:1810.10353.
21. Khan, D.M.; Yahya, N.; Kamel, N.; Faye, I. Effective Connectivity in Default Mode Network for Alcoholism Diagnosis. *IEEE Trans. Neural Syst. Rehabil. Eng.* **2021**, *29*, 796–808. [CrossRef]
22. Billinger, M.; Brunner, C.; Müller-Putz, G.R. Single-trial connectivity estimation for classification of motor imagery data. *J. Neural Eng.* **2013**, *10*, 046006. [CrossRef] [PubMed]
23. Wang, K.; Xu, M.; Wang, Y.; Zhang, S.; Chen, L.; Ming, D. Enhance decoding of pre-movement EEG patterns for brain–computer interfaces. *J. Neural Eng.* **2020**, *17*, 016033. [CrossRef] [PubMed]
24. Hu, S.; Wang, H.; Zhang, J.; Kong, W.; Cao, Y. Causality from Cz to C3/C4 or between C3 and C4 revealed by Granger causality and new causality during motor imagery. In Proceedings of the 2014 International Joint Conference on Neural Networks (IJCNN), Beijing, China, 6–11 July 2014; pp. 3178–3185.
25. Rathee, D.; Cecotti, H.; Prasad, G. Single-trial effective brain connectivity patterns enhance discriminability of mental imagery tasks. *J. Neural Eng.* **2017**, *14*, 056005. [CrossRef] [PubMed]
26. Shalbaf, A.; Maghsoudi, A. Mental Arithmetic Task Recognition Using Effective Connectivity and Hierarchical Feature Selection from EEG Signals. *Basic Clin. Neurosci.* **2020**. [CrossRef]
27. Ginter, J., Jr.; Blinowska, K.; Kamiński, M.; Durka, P. Phase and amplitude analysis in time–frequency space—Application to voluntary finger movement. *J. Neurosci. Methods* **2001**, *110*, 113–124. [CrossRef]
28. Ginter, J., Jr.; Blinowska, K.; Kamiński, M.; Durka, P.; Pfurtscheller, G.; Neuper, C. Propagation of EEG activity in the beta and gamma band during movement imagery in humans. *Methods Inf. Med.* **2005**, *44*, 106–113. [CrossRef]
29. Chung, Y.G.; Kim, M.K.; Kim, S.P. Inter-channel connectivity of motor imagery EEG signals for a noninvasive BCI application. In Proceedings of the 2011 International Workshop on Pattern Recognition in NeuroImaging, Seoul, Korea, 16–18 May 2011; pp. 49–52.
30. Lee, M.; Yoon, J.G.; Lee, S.W. Predicting motor imagery performance from resting-state EEG using dynamic causal modeling. *Front. Hum. Neurosci.* **2020**, *14*, 321. [CrossRef]
31. Chen, D.; Li, H.; Yang, Y.; Chen, J. Causal connectivity brain network: A novel method of motor imagery classification for brain-computer interface applications. In Proceedings of the 2012 International Conference on Computing, Measurement, Control and Sensor Network, Taiyuan, China, 7–9 July 2012; pp. 87–90.
32. Chen, C.; Zhang, J.; Belkacem, A.N.; Zhang, S.; Xu, R.; Hao, B.; Gao, Q.; Shin, D.; Wang, C.; Ming, D. G-causality brain connectivity differences of finger movements between motor execution and motor imagery. *J. Healthc. Eng.* **2019**, *2019*. [CrossRef]
33. Li, X.; Ong, S.H.; Pan, Y.; Ang, K.K. Connectivity pattern modeling of motor imagery EEG. In Proceedings of the 2013 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB), Singapore, 16–19 April 2013; pp. 94–100.
34. Physionet. EEG Motor Movement/Imagery Dataset. 2009. Available online: https://www.physionet.org/content/eegmmidb/1.0.0/ (accessed on 20 August 2021).
35. Schalk, G.; McFarland, D.J.; Hinterberger, T.; Birbaumer, N.; Wolpaw, J.R. BCI2000: A general-purpose brain-computer interface (BCI) system. *IEEE Trans. Biomed. Eng.* **2004**, *51*, 1034–1043. [CrossRef]
36. Chai, M.T.; Amin, H.U.; Izhar, L.I.; Saad, M.N.M.; Abdul Rahman, M.; Malik, A.S.; Tang, T.B. Exploring EEG effective connectivity network in estimating influence of color on emotion and memory. *Front. Neuroinform.* **2019**, *13*, 66. [CrossRef]
37. Solodkin, A.; Hlustik, P.; Chen, E.E.; Small, S.L. Fine modulation in network activation during motor execution and motor imagery. *Cereb. Cortex* **2004**, *14*, 1246–1255. [CrossRef]
38. Shen, L.; Dong, X.; Li, Y. Analysis and classification of hybrid EEG features based on the depth DRDS videos. *J. Neurosci. Methods* **2020**, *338*, 108690. [CrossRef]
39. Zhang, Z.; Duan, F.; Sole-Casals, J.; Dinares-Ferran, J.; Cichocki, A.; Yang, Z.; Sun, Z. A novel deep learning approach with data augmentation to classify motor imagery signals. *IEEE Access* **2019**, *7*, 15945–15954. [CrossRef]
40. Liu, R.; Zhang, Z.; Duan, F.; Zhou, X.; Meng, Z. Identification of Anisomerous Motor Imagery EEG Signals Based on Complex Algorithms. *Comput. Intell. Neurosci.* **2017**, *2017*, 2727856. [CrossRef]
41. Frolov, N.S.; Pitsik, E.N.; Maksimenko, V.A.; Grubov, V.V.; Kiselev, A.R.; Wang, Z.; Hramov, A.E. Age-related slowing down in the motor initiation in elderly adults. *PLoS ONE* **2020**, *15*, e0233942. [CrossRef]
42. Jin, J.; Miao, Y.; Daly, I.; Zuo, C.; Hu, D.; Cichocki, A. Correlation-based channel selection and regularized feature optimization for MI-based BCI. *Neural Netw.* **2019**, *118*, 262–270. [CrossRef]
43. Tariq, M.; Trivailo, P.M.; Simic, M. Mu-Beta event-related (de) synchronization and EEG classification of left-right foot dorsiflexion kinaesthetic motor imagery for BCI. *PLoS ONE* **2020**, *15*, e0230184. [CrossRef]
44. Tacchino, G.; Coelli, S.; Reali, P.; Galli, M.; Bianchi, A.M. Bicoherence interpretation in EEG requires Signal to Noise ratio quantification: An application to sensorimotor rhythms. *IEEE Trans. Biomed. Eng.* **2020**, *67*, 2696–2704. [CrossRef]
45. Awais, M.A.; Yusoff, M.Z.; Yahya, N.; Ahmed, S.Z.; Qamar, M.U. Brain Controlled Wheelchair: A Smart Prototype. *J. Physics Conf. Ser.* **2020**, *1529*, 042075. [CrossRef]
46. Kaminski, M.; Blinowska, K.J. Directed transfer function is not influenced by volume conduction—Inexpedient pre-processing should be avoided. *Front. Comput. Neurosci.* **2014**, *8*, 61. [CrossRef]

47. Omidvarnia, A. Time-Varying EEG Connectivity: A Time-Frequency Approach. 2011. Available online: https://www.mathworks.com/matlabcentral/fileexchange/33721-time-varying-eeg-connectivity-a-time-frequency-approach (accessed on 20 August 2021).
48. Friston, K.J. Functional and effective connectivity: A review. *Brain Connect.* **2011**, *1*, 13–36. [CrossRef] [PubMed]
49. Aertsen, A.; Preissl, H. Dynamics of activity and connectivity in physiological neuronal networks. In *Nonlinear Dynamics and Neuronal Networks*; VHC-Verlag: Weinheim, Germany, 1991.
50. Kus, R.; Kaminski, M.; Blinowska, K.J. Determination of EEG activity propagation: Pair-wise versus multichannel estimate. *IEEE Trans. Biomed. Eng.* **2004**, *51*, 1501–1510. [CrossRef] [PubMed]
51. Granger, C.W. Investigating causal relations by econometric models and cross-spectral methods. *Econom. J. Econom. Soc.* **1969**, *37*, 424–438. [CrossRef]
52. Bressler, S.L.; Seth, A.K. Wiener-Granger causality: A well established methodology. *Neuroimage* **2011**, *58*, 323–329. [CrossRef] [PubMed]
53. Baccalá, L.A.; Sameshima, K. Partial directed coherence: A new concept in neural structure determination. *Biol. Cybern.* **2001**, *84*, 463–474. [CrossRef] [PubMed]
54. Kaminski, M.J.; Blinowska, K.J. A new method of the description of the information flow in the brain structures. *Biol. Cybern.* **1991**, *65*, 203–210. [CrossRef] [PubMed]
55. Baccalá, L.A.; Takahashi, D.Y.; Sameshima, K. Directed transfer function: Unified asymptotic theory and some of its implications. *IEEE Trans. Biomed. Eng.* **2016**, *63*, 2450–2460. [CrossRef]
56. Aller, M.; Noppeney, U. To integrate or not to integrate: Temporal dynamics of hierarchical Bayesian causal inference. *PLoS Biol.* **2019**, *17*, e3000210. [CrossRef]
57. Selig, K.; Shaw, P.; Ankerst, D. Bayesian information criterion approximations to Bayes factors for univariate and multivariate logistic regression models. *Int. J. Biostat.* **2020**, *1*. [CrossRef]
58. Khan, D.M.; Yahya, N.; Kamel, N. Optimum Order Selection Criterion for Autoregressive Models of Bandlimited EEG Signals. In Proceedings of the 2020 IEEE-EMBS Conference on Biomedical Engineering and Sciences (IECBES), Langkawi Island, Malaysia, 1–3 March 2021; pp. 389–394. [CrossRef]
59. Zhang, Y.; Cao, G.; Wang, B.; Li, X. A novel ensemble method for k-nearest neighbor. *Pattern Recognit.* **2019**, *85*, 13–25. [CrossRef]
60. Karimi, F.; Sultana, S.; Babakan, A.S.; Suthaharan, S. An enhanced support vector machine model for urban expansion prediction. *Comput. Environ. Urban Syst.* **2019**, *75*, 61–75. [CrossRef]
61. Guan, S.; Zhao, K.; Yang, S. Motor imagery EEG classification based on decision tree framework and Riemannian geometry. *Comput. Intell. Neurosci.* **2019**, *2019*, 5627156. [CrossRef]
62. Vicino, F. The probabilistic neural network. *Subst. Use Misuse* **1998**, *33*, 335–352. [CrossRef]
63. Casale, F.P.; Gordon, J.; Fusi, N. Probabilistic neural architecture search. *arXiv* **2019**, arXiv:1902.05116.
64. Sameshima, K.; Baccala, L.A. *Methods in Brain Connectivity Inference through Multivariate Time Series Analysis*; CRC Press: Boca Raton, FL, USA, 2014.
65. Kamiński, M.; Ding, M.; Truccolo, W.A.; Bressler, S.L. Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. *Biol. Cybern.* **2001**, *85*, 145–157. [CrossRef]
66. Gweon, H.; Schonlau, M.; Steiner, S.H. The k conditional nearest neighbor algorithm for classification and class probability estimation. *PeerJ Comput. Sci.* **2019**, *5*, e194. [CrossRef]
67. El Emary, I.M.; Ramakrishnan, S. On the application of various probabilistic neural networks in solving different pattern classification problems. *World Appl. Sci. J.* **2008**, *4*, 772–780.
68. Satapathy, S.K.; Dehuri, S.; Jagadev, A.K.; Mishra, S. *EEG Brain Signal Classification for Epileptic Seizure Disorder Detection*; Academic Press: Cambridge, MA, USA, 2019.
69. Sagee, G.; Hema, S. EEG feature extraction and classification in multiclass multiuser motor imagery brain computer interface using Bayesian Network and ANN. In Proceedings of the 2017 International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICICT), Kerala, India, 6–7 July 2017; pp. 938–943.
70. Kim, Y.; Ryu, J.; Kim, K.K.; Took, C.C.; Mandic, D.P.; Park, C. Motor imagery classification using mu and beta rhythms of EEG with strong uncorrelating transform based complex common spatial patterns. *Comput. Intell. Neurosci.* **2016**, *2016*, 1489692. [CrossRef]
71. Dose, H.; Møller, J.S.; Iversen, H.K.; Puthusserypady, S. An end-to-end deep learning approach to MI-EEG signal classification for BCIs. *Expert Syst. Appl.* **2018**, *114*, 532–542. [CrossRef]
72. Lun, X.; Jia, S.; Hou, Y.; Shi, Y.; Li, Y.; Yang, H.; Zhang, S.; Lv, J. GCNs-Net: A Graph Convolutional Neural Network Approach for Decoding Time-resolved EEG Motor Imagery Signals. *arXiv* **2020**, arXiv:2006.08924.
73. Qiu, L.; Nan, W. Brain Network Constancy and Participant Recognition: An Integrated Approach to Big Data and Complex Network Analysis. *Front. Psychol.* **2020**, *11*, 1003. [CrossRef]
74. Stefano Filho, C.A.; Attux, R.; Castellano, G. EEG sensorimotor rhythms' variation and functional connectivity measures during motor imagery: Linear relations and classification approaches. *PeerJ* **2017**, *5*, e3983. [CrossRef]
75. Onay, F.K.; Köse, C. Assessment of CSP-based two-stage channel selection approach and local transformation-based feature extraction for classification of motor imagery/movement EEG data. *Biomed. Eng. Technol.* **2019**, *64*, 643–653. [CrossRef]

sensors

MDPI

Article

Comparing Methods of Feature Extraction of Brain Activities for Octave Illusion Classification Using Machine Learning

Nina Pilyugina [1,*], Akihiko Tsukahara [2] and Keita Tanaka [2]

[1] Graduate School of Advanced Science and Technology, Tokyo Denki University, Hiki-gun, Saitama 350-0394, Japan
[2] Graduate School of Science and Engineering, Tokyo Denki University, Hiki-gun, Saitama 350-0394, Japan; tsukahara@mail.dendai.ac.jp (A.T.); ktanaka@mail.dendai.ac.jp (K.T.)
* Correspondence: 18udq91@ms.dendai.ac.jp

Abstract: The aim of this study was to find an efficient method to determine features that characterize octave illusion data. Specifically, this study compared the efficiency of several automatic feature selection methods for automatic feature extraction of the auditory steady-state responses (ASSR) data in brain activities to distinguish auditory octave illusion and nonillusion groups by the difference in ASSR amplitudes using machine learning. We compared univariate selection, recursive feature elimination, principal component analysis, and feature importance by testifying the results of feature selection methods by using several machine learning algorithms: linear regression, random forest, and support vector machine. The univariate selection with the SVM as the classification method showed the highest accuracy result, 75%, compared to 66.6% without using feature selection. The received results will be used for future work on the explanation of the mechanism behind the octave illusion phenomenon and creating an algorithm for automatic octave illusion classification.

Keywords: feature selection; machine learning; octave illusion; auditory illusion; MEG

Citation: Pilyugina, N.; Tsukahara, A.; Tanaka, K. Comparing Methods of Feature Extraction of Brain Activities for Octave Illusion Classification Using Machine Learning. *Sensors* **2021**, *21*, 6407. https://doi.org/10.3390/s21196407

Academic Editor: Manuel José Cabral dos Santos Reis

Received: 31 August 2021
Accepted: 22 September 2021
Published: 25 September 2021

1. Introduction

The auditory illusion is a false perception of real auditory stimuli. Unlike hallucinations, which have no sensory base, auditory illusions are always caused by external stimuli. Compared with visual illusions, people who perceive auditory illusions are not always aware of them. It is difficult for the human brain to separate the real and perceived sounds. In addition to physical pathologies, the ability to classify them depends on the mental status of the subject. Not all auditory illusions are symptoms of psychological disorders; the main characteristics of pathological illusions are their connection with the subject's painful experiences and worries and the absence of the context of the situation. Auditory illusions can accompany depression, panic disorders, delirium, and other mental problems [1]. Therefore, understanding the mechanisms underlying auditory illusions will contribute to our knowledge of pathological mental issues.

The octave illusion is one of the less-studied types of auditory illusions. It is induced by two dichotic sounds (400 and 800 Hz) played simultaneously and constantly in both ears [2]. The main characteristic of this phenomenon is the perception that occurs only in one ear at a time. A low tone presented to the right ear and a high tone to the left ear can be perceived in four main patterns. Perception is described as a single high-pitch tone in one ear, alternating with a single low-pitch one in another ear (Figure 1).

The behavioral explanation for octave illusion is the "what" and "where" model, in which "what" is a perceived sound determined by the dominant ear and "where" is a sound location defined in the ear receiving a high tone [3].

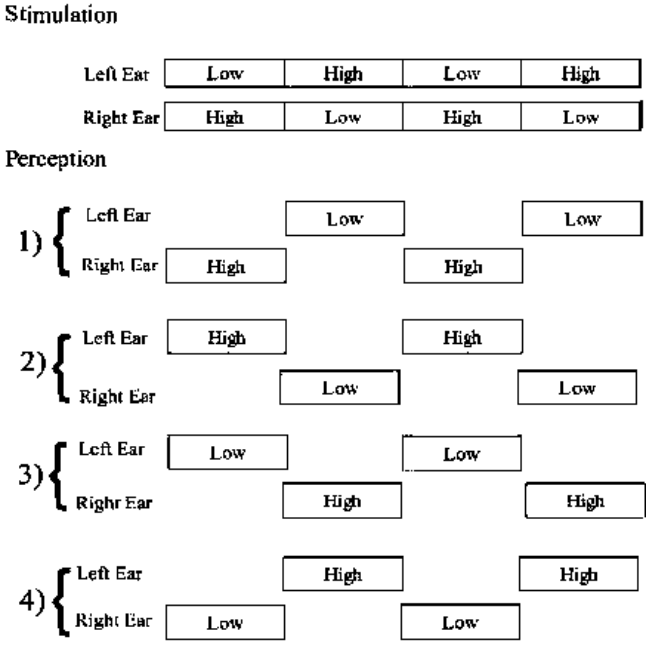


Figure 1. Octave illusion stimulation and four perception patterns.

However, the difference in brain responses between subjects who experience the illusion and those who do not is unclear. Although it has been proven that the pitch of the illusion perception has a main neural counterpart bilaterally in Heschl's gyrus (primary auditory cortex), the processes underlying the octave illusion have not been clarified yet. We suggest that it is possible to separate illusion (ILL) and nonillusion (non-ILL) groups based on the difference in auditory steady-state responses (ASSRs) using machine learning methods.

Machine learning has a wide application in the biomedical field. It is being used for EEG-based BCI for classification person intentions [4], gait decoding [5], or short/zero-calibration calibration [6]. Along with EEG data, machine learning is being used for the analysis of fMRI data [7] and MEG data. There are studies dedicated to applying deep learning to MEG data source localizations [8] and decoding signals [9]. Machine learning has also proved itself as a powerful tool for recognizing subtle patterns in complex data, such as ASSR [10,11].

ASSRs are auditory evoked potentials that arise in response to rapid auditory stimulation. They can be used as a measurement to estimate the brain's ability to generate responses, which can be used to differentiate subjects with normal and pathological hearing sensitivity [12]. Pathological hearing sensitivity often corresponds with mental diseases, such as schizophrenia, and researching ASSRs can help understand these problems as well.

In this study, there was no evident gap between the average ASSR responses (Figure 2) of the ILL and non-ILL groups during the octave illusion stimulation (Figure 3). Therefore, it is impossible to distinguish them by simple comparison, and we hypothesized that the selected features of ASSR patterns of the left and right hemispheres would provide enough information for binary classification. However, the usual process of data selection, when certain features are added or removed individually depending on the results, is more difficult to implement for this task because of the lack of information about the octave illusion. Because we do not know what exactly defines the octave illusion, limiting the size of the dataset risks losing valuable features without resolving problems with overfitting or

improving accuracy. Therefore, we investigated how the use of automatic feature selection accomplishes octave illusion classification using machine learning.

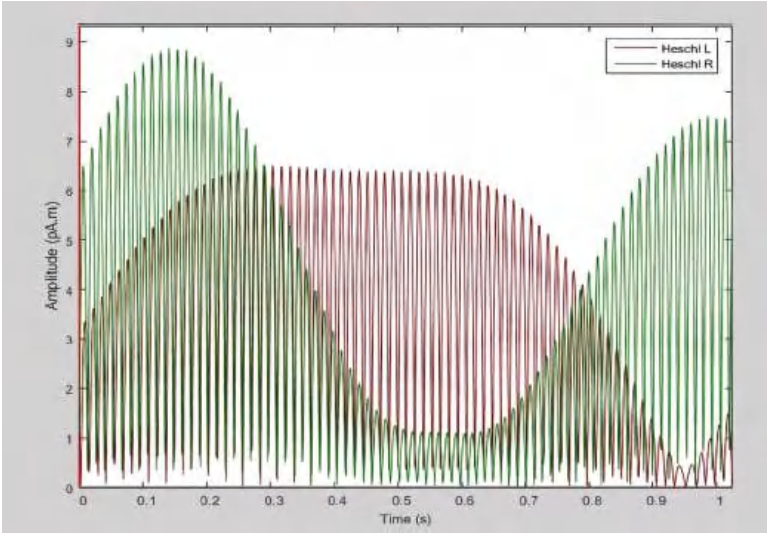


Figure 2. The random example of ASSR data used in this study.

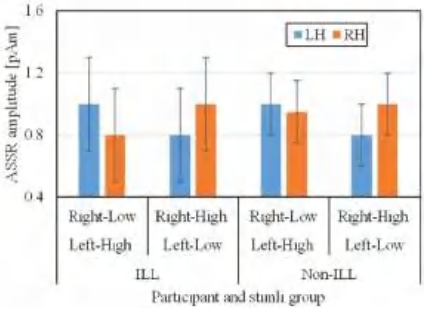


Figure 3. ASSR amplitude in nonillusion and illusion groups. RH and LH show the ASSR amplitude in the right and left hemispheres.

In machine learning, gathering a sufficient number of features is a vital requirement for classification tasks. However, increasing the number of features improves the classification abilities only to a certain point. This is called the curse of dimensionality. The curse of dimensionality is a common problem in machine learning caused by exponentially increasing errors with the number of features. A larger number of features requires a larger dataset, but because practically the number of training data is fixed, the classifier's performance will drop after the number of features reaches a certain point, depending on the size of the dataset.

The excessive number of features also leads to other problems:

1. Overfitting. It is a condition when the model has learned so many random fluctuations and noise that it cannot learn from new data.
2. Too many features make each observation in the dataset equidistant from all others. However, if all data are approximately equidistant from each other, then all data look equal, and no significant predictions can be made.

Feature selection refers to several methods that resolve this fundamental problem by the dimensional reduction of unnecessary variables. From a set of features $F = \{f1, f2, \ldots, fn\}$, feature selection methods define the ones that contribute the most to the learning ability. Feature reduction helps to improve the classifier's learning abilities, reduces overfitting and training time, and removes unnecessary noise.

Automatic feature selection is a popular method for the inclusion of brain data features, such as certain features of motor imagery [13] and features important for diagnosis using PET data [14], or for automated electroencephalography (EEG) data classification [15]. However, auditory illusion data have not been studied sufficiently, and there is no universal approach or strong basis for applying feature selection algorithms. Analyzing selected features that define octave illusion classification will contribute to the general understanding of auditory illusion mechanisms and, accordingly, the mental issues' processes.

We used magnetoencephalography (MEG) data because, unlike EEG, it provides the origins of brain functional activity. Functional magnetic resonance imaging (fMRI), which simply measures blood flow instead of directly measuring the brain's signals, also does not provide the necessary information. Moreover, the combination of MEG and the frequency-tagging method provides access to the contribution of each ear to the responses in the auditory cortices of each hemisphere. Therefore, we suggest that analyzing ASSR through MEG data will reveal the difference in auditory cortex activity between the auditory cortex of the ILL and non-ILL groups.

In this study, we aimed to find the most efficient union of the automatic selection method and machine learning method by comparing their various combinations. Considering the analyzed literature, to the best of our knowledge, this is the first study dedicated to the automatic feature extraction of octave illusion data for the classification of ILL and non-ILL groups.

2. Materials and Methods

2.1. Experimental Paradigm

This study involved MEG data of 17 male right-handed participants (9 ILL and 8 non-ILL) with a mean age and standard deviation of 21.4 ± 1.09 years. All participants were right-handed and had no history of otolaryngological or neurological disorders. All participants provided written consent after being informed of the nature of the study. This study was performed in accordance with the Declaration of Helsinki and was approved by the Research Ethics Committee of Tokyo Denki University.

MEG was recorded using a 306-channel whole-head-type brain magnetic field measurement device (VectorView 306, Elekta Neuromag, Neuromag, Helsinki, Finland). The brain magnetic field measurement device was installed in a magnetically shielded room, and the octave illusion tones were presented to the participants from the stimulation computer, after which the MEG was measured. After the analog-to-digital conversion of the measured MEG, the data were loaded into a computer at a sampling frequency of 1000 Hz.

Adobe Audition CS6 (Adobe Systems Incorporated) was used to generate tones. For a higher tone, the sound intensity was set to 3 dB, which is lower than the sound pressure level of the low tone.

2.2. Behavioral Testing

To classify the participants into ILL and non-ILL groups, we conducted a behavioral experiment in which each participant was equipped with headphones (E-A-RTONE 3A, Aearo Company Auditory Systems, Indianapolis, IN, USA) and presented with an octave illusion sound from a computer (ThinkPad Lenovo). Tones that were 513 ms long with modulation frequencies of 400 and 800 Hz (Figure 4) were played to the left and right ears, and each participant wrote on paper the perceptual pattern while the combination of the first stimulus scales and modulation frequencies of the left and right ears were changing. All participants listened to the sounds until they fully understood the perceptual pattern.

As instructed, the participants wrote "Low" when they perceived a low tone, wrote "High" when they perceived a high tone, or left the space blank in the case of zero perception.

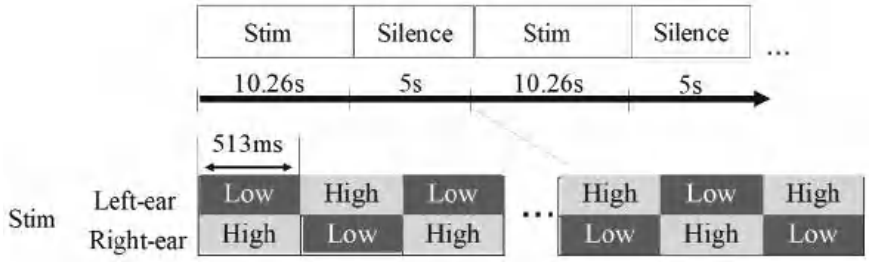


Figure 4. Experimental design. Audio stimuli had low (400 Hz) and high (800 Hz) tones, with a duration.

The behavioral experiment results showed that the participants who alternately perceived the high and low tones from each ear were classified as the ILL group, and those who perceived the actual sounds were classified as the non-ILL group.

2.3. Preprocessing

Each participant's session was cut to 1026 ms, one switch of 513 ms long tone, and averaged 96 times. All data were then subjected to bandpass filtering at 36–38 Hz. In the last preprocessing step, we applied source estimation to Heschl's gyrus (primary auditory cortex).

In this way, one set of features consisted of 2052 features of ASSR values (1026 values of each ms from each left and right hemisphere). Because we only have data of 17 subjects, 2052 features for one sample was considered as an extremely large number, which can lead to overfitting. Therefore, we decided to use the ASSR signals only after 513 ms because the change of tones gives the most impact on ASSR amplitudes. This leads us to 1026 features for each subject. This number still can be considered large; however, since we do not have reliable information about which features differ between illusion and nonillusion groups, in order to not lose important features, we decided to use 513 features for each left and right hemisphere—1026 in total for each subject.

2.4. Machine Learning

Machine learning is a number of algorithms aimed to learn data, find specific patterns according to task, and make decisions without human intervention. There are three main categories of machine learning: supervised machine learning, which is characterized by using labeled data to train the model; unsupervised machine learning, which is characterized by analyzing and clustering unlabeled datasets; and reinforcement learning, which is based on rewarding and punishing the model according to its desired behavior.

Since we have labeled data and we are interested in using the trained model with other similar data in the future, we used supervised machine learning in our study.

All algorithms require a set of related data to extract features that characterize the problem. The structure and quality of data is the most important factor for receiving reliable results of models' performance. The more various and clean data, the more accurate the performance will be. Moreover, for any given task, some specific models can show better results than others. There are no exact rules on how to choose the best model for a task. It is necessary to test several algorithms to find the one that gives the most accurate results.

In our study, we had data of only 17 samples, which did not provide enough variety of data. This is why we focused on using several simple models with L2 regularization to avoid overfitting and compared the results to find the most optimal one.

In this study, we compared the results of four main classification approaches: logistic regression, random forest, support vector machine, and convolutional neural network.

2.4.1. Logistic Regression

Logistic regression (LR) is a classification algorithm used to estimate binary values as true/false or 0/1 (Figure 5). The function quantifies the likelihood that a training sample point is correctly classified by the model. Therefore, the average for the entire training set indicates the probability that a random data point will be correctly classified by the system, regardless of the possible class. LR tries to maximize the mean of the data. For binary classification, the logistic regression model can be expressed as

$$P(y) = 1 - \frac{1}{1 + \exp(w\,x + b)} \tag{1}$$

where P is the probability, y is the outcome of interest, w is the weight, and b is a bias term [16]. LR is a simple, fast-training, feature-derivation classification method that gives good results on a small dataset with many features, which is the case in this study.

Since there is a literature gap in verified knowledge about octave illusion data features, in our study, to identify the dependencies between variables, we used a trial-and-error approach to choose the most accurate parameters for our model. The model with the following parameters showed the most accurate results for LR: gamma (inverse of the standard deviation) = 0.0001, C (inverse of regularization strength) = 1.0, L2 regularization.

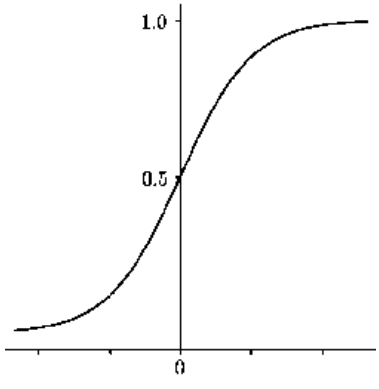


Figure 5. Logistic regression.

2.4.2. Random Forest

Random forest (RF) is an ensemble version of the decision tree algorithm. Each decision tree in the ensemble "votes" for a certain classification decision, and the prediction with the majority of voices "wins" (Figure 6). RF uses multiple trees to compute the majority of votes in the last leaf nodes to make a decision. Using decision trees, random forest models have resulted in significant improvements in prediction accuracy compared with a single tree by increasing the number of trees. In addition, each tree in the training set was randomly sampled without replacement. Each decision tree in the forest presents a simple structure in which the top node is considered the root of the tree that is recursively split at a series of decision nodes from the root until the decision node is reached [16]. Compared with other methods, RF is less prone to overfitting and works well with an automated feature interaction, making it a suitable method for classifying octave illusion data.

As with LR, to identify the dependencies between variables we used a trial-and-error approach to choose the most accurate parameters for our model. The model with the following parameters showed the most accurate results for RF: number estimators = 100, samples = 2, and the maximum number of features.

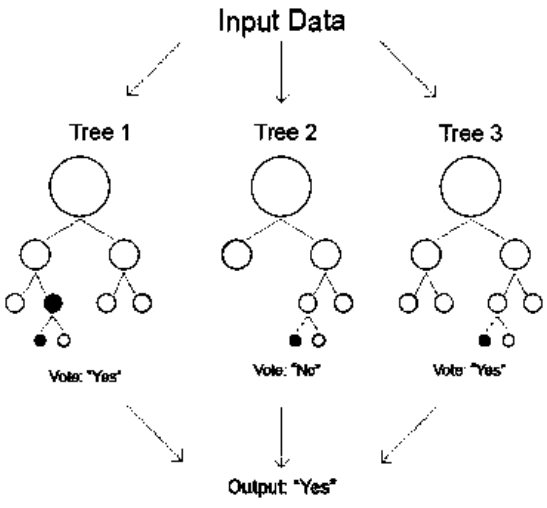


Figure 6. Random forest.

2.4.3. Support Vector Machine

Support vector machine (SVM) is a two-class classification method that finds the optimal linear hyperplane in the feature space that maximally separates the target classes, saving space for misclassification (Figure 7). The common formula for the linear classifier is:

$$f(x) = \sum_{i}^{n} \alpha_i k(x\ , x_i) + b \tag{2}$$

where α is the margin hyperplane, x and x_i are separable classes, k is a kernel function, b is a linear parameter, and $i = 1, 2, 3, \ldots, n$ [17]. There could be an infinite number of hyperplanes separating classes, but because it is a two-dimensional space, any hyperplane will always have one dimension, which can be represented by a simple regression line.

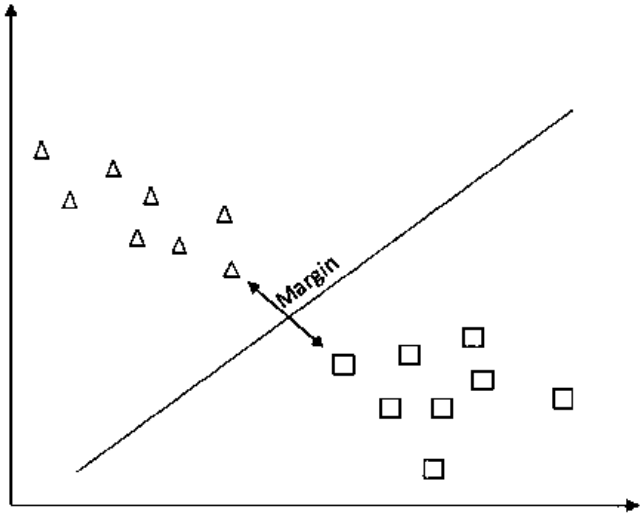


Figure 7. Support vector machine.

Although SVM, same as LR, shows good results on a small dataset with many features, unlike LR, it handles outliers better.

Along with the approach for LR and RF, we again used a trial-and-error approach to choose the most accurate parameters for our model. The model with the following parameters showed the most accurate results for SVM: gamma = 0.0001, C = 1.0, L2 regularization.

2.4.4. Convolutional Neural Network

Neural networks would be an effective method even for data with this high level of impurity; however, on a small dataset with many features, neural networks are more easily overfitted than other methods. Nevertheless, in order to explore possibilities of deep learning as well, we investigated the application of a convolutional neural network (CNN) for classification of octave illusion data. Although CNNs are commonly used for computer vision tasks, they have proved their efficiency in other fields, such as signal processing or medical applications. Furthermore, CNNs can be especially effective for biomedical data because they are tolerant to the input data transformations such as scaling or distortion and they can adapt to different input sizes [18].

Because our original dataset consists of 2D arrays of size 1026×2 (time/hemisphere), in order not to lose data, we could not use anything except 1D CNN structure, and to avoid overfitting, we used extremely simple CNN architecture of only three layers: convolutional, max-pooling, and fully connected (Table 1).

However, since it is impossible to establish exactly which features the CNN has used for its training and classification, in our study, we used the CNN as a tool to find that there is a difference in ASSR response between two groups and to compare its classification results with simpler methods of machine learning.

Table 1. CNN architecture.

Layer	Kernel Size	Filter	Stride
Convolutional	1×1	4	1
Max-Pooling	1×2	-	4
Fully Connected	-	-	-

2.5. Feature Selection

In machine learning, gathering a sufficient number of features is a vital requirement for classification tasks. However, increasing the number of features improves the classification abilities only to a certain point. This is called the curse of dimensionality. The curse of dimensionality is a common problem in machine learning caused by exponentially increasing errors with the number of features. A larger number of features requires a larger dataset, but because practically the number of training data is fixed, the classifier's performance will drop after the number of features reaches a certain point, depending on the size of the dataset [19]. Since in our study we have the dataset of only 17 subjects with 1026 features for each subject, using feature selection is necessary for effective classification of octave illusion and nonillusion data.

Feature selection should not be confused with feature extraction. Feature extraction creates a new set of features by mapping the original set of features. In contrast, feature selection takes a subset of the existing features without creating a new one. The overall feature selection process used in this study is shown in Figure 8.

In our study, we compared the results of four feature selection methods, which are univariate feature selection, recursive feature elimination, principal component analysis, and feature importance.

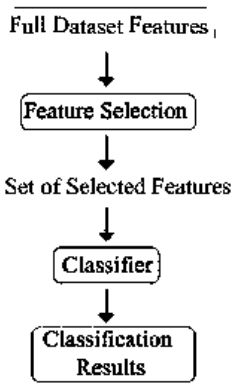


Figure 8. Feature selection process.

2.5.1. Univariate Feature Selection

Univariate feature selection (US) is a method of selecting features that contribute the most to the classification using univariate statistical tests. It returns a ranked list of features based on different statistical scoring functions. The main characteristic of a univariate approach is that it does not consider the dependencies between the features, and in the end, features of the dataset are independent of each other [20].

2.5.2. Recursive Feature Elimination

Recursive feature elimination (RFE) works by recursively removing values and uses the remaining attributes to build the model. First, the classifier is trained on the initial set of features, and the importance of each feature is written. The least important features are then cut from the features list. This procedure is recursively repeated until the desired number of quality features is reached [21].

2.5.3. Principal Component Analysis

Principal component analysis (PCA) chooses variables based on the magnitude (from largest to smallest absolute values) of their coefficients. PCA is fast and easy to implement, but it does not count the potential multivariate nature of the data structure, which leads to the loss of potentially valuable features [21].

2.5.4. Feature Importance

Feature importance (FI) (or variable importance) is a method to calculate scores for each feature for a given model. A feature is considered "important" if the accuracy of the model drops, and its change causes an increase in errors. However, a feature is considered "unimportant" if the shuffling of its values does not affect the accuracy of the model. There are several approaches to calculate the importance of features; in our study, we used an ensemble of decision trees (random forest) with mean decrease impurity. The algorithm randomly rearranges or shuffles one column of the validation dataset, leaving all other columns untouched [22]. It is a quick and easy-to-implement method with a tendency to prefer features with high cardinality, which is one of the important characteristics of our dataset.

3. Results

The main problem with using feature selection is its stochastic nature, which could lead to different results. To eliminate all possible ambiguities, each combination of machine learning and feature selection methods was run 10 times. Owing to the size of 17 of the entire dataset, we set the size of the training dataset to 11 (6 illusion and 5 nonillusion data), with the validation dataset of 4, and the test dataset to 6 (3 illusion and 3 nonillusion data)

(Table 2). Again, based on the relatively small size of the training dataset (11 data in total), we focused on choosing the appropriate number of features.

Table 2. Dataset.

Parameters	Dataset
Stimuli (low tone/high tone) (Hz)	400/800
Number of participants (ILL/non-ILL)	17 (9/8)
Training data (ILL/non-ILL)	11 (16/5)
Validation data (ILL/non-ILL)	6 (3/3)
Test data (ILL/non-ILL)	6 (3/3)

First, we ran LR, RF, and SVM without using feature selection. Since we decided to use data from the time period between 513 and 1026 ms, we have 1026 features of ASSR signals from both the left and right hemispheres for the dataset of 17 subjects. To provide an understanding of used features, the short list of original features is shown in Table 3. Using the trial-and-error approach, the best parameters for each classifier are as follows:

- LR: gamma (inverse of the standard deviation) = 0.0001, C (inverse of regularization strength) = 1.0, L2 regularization.
- RF: number estimators = 100, samples = 2, maximum number of features.
- SVM: gamma = 0.0001, C = 1.0, L2 regularization.

Table 3. Short list of original features.

Label	LH_513	LH_514	LH_515	RH_513	RH_514	RH_515
NILL	7.29×10^{-13}	4.68×10^{-13}	1.86×10^{-13}	6.84×10^{-12}	6.84×10^{-12}	6.47×10^{-12}
NILL	3.26×10^{-12}	3.37×10^{-12}	3.30×10^{-12}	2.60×10^{-12}	1.21×10^{-12}	2.63×10^{-13}
ILL	4.50×10^{-12}	3.45×10^{-12}	2.24×10^{-12}	8.76×10^{-12}	2.38×10^{-12}	4.73×10^{-12}
NILL	6.21×10^{-12}	5.19×10^{-12}	3.87×10^{-12}	4.05×10^{-13}	3.62×10^{-13}	1.12×10^{-12}
ILL	2.59×10^{-12}	1.89×10^{-12}	1.09×10^{-12}	4.45×10^{-13}	1.97×10^{-12}	3.38×10^{-12}
ILL	2.48×10^{-12}	3.09×10^{-12}	3.53×10^{-12}	1.82×10^{-12}	1.35×10^{-12}	8.09×10^{-13}
ILL	2.58×10^{-12}	3.28×10^{-12}	3.81×10^{-12}	3.81×10^{-12}	3.41×10^{-12}	2.82×10^{-12}
ILL	2.46×10^{-13}	4.62×10^{-13}	6.56×10^{-13}	1.45×10^{-12}	2.33×10^{-12}	3.07×10^{-12}
ILL	2.71×10^{-12}	2.56×10^{-12}	2.28×10^{-12}	3.17×10^{-12}	2.66×10^{-12}	2.01×10^{-12}
ILL	3.20×10^{-12}	4.44×10^{-12}	5.45×10^{-12}	1.35×10^{-12}	1.75×10^{-12}	2.05×10^{-12}
ILL	2.60×10^{-12}	1.98×10^{-12}	1.26×10^{-12}	4.98×10^{-12}	5.11×10^{-12}	4.97×10^{-12}
NILL	8.77×10^{-13}	1.50×10^{-12}	2.04×10^{-12}	6.66×10^{-13}	5.40×10^{-13}	3.94×10^{-13}
NILL	2.45×10^{-12}	2.41×10^{-12}	2.24×10^{-12}	3.74×10^{-12}	4.63×10^{-12}	5.28×10^{-12}
NILL	3.56×10^{-12}	4.02×10^{-12}	4.25×10^{-12}	5.04×10^{-12}	4.22×10^{-12}	3.16×10^{-12}
NILL	3.25×10^{-12}	3.24×10^{-12}	3.05×10^{-12}	3.96×10^{-12}	5.27×10^{-12}	6.29×10^{-12}
ILL	6.08×10^{-12}	6.07×10^{-12}	5.71×10^{-12}	9.50×10^{-12}	9.54×10^{-12}	9.08×10^{-12}
NILL	1.71×10^{-12}	1.60×10^{-12}	1.40×10^{-12}	1.09×10^{-12}	1.07×10^{-12}	1.01×10^{-12}

LH is left hemisphere; RH is right hemisphere; 513, 514, and 515 are time codes in ms; ILL is illusion data; NILL is nonillusion data.

The results for each algorithm are listed in Table 4.

Table 4. Classification results for machine learning and CNN methods.

Method	TP	TN	FP	FN
CNN	3	3	0	0
LR	2	1	2	1
RF	1	2	1	2
SVM	2	2	1	1

TP: true positive, TN: true negative, FP: false positive, FN: false negative.

Both LG and RF showed the same unsatisfactory results, with only 50% accuracy. In contrast, the SVM steadily showed 66.6% accuracy, which, considering the size of the dataset, can be called a satisfactory result.

In order to find the appropriate amount of features that will give stable classification results without losing too many data, we decided to remove the number of selected features gradually, by dividing it in half, and test the results. From the original dataset for machine learning of 1026, we stopped on datasets of 200, 30, and 40 features selected by each feature selection method. The results for 200 and 30 features, which showed some improvements, are presented in Tables 5–8. Those steps were selected as turning points to choose a direction for increasing or decreasing the number of features. The dataset of 40 selected features (Tables 9 and 10) showed the best and most stable results.

Table 5. Classification results for machine learning methods with US of 200 features.

Method	TP	TN	FP	FN
LR US/FI	2	1	2	1
RF US/FI	2	2	1	1
SVM US/FI	2	2	1	1

TP: true positive, TN: true negative, FP: false positive, FN: false negative.

Table 6. Classification results for machine learning methods with RFE/PCA/FI of 200 features.

Method	TP	TN	FP	FN
LR RFE/PCA	2	1	2	1
RF RFE/PCA	1	2	1	2
SVM RFE/PCA	2	2	1	1

TP: true positive, TN: true negative, FP: false positive, FN: false negative.

Table 7. Classification results for machine learning methods with US/FI of 30 features.

Method	TP	TN	FP	FN
LR US/FI	2	1	2	1
RF US/FI	2	2	1	1
SVM US/FI	2	2	1	1

TP: true positive, TN: true negative, FP: false positive, FN: false negative.

Table 8. Classification results for machine learning methods with RFE/PCA of 30 features.

Method	TP	TN	FP	FN
LR RFE/PCA	2	1	2	1
RF RFE/PCA	2	2	1	1
SVM RFE/PCA	2	2	1	1

TP: true positive, TN: true negative, FP: false positive, FN: false negative.

Table 9. Classification results for machine learning methods with US/FI of 40 features.

Method	TP	TN	FP	FN
LR RFE/PCA	2	1	2	1
RF RFE/PCA	2	2	1	1
SVM RFE/PCA	3	2	1	0

TP: true positive, TN: true negative, FP: false positive, FN: false negative.

Table 10. Classification results for machine learning methods with RFE/PCA of 40 features.

Method	TP	TN	FP	FN
LR RFE/PCA	2	1	2	1
RF RFE/PCA	2	2	1	1
SVM RFE/PCA	2	2	1	1

TP: true positive, TN: true negative, FP: false positive, FN: false negative.

The results of the univariate feature selection for the dataset of 200 features are shown in Table 5. The method showed no improvement in combinations with LR and SVM but showed better results of RF: 66.6% instead of 50%. The results for RFE, PCA, and FI are shown in Table 6. None of the methods showed any difference from the original dataset.

The results of the univariate feature selection and feature importance for the dataset of 30 features are listed in Table 7. The method showed no improvement in combinations with LR and SVM from the original dataset, and the same RF results of 66.6% as for the dataset of 200 features. The results for RFE and PCA are shown in Table 9. It showed better results again of RF 66.6% instead of 50%. The results for other methods stayed the same.

The results of the univariate feature selection and feature importance for the dataset of 40 are listed in Table 8. Both methods showed no improvement in combination with LR, same results with RF, and better results of SVM with 75% accuracy. Although both methods showed the same accuracy results, US proved to be a more stable approach owing to its constant selection of the same features. Values selected by FI were different for each run, which is an expected behavior considering its stochastic nature. However, since we are interested in defining features of ASSR that contribute to octave illusion classification, a large variety in selection is unsatisfying data.

The results for both RFE and PCA showed no difference for LG, RF, and SVM and are listed in Table 10. As in the case of applying feature importance, we faced the problem with different sets of selected features after every run. Because the selected features were different every time and did not match even once after 10 runs, we created 10 sets of selected features for each RFE and PCA and ran them multiple times to obtain reliable results. The results shown are the majority of accurate results (7 out of 10) for all methods.

The dataset of 40 features received by using the US, which showed the best classification results aside from CNN, almost entirely consists of data from the left hemisphere. Datasets created by using RFE, PCA, and FI consist more of data from the right hemisphere, but the majority is left for the left one. Timecodes of those features are scattered along the time axis, and it is difficult to make a statement about when the difference between illusion and nonillusion groups happens exactly, but we can say that it takes place in the left hemisphere.

Compared to using other machine learning methods with or without feature selection, applying deep CNN to the original dataset of features gave the best results of 100% accuracy, sensitivity, and specificity (Tables 9 and 11). This made the used CNN structure the most efficient tool to classify octave illusion and nonillusion data using ASSR, but it does not contribute to our knowledge about exactly which features of ASSR make the difference between the two groups. Since there are no false positive or false negative results and the training and validation losses are both small after epoch 10 (Figure 9), we can say that features were extracted successfully and no overfitting had happened.

Table 11. Overall classification results.

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)
CNN	100	100	100
LR	50	66.6	33.3
RF	50	33.3	66.6
SVM	66.6	66.6	66.6
LR US/FI (40 features)	50	66.6	33.3
RF US/FI (40 features)	66.6	66.6	66.6
SVM US/FI (40 features)	75	100	66.6
LR RFE/PCA (40 features)	50	66.6	33.3
RF RFE/PCA (40 features)	66.6	66.6	66.6
SVM RFE/PCA (40 features)	66.6	66.6	66.6

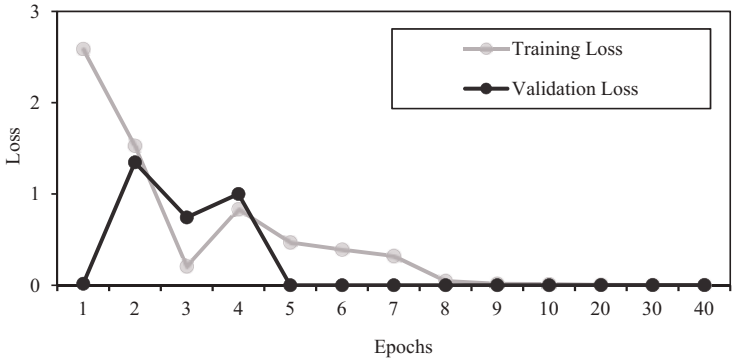


Figure 9. Training and validation losses.

4. Discussion

In this study, we aimed to find the most efficient combination of feature selection and machine learning methods for classifying octave illusion data. Machine learning has been widely used for the classification of various brain data, from classifying brain–computer interface (BCI) data to decoding MEG signal processing. ASSR signals are quite difficult to define owing to their small amplitudes and high levels of brain noise. The combination of SVM as a machine learning algorithm with univariate selection and feature importance as feature selection methods showed the highest classification results with 75% accuracy, 100% sensitivity, and 66.6% specificity (Table 10), which, considering the small size of the training dataset, are satisfactory results. Applying CNN gives even better results with 100% accuracy, 100% sensitivity, and 100% specificity, which makes it the best classification method (Figure 10).

However, considering the big picture, we are not interested in simple classification of illusion and nonillusion data, but in obtaining information about exactly which ASSR values differentiate those two groups to obtain a deeper understanding of the auditory illusion mechanism. Since the FI, due to its stochastic nature, gives various results and requires several runs to average it, and the US in its turn always presents the same results, the combination of SVM with the US is preferable.

Univariate selection is the only feature selection method that provides almost the same set of ASSR features in every run. Using this method, we received the set of features that most clearly define the difference between octave illusion and nonillusion groups, which will help in constructing the classification tool for octave illusion and nonillusion data. Since this set almost entirely consists of data from the left hemisphere, we suggest that for the right-handed group of people, the mechanisms that cause the octave illusion are lying there. However, since no dependencies were found between these features, at this

moment, it is difficult to define the pattern of auditory cortex activity for octave illusion and nonillusion groups.

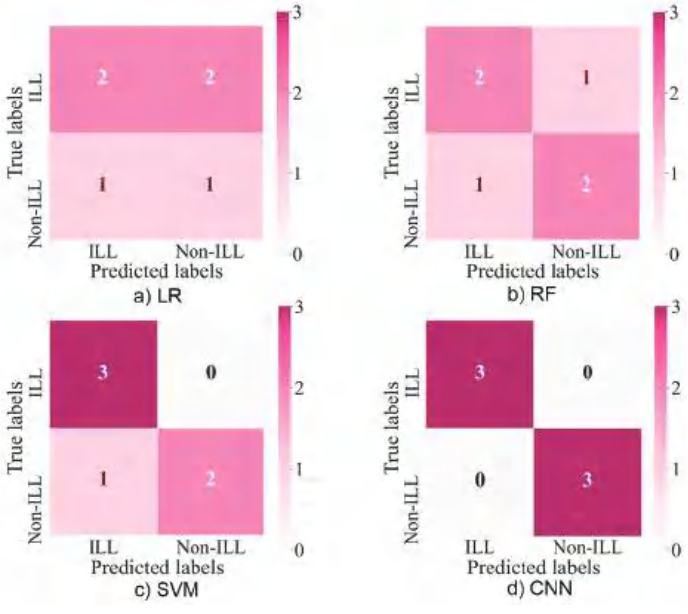


Figure 10. Confusion matrix for four methods.

In addition to using the developed machine learning methods for the classification of octave illusion and nonillusion data, information about selected features can be used to understand the underlying mechanisms of auditory illusions, which can contribute to managing mental diseases. In the future, we plan to build a universal tool for the classification of various types of auditory illusions based on the differences in the ASSR signals.

Author Contributions: Methodology, N.P.; Supervision, A.T.; Writing—review & editing, K.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted according to the Declaration of Helsinki and was approved by the Research Ethics Committee of Tokyo Denki University.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Zhmurov, V.A. *Psychopathology. Part I*, 2nd ed.; Irkutsk Publishing House State University: Irkutsk, Russia, 2011; Volume 1, pp. 19–20.
2. Tanaka, K.; Kurasaki, H.; Kuriki, S. Neural representation of octave illusion in the human cortex revealed with functional magnetic resonance imaging. *Hear. Res.* **2018**, *359*, 85–90. [CrossRef] [PubMed]
3. Deutsch, D.; Roll, P.L. "Separate "what" and "where" decision mechanisms in processing a dichotic tonal sequence. *J. Exp. Psychol. Hum. Percept. Perform.* **1976**, *2*, 23–29. [CrossRef] [PubMed]
4. Alomari, M.H.; Samaha, A.; AlKamha, K. Automated Classification of L/R Hand Movement EEG Signals using Advanced Feature Extraction and Machine Learning. *Int. J. Adv. Comput. Sci. Appl.* **2013**, *4*, 207–212.
5. Nakagome, S.; Luu, T.P.; He, Y.; Ravindran, A.S.; Contreras-Vidal, J.L. An empirical comparison of neural networks and machine learning algorithms for EEG gait decoding. *Sci. Rep.* **2020**, *10*, 1–17. [CrossRef] [PubMed]

6. Ko, W.; Jeon, E.; Jeong, S.; Phyo, J.; Suk, H.-I. A Survey on Deep Learning-Based Short/Zero-Calibration Approaches for EEG-Based Brain–Computer Interfaces. *Front. Hum. Neurosci.* **2021**, *15*, 258–280. [CrossRef] [PubMed]
7. Khosla, M.; Jamison, K.; Ngo, G.H.; Kuceyeski, A.; Sabuncu, M.R. Machine learning in resting-state fMRI analysis. *Magn. Reson. Imaging* **2019**, *64*, 101–121. [CrossRef] [PubMed]
8. Zubarev, I.; Zetter, R.; Halme, H.-L.; Parkkonen, L. Adaptive neural network classifier for decoding MEG signals. *NeuroImage* **2019**, *197*, 425–434. [CrossRef] [PubMed]
9. Hasasneh, A.; Kampel, N.; Sripad, P.; Shah, N.J.; Dammers, J. Deep Learning Approach for Automatic Classification of Ocular and Cardiac Artifacts in MEG Data. *J. Eng.* **2018**, *2018*, 1–10. [CrossRef]
10. Montilla-Trochez, D.; Salas, R.; Bertin, A.; Griskova-Bulanova, I.; Lisboa, P.; Saavedra, C. Convolutional neural network for cognitive task prediction from EEG's auditory steady state responses. Congress on Robotics and Neuroscience. In Proceedings of the 5th Congress on Robotics and Neuroscience, Valparaíso, Chile, 27–29 February 2020.
11. Hwang, E.; Han, H.-B.; Kim, J.Y.; Choi, J.H. High-density EEG of auditory steady-state responses during stimulation of basal forebrain parvalbumin neurons. *Sci. Data* **2020**, *7*, 288. [CrossRef] [PubMed]
12. O'Donnell, B.F.; Vohs, J.L.; Krishnan, G.P.; Rass, O.; Hetrick, W.P.; Morzorati, S.L. The auditory steady-state response (ASSR): A translational biomarker for schizophrenia. *Suppl. Clin. Neurophysiol.* **2013**, *62*, 101–112. [PubMed]
13. Rejer, I. EEG feature selection for BCI based on motor imaginary task. *Found. Comput. Decis. Sci. Dec.* **2012**, *37*, 283–292. [CrossRef]
14. Garali, I.; Adel, M.; Bourennane, S.; Guedj, E. Histogram-based features selection and volume of interest ranking for brain PET image classification. *IEEE J. Transl. Eng. Health Med.* **2018**, *6*, 99–112. [CrossRef] [PubMed]
15. Nanthini, B.S.; Santhi, B. Electroencephalogram signal classification for automated epileptic seizure detection using genetic algorithm. *J. Nat. Sci. Biol. Med.* **2017**, *8*, 159–166. [CrossRef] [PubMed]
16. Kirasich, K.; Smith, T.; Sadler, B. Random forest vs logistic regression: Binary classification for heterogeneous datasets. *SMU Data Sci. Rev.* **2018**, *1*, 9.
17. Hasan, M.A.M.; Nasser, M.; Pal, B.; Ahmad, S. Support vector machine and random forest modeling for Intrusion Detection System (IDS). *J. Intell. Learn. Syst. Appl.* **2014**, *6*, 45–52. [CrossRef]
18. Kiranyaz, S.; Ince, T.; Gabbouj, M. Real-time patient-specific ECG classification by 1-D convolutional neural networks. *IEEE Trans. Biomed. Eng.* **2016**, *63*, 664–675. [CrossRef] [PubMed]
19. Subho, M.R.H.; Chowdhury, M.R.; Chaki, D.; Islam, S.; Rahman, M.M. A univariate feature selection approach for finding key factors of Restaurant Businessssklearn default parameters. In Proceedings of the 2019 IEEE, Region 10 Symposium (TENSYMP), Kolkata, India, 7–9 June 2019; pp. 605–610.
20. Chen, X.; Jeong, J.C. Enhanced recursive feature elimination. In Proceedings of the Machine Learning and Applications. ICMLA 2007, Cincinnati, OH, USA, 13–15 December 2007; pp. 429–435.
21. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. *Int. J. Livest. Res.* **2017**, *7*, 60–78. [CrossRef]
22. Greenwell, B.M.; Boehmke, B.C. Variable Importance Plots—An Introduction to the vip Package. *R J.* **2020**, *12*, 343–366. [CrossRef]

sensors

MDPI

Article

IndoorCare: Low-Cost Elderly Activity Monitoring System through Image Processing

Daniel Fuentes [1], Luís Correia [1], Nuno Costa [1], Arsénio Reis [2], José Ribeiro [1], Carlos Rabadão [1], João Barroso [2] and António Pereira [1,3,*]

[1] Computer Science and Communication Research Centre, School of Technology and Management, Polytechnic Institute of Leiria, 2411-901 Leiria, Portugal; daniel.fuentes@ipleiria.pt (D.F.); luis.correia@ipleiria.pt (L.C.); nuno.costa@ipleiria.pt (N.C.); jose.ribeiro@ipleiria.pt (J.R.); carlos.rabadao@ipleiria.pt (C.R.)
[2] INESC TEC, University of Trás-os-Montes e Alto Douro, Quinta de Prados, 5001-801 Vila Real, Portugal; ars@utad.pt (A.R.); jbarroso@utad.pt (J.B.)
[3] INOV INESC Inovação, Institute of New Technologies, Leiria Office, Campus 2, Morro do Lena-Alto do Vieiro, Apartado 4163, 2411-901 Leiria, Portugal
* Correspondence: apereira@ipleiria.pt

Abstract: The Portuguese population is aging at an increasing rate, which introduces new problems, particularly in rural areas, where the population is small and widely spread throughout the territory. These people, mostly elderly, have low income and are often isolated and socially excluded. This work researches and proposes an affordable Ambient Assisted Living (AAL)-based solution to monitor the activities of elderly individuals, inside their homes, in a pervasive and non-intrusive way, while preserving their privacy. The solution uses a set of low-cost IoT sensor devices, computer vision algorithms and reasoning rules, to acquire data and recognize the activities performed by a subject inside a home. A conceptual architecture and a functional prototype were developed, the prototype being successfully tested in an environment similar to a real case scenario. The system and the underlying concept can be used as a building block for remote and distributed elderly care services, in which the elderly live autonomously in their homes, but have the attention of a caregiver when needed.

Keywords: computer vision; image analysis; internet of things; monitoring of elderly; low cost

Citation: Fuentes, D.; Correia, L.; Costa, N.; Reis, A.; Ribeiro, J.; Rabadão, C.; Barroso, J.; Pereira, A. IndoorCare: Low-Cost Elderly Activity Monitoring System through Image Processing. *Sensors* **2021**, *21*, 6051. https://doi.org/10.3390/s21186051

Academic Editor: Mario Munoz-Organero

Received: 28 July 2021
Accepted: 7 September 2021
Published: 9 September 2021

1. Introduction and Motivation

Portugal has an aging population with a tendency to increase [1], particularly, in rural areas, where the migration of the active population to large urban centers, in search of better job opportunities and quality of life, has led the remaining resident population in rural environments, mostly elderly, suffering social exclusion, and often ending up living in isolation.

As the economic factor is sometimes a barrier for technology adoption, especially by the elderly population with limited financial resources [2], cost-effective solutions to monitor the elderly and the isolated population in general have been researched for a while. The Ambient Assisted Living Joint Programme [3] is a European initiative and is an example of the needs that exist in support for the elderly. Since then, more and more research has been carried out [4], and every year new solutions appear make their contribution.

In this article, a low-cost solution for monitoring the movement of elderly people living alone in their homes, based on the AAL paradigm, is presented with the name IndoorCare. The system is based on a distributed architecture, where low-cost IoT devices acquire and process video images to export non-personal/private data through a gateway to a server. Then, the server aggregates all this information and makes it available, in a simple

way, to the user or caregiver. This proposed solution is based in technologies increasingly used in the area of smart everything [5] and provides a non-invasive monitoring system to a caregiver or a family member. The IndoorCare system records the person's physical movements over time and allows that information to be analyzed later by the caregiver to assess the person's health. One other advantage of a solution such as this is to allow the detection of any anomalous situations, such as emergencies or possible falls, in time.

The paper is organized as follows: Section 2 presents an overview of the related work; Section 3 describes the solution's architecture; Section 4 presents the prototype that was developed to validate the concept; Section 5 presents the system's evaluation and optimizations; and Section 6 presents the work's general conclusions.

2. Related Work

There are several solutions and projects focused on the detection, spatial location, and monitoring of the daily activity of people in indoor environments. In this section, we discuss some of the founding technologies for this type of solution, namely image analysis and infrared sensors.

There are several articles that analyze the AAL solutions that have been appearing in recent years [6,7] and some that expose the challenges that must be resolved in the future [8], especially in a post COVID-19 era [9]. There are also recent studies that focus on the analyses and comparison, in various ways, of the created applications and architectures of recent AAL solutions, exposing the trends in the solutions' implementation that most works have followed [10]. On a more practical level, there are several interesting implementations that have used various methods to monitor and interact with older people; these solutions typically use IoT devices to perform this monitoring [11], whether using sensors attached to the person [12], monitoring furniture [13], using video systems that analyze in real time what is happening [14,15], using face recognition to detect people and who they are [16], or even through the analyses of the sound [17]. The use of computer vision together with artificial intelligence is an increasingly common practice, using the best that these technologies allow to better monitor the elderly in their homes [18]. In addition, there is a growing need to transfer the information processing from data centers to the periphery of the systems, namely to the source where the data is acquired, to reduce the traffic sent by the equipment at the edge. This concept is called fog computing, and in addition to being a great advantage for computer vision solutions, it is also already being used in monitoring solutions for the elderly that use wearable devices, among others [19].

The extraction of information based on image analysis is a relatively recent topic that has enabled the development of technologies that allow the automatization of the information gathering process. Image recognition solutions, such as the Open-Source Computer Vision Library (OpenCV) [20], combined with ubiquitous computing, using microcomputers, such as Raspberry Pi [21] or Arduino [22], allow the creation of environments that can act intelligently, according to the information extracted and collected.

The OpenCV software library uses image analysis to recognize the various types of information in an image, such as: detection of hand gestures, as set out in [23]; human facial recognition [24], where in this particular paper [25] the authors implemented a prototype using the Java CV library [26], which analyzes the camera stream from a IP security camera and detects human presence; and recognition and extraction of vehicle registration information [27] or surveillance security systems [28], in which the authors coupled common web cameras to devices of small processing power, e.g., Raspberry Pi, which acquires the image from the camera and uses a cloud platform to process the image for movement detection.

The research in the area of human monitoring and human location has resulted in several interesting works, such as the one presented in [29], where the authors propose a system that uses two modes of monitoring, inside the residence (indoor) and outside the residence (outdoor). For the detection of the indoor position, the users must wear RFID (Radio-Frequency Identification) tags, which are detected and read whenever the

user enters a new division, similarly to the RFID tagging systems used in logistic solutions to track items. For the detection of the outdoor position, the user must wear a GPS (Global Positioning System) device for position tracking. The GPS mode (outdoor) is activated automatically whenever the user leaves the room three meters away.

In ref. [30], the authors used infrared (IR) sensors to calculate the number of people inside a building, installing sensors on the doors' tops to detect transit movement between rooms, so they could calculate how many people were in each division.

In the work developed in ref. [31], the authors propose a solution that addresses some of the problems enunciated in this work. A system is proposed that constantly monitors the security of a home. It uses several Raspberry Pi devices, connected to surveillance cameras, and uses the OpenCV library for image analysis. The system can detect various types of events, such as opening and closing doors and windows, movement in the rooms, and breaking windows.

Table 1 summarizes and explains why the solutions previously presented are considered interesting for the development of this solution.

Table 1. Comparison between the solutions presented in the related work section.

Reference	Case Study	Why Was Chosen
[23]	Hand gesture detection	Image data extraction using OpenCV
[24]	Video processing on Raspberry PI	OpenCV on Raspberry PI
[25]	Human facial recognition	OpenCV on Raspberry PI
[27]	Extraction of vehicle information	OpenCV on Raspberry PI
[28]	Surveillance system	Image acquisition on Raspberry PI
[29]	Indoor/outdoor person detection	Uses RFID tags to detect humans
[30]	Indoor human detection	Uses infrared sensor to detect humans
[31]	Indoor monitoring	Uses IoT devices with AI
IndoorCare	Indoor human monitoring	Uses IoT devices with Computer Vision

The work that identifies most of the requirements for this intended solution is ref. [31], mainly because of the image analysis using computer vision, artificial intelligence, and IoT devices. Although the solution works as intended, according to the authors, it requires too many processing resources from the IoT devices, which means that a robust IoT device must be used, thus increasing the solution's cost. The objective of this work is to develop a solution that can monitor a person's movements inside the house, in the various rooms, also using computer vision, as in ref. [31], using IoT devices, while keeping the cost reasonably low.

3. IndoorCare System Architecture

The solution IndoorCare, proposed in this article, is based on some principles used in other solutions, namely using only one device per room for human detection [29], detecting the presence of people through motion capture analysis [30], and using low-cost microcomputers to analyze and process the collected information [31]. The system has a distributed and multi-agent architecture [32], which implements the client–server model, having a gateway module to ensure information security, as presented in Figure 1.

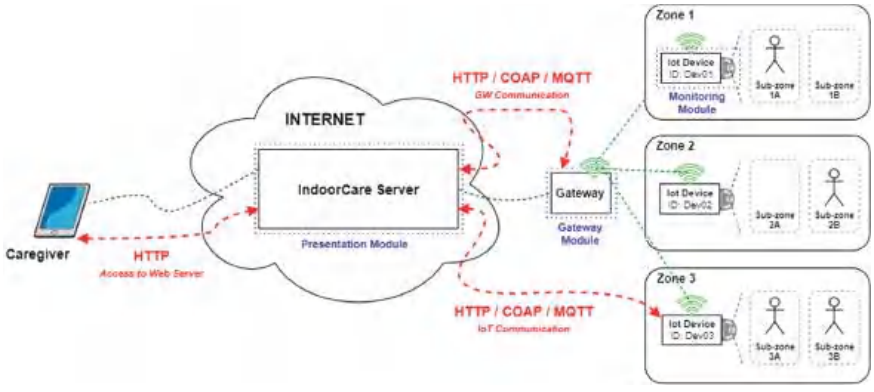


Figure 1. IndoorCare conceptual architecture.

The architecture comprises three modules, highlighted in Figure 1, where each module has a specific agent, a software-based entity, which performs the various tasks on the equipment to ensure proper operation of the system:

- The monitoring module, which translates to the several IoT devices at home that capture images and pre-process the data obtained from the images, in order to send this information to the presentation module.
- The gateway module, responsible for ensuring WiFi network availability and data communication security from the clients to the server (presentation module).
- The presentation module, which is the server that receives the data from the monitoring modules. It provides a presentation layer for the users (caregivers) to visually perceive the dynamics of the elderly person's activities inside the house over time.

The option to use image acquisition equipment and computer vision, instead of infrared sensors, although the latter in theory are cheaper, was because with computer vision it is possible to analyze several subzones within the same zone with a single IoT device, while with affordable infrared sensors, typically, one sensor is necessary to detect motion in each subzone. Although there are infrared sensors with this capability, such as the temperature detection cameras used to detect possible cases of COVID-19 [33], these are not low-cost IoT devices as their cost is in the range of thousands of euros per device.

3.1. Monitoring Module (IoT Device)

The IoT devices present in the elderly person's home are responsible for processing the information they acquire; namely, image analysis through OpenCV, generating processed data ready to be sent to the server. This functionality is in line with the Edge Computing concept [34], where the processing is executed close to the data source, in this case at the home of the elderly, being a paradigm increasingly used in the IoT universe and in smart systems.

The monitoring module works as a black box system that receives video feeds, and outputs the extracted data from the image analysis in the form of movement events. No other information is fed to the user/caregiver. This module encompasses the devices installed in the distinct areas (zones) of the elderly home to perform the movement detection. The software agent defined for this module and presented in Figure 2 is responsible for the image acquisition and hotspot calculation, using the available resources on the IoT device. A hotspot is defined as an area where movement is detected in the image, and for which the device must compare several image frames to be able to confirm if there is movement in a zone or not.

In Figure 2 are detailed all the components of each of the modules and how they communicate with each other.

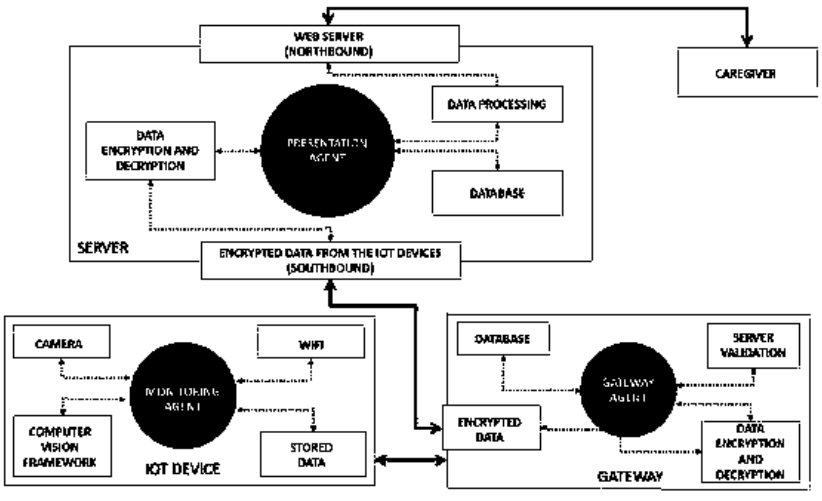


Figure 2. IndoorCare detailed architecture.

Each equipment has a unique identifier (ID) that identifies with which zone the device is associated. The ID is also used to identify the hotspot's information in the server's database and establish a relation between the device and its location in the house.

3.2. Gateway Module (Gateway)

This module is responsible for the confidentiality and integrity of the data transmission between the monitoring module and the presentation module. The module's agent, also presented in Figure 2, ensures that all communications go through an encrypted tunnel, from the network access point to the server. The agent guarantees the data encryption, as well as the server's address validation.

3.3. Presentation Module (Server)

This module acts as the server element of the client–server model in the communication and receives data from the clients, which are represented by the monitoring module. The data are saved, processed, and presented to the caregiver user. The software agent in this module, as presented in Figure 2, implements the features for data reception, decryption, saving, and presentation on a web platform. The user interaction is minimalistic and the agent basically combines the data streams from the several clients into a unique event feed.

To create the event stream from the data, the agent uses the subzones, as previously defined in the server's configuration, to check whether there is movement within them, using the hotspots sent by the clients. A subzone corresponds to a part of the image (the entire zone) acquired by a particular IoT device and corresponds to a specific area inside the elderly person's home.

This module provides a web portal for the caregiver to monitor the elderly and browse the daily activities inside the house, via Northbound [35] access. It also provides communication using Application Programming Interfaces (API) [36] via Southbound [35] for communication with the IoT devices and gateways.

3.4. Communication

As seen previously in Figure 1, there are three different types of communications:

- Caregiver with the server;
- Gateway with the server;
- IoT devices with the server.

In Northbound communication, between the caregiver and the server, as it typically occurs in a web environment (via the Internet), the most suitable communication protocol will be HTTP (Hypertext Transfer Protocol), in its secure version (HTTPS) [37]. This is one of the most used protocols for accessing online platforms and is widely used in the IoT environment for the same purpose.

In Southbound communication, between the server, gateways, and IoT devices, since it is a communication between IoT and network devices (if supported by the hardware), several protocols focused on the IoT environment can be used:

- HTTP (Hypertext Transfer Protocol): The most used client–server communication protocol on the Web which is also widely used in the IoT world due to its simplicity and efficiency in delivering information.
- COAP (Constrained Application Protocol): A communication protocol designed for devices that have limited processing capabilities, much like HTTP, but that uses much less data to send messages.
- MQTT (Message Queuing Telemetry Transport): One of the lightest communication protocols, it uses the Publisher/Subscriber model to exchange messages and is widely used in scenarios where network connectivity is not ideal.

These are just a few examples of communication protocols that can be implemented in this architecture, with HTTP still being one of the most used [38].

4. Implemented Prototype

In this section is presented the prototype developed to validate the proposed architecture. Low-cost IoT devices, widely used by the community, were used to implement the solution to validate the fulfilment of the objectives set out in the previous section.

In Figure 3, the general architecture of the prototype is illustrated, showing the modules and devices used.

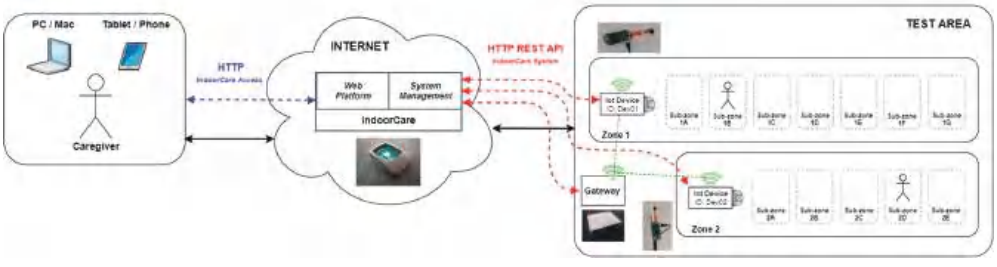


Figure 3. Prototype conceptual architecture.

The implemented prototype incorporates all the modules described in the architecture to demonstrate the intended functionality with the proposed system. Links at the right side are the interaction that takes place between the server, gateways and IoT devices (Southbound). The link at the left side represents the user/caregiver interaction with the web platform to access the IndoorCare system (Northbound). It should be noted that in this prototype, at the gateway level, only the basic static mechanisms were implemented for the system to work correctly, namely the VPN connection and server address validation.

4.1. Equipment Used

The equipment selection for this prototype project considered the costs in order to keep the solution effective and as low cost as possible, targeted at people with modest economic resources.

For the client IoT devices, we opted to use Single Board Computers (SBC) with an Operating System (OS) based in Linux, specifically the from the Raspberry PI family [39], due to its low price and good technical characteristics. To capture and analyze the images, we chose the Raspberry Pi Zero W [40] combined with a Fisheye 160° camera (including a

5 V 2.1 A power supply and Micro SD card), as presented in Figure 4. The total cost per device was around EUR 45 + VAT.

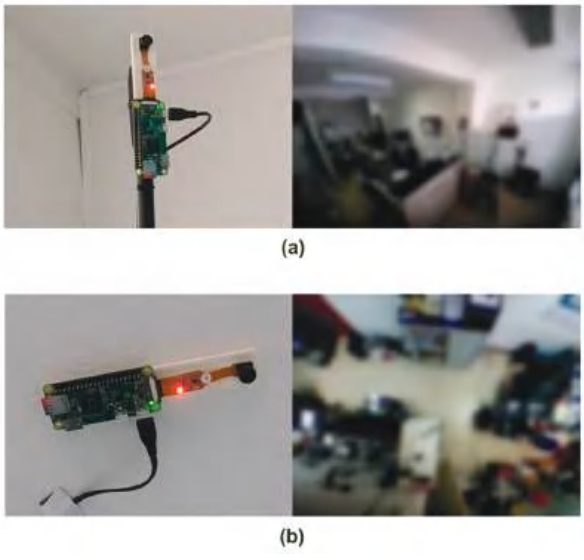


Figure 4. Raspberry PI Zero W #1 (**a**) and #2 (**b**).

This equipment is reasonably compact and has a set of ideal characteristics, such as a single-core processor at 1 GHz, 512 MB of RAM, and built-in WiFi, thus enabling the creation of client equipment capable of collecting, processing, and sending images to the system server over a WiFi network. It should be noted that initially the cameras used were normal Raspberry Pi Camera Modules [41], which were replaced by fisheye cameras only after testing the system, as described in Section 5.

For the server equipment, to receive, store, process, and present data on a web portal, we opted for the Raspberry Pi 3 B [42] (including a 5 V 2.1 A power supply and a Micro SD card). The total cost was around EUR 50 + VAT. The characteristics of this equipment, despite being an IoT device, meet the requirements for server equipment, as it has a quad-core processor at 1.4 GHz and 1 GB of RAM. Although this device has a reasonable performance, in a real or production environment, a more robust computational node should be used, namely a dedicated server (PC) or an online VPS (Virtual Private Server).

For network gateway equipment, we chose a Mikrotik Routerboard RB951Ui-2ND [43], mainly because of the possibility to create internal scripting for network management, and the ability of this scripting to communicate with platforms via REST API. This device had a total cost around EUR 30 + VAT. This equipment can work as a WiFi access point for the client devices, allowing the creation of a secure bridge Virtual Private Network (VPN) [44] between client and server equipment, thus ensuring client–server end-to-end confidentiality.

One major concern is the device intrusion that can lead to the visualization of the images captured by the camera by unauthorized persons. A way to guarantee the privacy of the residents is by physically blurring the lens of the equipment. Figure 5 shows the differences between a focused and an unfocused lens, and it is possible to notice that in the image with the lens out of focus, objects and people are not perceptible, thus ensuring the privacy required by the GDPR [45].

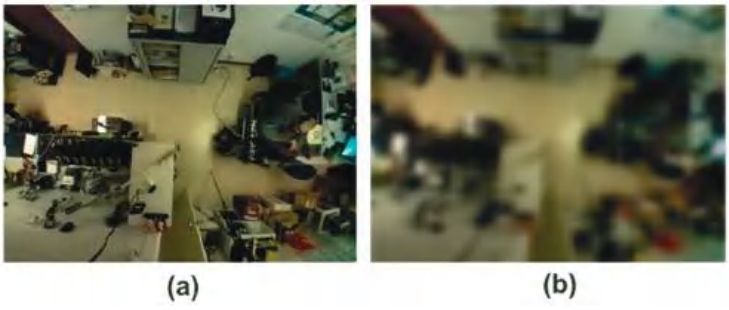


Figure 5. Focused (**a**) and unfocused (**b**) lens.

4.2. System Operation

To acquire the location of an individual person from video camera images, it is necessary to analyze and extract information from the images. We used the OpenCV software library [20] and the ImUtils library [46] to recognize movement in the images and Python [47] as the programming language for the software agent. It is possible to compare two images, one that serves as base reference for comparison and the other to check for changes, converting the captured images to arrays of pixels and comparing the different values of their respective positions. To perform the comparison, the absolute value in the subtraction of each of the respective pixels is obtained, thus creating an image that presents the differences found in the pixel array, which in this case shows the complete changes that occurred between the images. Then, a threshold is applied to the resultant image, by defining a change limit between pixels, where the pixels below the threshold are discarded and those above are saved, to create an image, commonly known as threshold, which contains only the pixels where there is a significant difference or, in this case, movement detection.

In Figure 6, on the right side, the threshold which corresponds to the movement detected on the left side of the image is displayed, with the movement area defined by a blue rectangle.

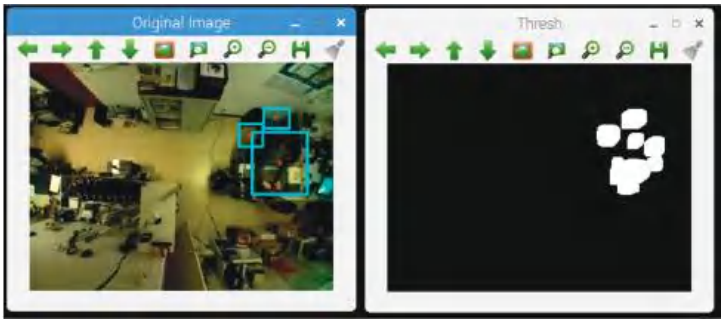


Figure 6. Motion detection in the image.

To effectively calculate the threshold, as stated in [24], it is necessary at an early stage to convert the color image to a grayscale image, so the only differentiating factor is the pixel brightness. Then, it is necessary to blur the image so that there are no sudden changes in the pixel tones. Figure 7 shows the different types of blurs supported by OpenCV: Gaussian Blur, Median Blur, and Normalized Block Blur [48].

Figure 7. Some types of image blurring.

Each blur type uses a different approach, yielding different results. The tests performed consisted of acquiring images where there was always the same human movement, walking from one end of the room to the other. Several threshold values were tested with the different types of blurs, which led to the following conclusion:

- Median Blur and Normalized Block give less false positives in the motion detection;
- Gaussian Blur detects more movement, as it provides more image detail after blurring.

In the prototype, Median Blur was used, but any other blur could be used as well.

Figure 8 describes the algorithm implemented by the Monitoring Agent to collect and compare images, find hotspots, and send them to the server (explained in the communication subsection). The device starts up and initially acquires an image to use as a base. In the following instant, the device acquires another image, and then creates a threshold for it. It analyzes if there is movement or not and if so, creates a hotspot entry and saves it into the log. Every 30 s the IoT device tries to upload all the hotspots it finds in that period of time.

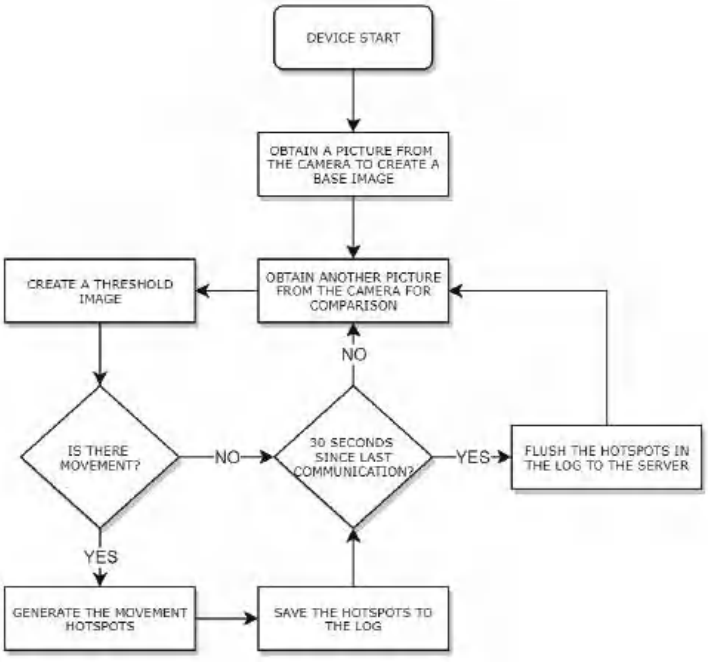


Figure 8. Client operation algorithm.

On the server side, the Presentation Agent receives and inserts the data into a database, after which it is processed and presented to the user/caregiver. To detect movement in an area, it is necessary to create sub-areas that will work as baselines for comparison with the detected hotspots by the devices. The server prototype provides the management feature to define and manage zones and subzones, as exemplified in Figure 9.

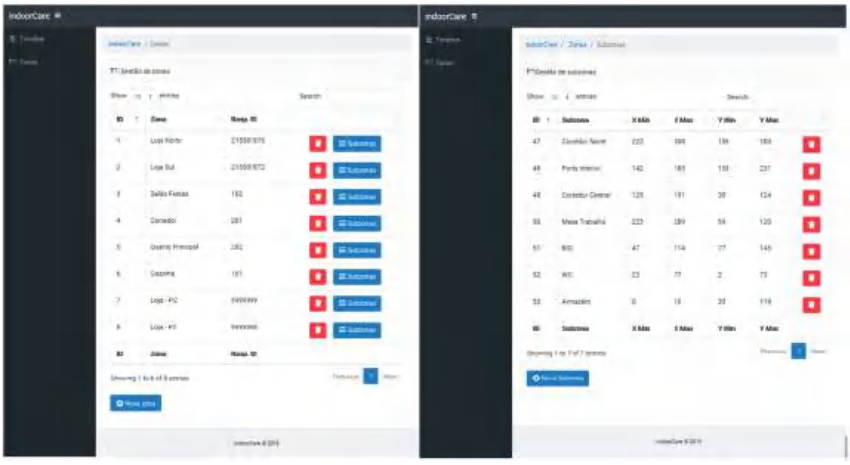


Figure 9. Management of zones and subzones.

This management feature allows the administrator/system installer to create the zones and the respective subzones that the caregiver want to supervise. It should be noted that to be able to acquire an image of the IoT device used as a monitoring module, a physical action on the equipment is required, namely the junction of two GPIO pins to activate the device's configuration mode. In Figure 10 are displayed two zones used in the system's prototype and its subzones, each one identified by an ID. Any hotspot detected within one of the delimited areas corresponds to movement in that subzone.

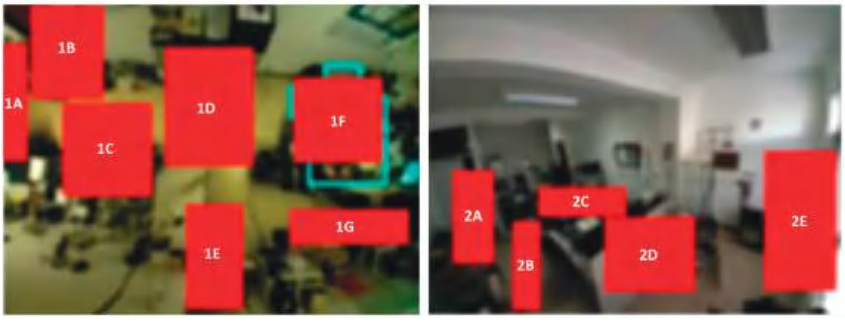


Figure 10. Two zones and several subzones defined for the prototype.

Following the hotspot detection, this information must be transmitted to the caregiver through a simple and effective interface, mainly because if the interface is too complex, the caregiver may not feel comfortable using it. An example of a simplistic visual interface is the timeline feed of events shown in Figure 11, which is a summary of the events that occurred each day at a certain time. The timeline provides a perception feed of the activity in the elderly home spaces under monitoring, by combining the feeds from the zones into a unique feed, formatted as a timeline grid. For the human caregiver/user, it is very simple and effective to check for specific events [49] and general activity in the house. In the timeline, each blue vertical stripe represents a hotspot detection in that respective subzone, signaling that movement was detected at that time in that area.

Figure 11. Event timeline.

With the timeline display, the caregiver can follow the daily life of the elderly and check his routine in a simple and non-intrusive way.

4.3. Communication

To be able to send the "converted images" transformed into data to the server, it is necessary for the client to be able to structure this information in such a way that it will be well interpreted at the destination. XML (Extensible Markup Language) [50] is a markup language that allows structuring information in a simple and easily readable way by human beings. It is one of the standards used in the communication of information between information systems and has great flexibility, allowing the creation of the most varied message structures. Another format also widely used in information communication is JSON (JavaScript Object Notation) [50], a compact message format that has less overhead than XML and has been also widely implemented in the industry.

In this prototype, we chose to use XML only because it allows easier reading of messages and facilitates the query of logs, but JSON could also be used. The messages sent in XML from the IoT devices to the server have the following fields:

- datetime: date and time of registration of hotspots, will be grouped in intervals of 30 s for better organization on the server.
- loggerid: the unique identifier of the IoT device that is collecting the information.
- framewidth: the original width of the image that generated the hotspot.
- frameheight: the original height of the image that generated the hotspot.
- matrixwidth: the scale of the matrix width used in this device (to normalize the different resolutions of different cameras).
- matrixheight: the height scale of the matrix used in this device (to normalize the different resolutions of different cameras).
- hotspot: with the x and y coordinates of a hotspot detected at that moment, there may be several at the same moment.

Figure 12 presents an example of one of these messages, sent periodically to the server.

```
<motion datetime”2021-01-06 16:45:00” loggerid="monitor01" framewidth="1024"
frameheight="600" matrixwidth="100" matrixheight="60">
        <hotspot>
                <x>450</x>
                <y>151</y>
        </hotspot>
        <hotspot>
                <x>461</x>
                <y>147</y>
        </hotspot>
        <hotspot>
                <x>468</x>
                <y>154</y>
        </hotspot>
</motion>
```

Figure 12. Example of an XML message used.

Client devices calculate hotspots and store this information in a log to be sent every 30 s. In case of communication failure, the Monitoring Agents themselves save the information that was not successfully sent to the server in the log and in the next iteration they try to resend all the pending information.

Initially, it was decided to encrypt the data using symmetric encryption on the clients, but this required unnecessary processing by the IoT devices, so the solution was to delegate this task to the gateway, which would be responsible for creating the VPN bridge with the server and ensure information security and confidentiality.

4.4. Movement Data History

One of the advantages of the way the system is designed and implemented is that there is a history of every hotspot detected, and so the processing of movement in the subzones is carried out in the server, and new subzones can be added or rearranged long after the system's first initialization.

In Figure 13 is presented an example of how this works. On the left side there are five subzones defined and all the hotspots registered since the system startup; on the right side there is a new subzone defined (2F) after the system initialization. Because a hotspot history exists for each zone, every movement that occurred in that subzone, even before its creation, can be fully visualized in the timeline.

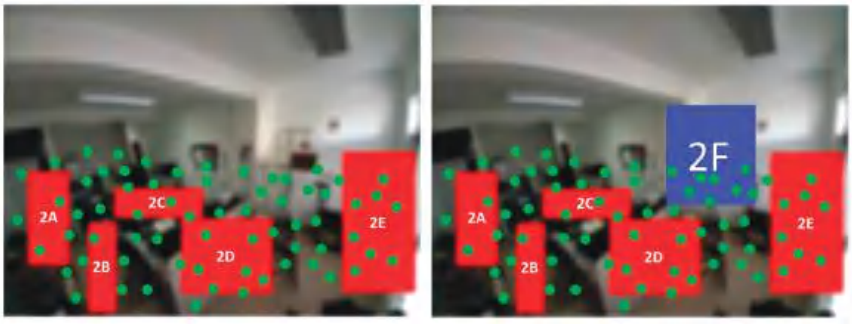


Figure 13. Hotspot's history.

Doing this allows the visualization of all the movement in the new/rearranged subzones, which were not contemplated in the system before, because all the data history related to the detected movement of the entire zone is saved.

5. Tests and Optimizations

Due to the current COVID-19 pandemic, it was not possible to carry out tests in real situations with the elderly. All tests performed were simulated in the same house division/area, with specific tests focused on the correct functioning of each module. During the tests, some optimizations were made, namely in the agent present in the IoT devices.

5.1. Client Testing

The tests performed on the IoT devices consisted of analyzing the code of the agent developed in python and its ability to perform the necessary operations, namely:

- Acquire images from a camera connected to the IoT device;
- Process the image using OpenCV for motion detection;
- Creation of hotspots for later upload to the server;
- Sending collected hotspots to the server.

In Figure 14 is shown the output of the Monitoring Agent, while in debug mode, displaying the image coordinates, where the movement was detected, and the XML message generated to be sent to the server.

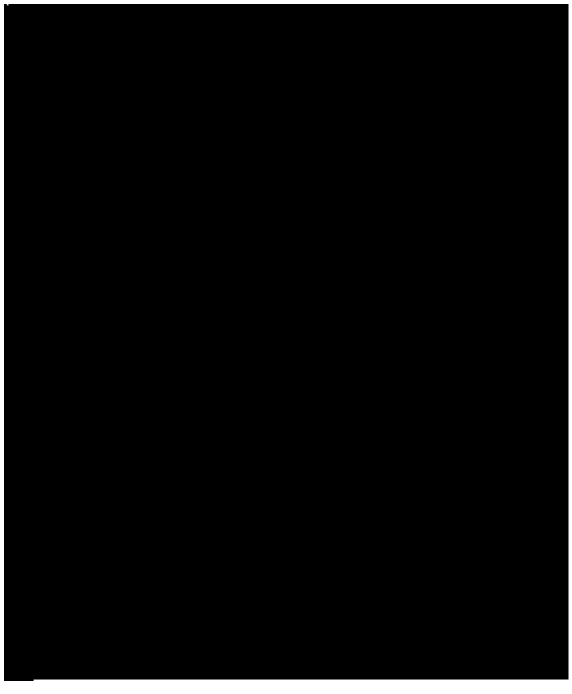


Figure 14. Monitoring Agent output in debug mode.

Another test involving the IoT devices was to verify if the hotspots generated by the devices were accurate or not; that is, if the motion detected by the devices was consistent with the motion points that appeared on the server's timeline. In Figure 15 is shown the timeline of the event stream, as displayed by the server, showing movement detection, while on the right side of the figure are shown the outputs of the equipment and the zones they are monitoring.

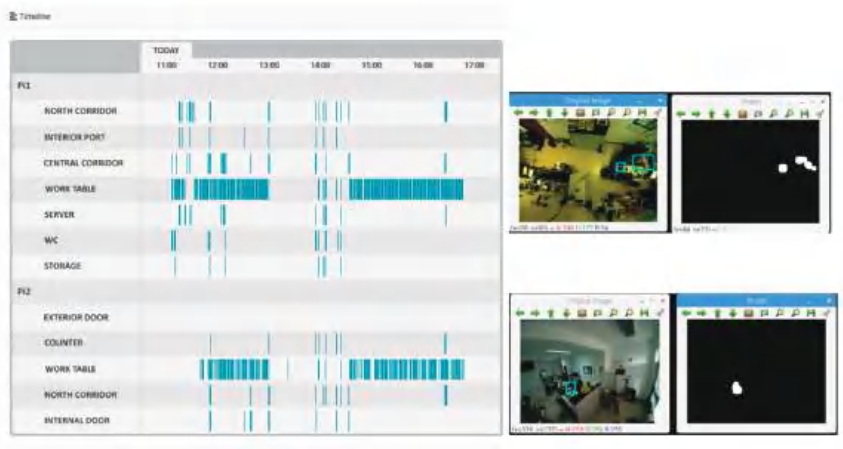


Figure 15. Timeline of events and the respective devices.

The tests performed on the system confirmed that the points collected by the device and the movements present in the timeline matched.

5.2. Server Testing

The tests carried out on the server focused on the reception and processing of data from the servers and their presentation to the user/caregiver, including:

- Reception of the hotspot in the IoT devices;
- Hotspot data processing and timeline generation;
- Creation and editing of zones and subzones.

In Figure 16 is shown an example of how the subzone creation tool of the server was tested; this was accomplished by creating subzones with specific x and y limits and by sending static hotspots generated manually on the IoT device with the corners of the subzone, to verify that the server was placing the hotspot point in the correct pixels on the image and thus generating movement correctly in that subzone.

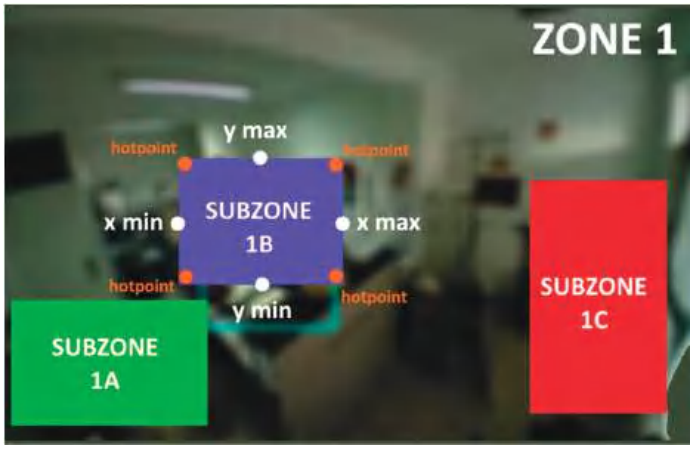


Figure 16. Subzone creation testing.

This test, in particular, served to verify if there was any deviation in the hotspot calculation due to the scale applied to the different image sizes of different types of cameras.

Different image capture resolutions were used to see if the same movement coincided in the same subzone, a result that was confirmed at the end.

5.3. Timeline Interpretation

The timeline event feed is a key element of the system, for which were conducted some tests to verify if an ordinary person (after very brief training) could understand the information, as presented, and perceive the events that might have generated those data. Due to the current pandemic situation, the tests were executed with only five persons simulating caregivers.

A test protocol was designed under which the subjects received a hypothetical timeline of the event feed of a day in a hypothetical house. While visualizing the timeline, the subjects were questioned about what they perceived had happened in the house during the day. In Figure 17 is shown a timeline, created for testing purposes only, in which specific numbered points correspond to specific events.

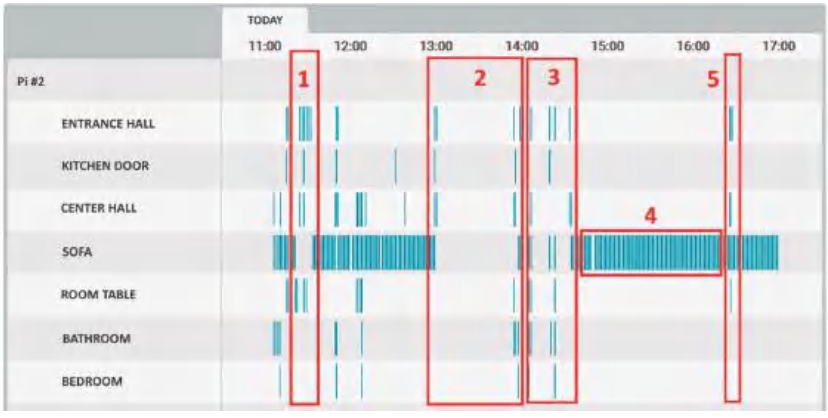


Figure 17. Timeline interpretation.

The events were then presented, but not numbered, and the subjects had to match the event with the event number on the timeline. These events were as follows:

- "*Mr. João spent the morning watching television on the sofa and then went to lunch.*";
- "*Mr. João went to drink water in the kitchen.*";
- "*Mr. João went to the bathroom.*";
- "*Mr. João was watching TV for almost 2 h.*";
- "*Someone knocked on the door and Mr. João went to see who it was.*".

The results, with a test group of five individuals, are quite positive, with all the individuals confirming that they were able to perceive what happened by reading the timeline. The only exceptions were events 1 and 5, which are very similar in the timeline, and two of the five individuals misinterpreted these two.

5.4. Hotspot Detection Optimization

During the tests on IoT devices, it was noticed that when using an outline rectangle for the movement detection, as the example in Figure 18 shows, the calculation of hotspots sometimes covered two or more subzones, leading to quite a few false positives in subzones where the movement was not happening.

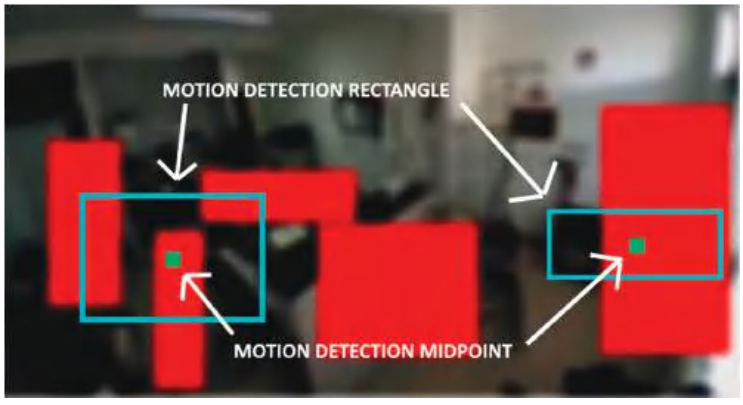


Figure 18. Movement detection optimization.

To reduce the number of false positives, it was decided to create a central point in the motion detection rectangle that would represent the midpoint of all the movement that occurred in that specific area of the image. By performing this optimization, and after several tests, it was concluded that when using this midpoint technique, the number of false positives decreased significantly, creating a timeline with much less scattered movement points.

Another advantage of this optimization was the significant decrease (about 50%) in the network traffic to send a hotspot data message, mainly because the messages are in XML and the overhead becomes much smaller, when, in this case, only a pair of X + Y coordinates are transmitted per movement, instead of the two pairs of coordinates to send the rectangle.

5.5. Automatic Background Adaptation

The analysis of whether there is movement or not is performed by comparing an image with a previous image (base image), in order to verify if there are differences between them. There is a problem when the base image no longer corresponds to the actual scenario, and small changes were introduced due to non-motion pixel changes. These changes, although not caused by motion, were detected as motion because the pixels in the image changed since the base image.

One of the identified problems was the constant change in the environment that the client was analyzing, either due to the presence of new objects or changes in lighting scenario, such as a light being turned on. To address this problem, in the IoT device was implemented a compensation algorithm that modified the baseline image when required. In Figure 19, two of the problems encountered while testing the system are presented. The first one is when an object enters the background scenario, and it does not exist in the base image. The second is when the scenario lighting changes, and all the pixels change, creating movement in the entire image.

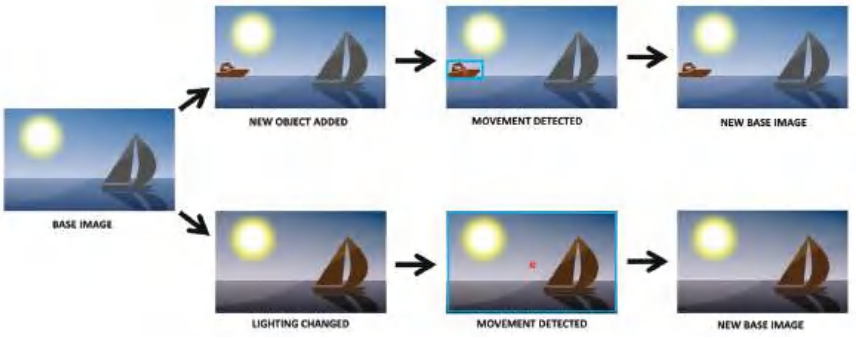


Figure 19. Creation of the new base image.

After several tests, an algorithm was created based on the work of [51], to optimize movement detection, that changed the baseline image based on two conditions:

- When no significant movement is detected for more than X seconds (X is variable), which allows the baseline image to be updated to ambient lighting throughout the day.
- When significant movement is detected for more than Y seconds (Y is variable), which happens in at least three different cases: when there is a sudden change in the ambient lighting, when new objects are introduced in the scenario, and when there is real movement in the image.

Regarding the second assumption, when a new base image is created, and if the movement detection continues, then it is because there is real movement detected. In case of a change in lighting or a new object in the scene, after the new base image is created the movement stops.

5.6. Cameras with Fisheye Lens

During the initial development of the system, cameras with regular lenses were used, with an about 72° viewing angle, which greatly limited the area to be monitored, especially if viewed from above (ceiling of the room). Later, fisheye lenses, with an approximately 160° viewing angle, were installed on the IoT devices, allowing a much larger monitoring area. In Figure 20, there is a lens comparison with normal lenses (72°) on the left and fisheye lenses (160°) on the right.

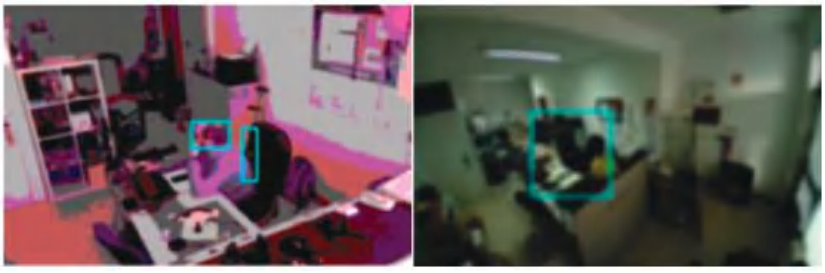


Figure 20. Angle viewing of lens: 72° (**Left**) vs. 160° (**Right**).

The equipment is positioned exactly in the same place and the only difference is the camera lens. On the right side, the amount of area is much bigger, which allows a better use of the image for motion detection. With this type of lens, it is also possible to place the equipment on the ceiling of the room and monitor the entire area, as seen in Figure 8.

With this optimization, and after several tests, the fisheye lens installed in the ceiling proved to be the ideal place to position the IoT device and monitor the room. This setup also produced fewer false positives in the movement detection in the subzones, mainly because the line of sight between the camera and the subzone is less likely to be obstructed.

5.7. Movement Detection Performance

To test the system's performance in real motion detection, various sizes of minimum motion area were tested to check how this would influence the system's execution. Figure 21 shows the test scenario with a defined sub-area (a) to detect the movement of the door (b). Three sizes of minimum area were used to detect the same movement of the door (c), which were 20 px, 50 px, and 100 px. These sizes were defined in particular for this test and another set could be defined; the purpose was only to check whether the minimum detection area influenced the detection of the same movement.

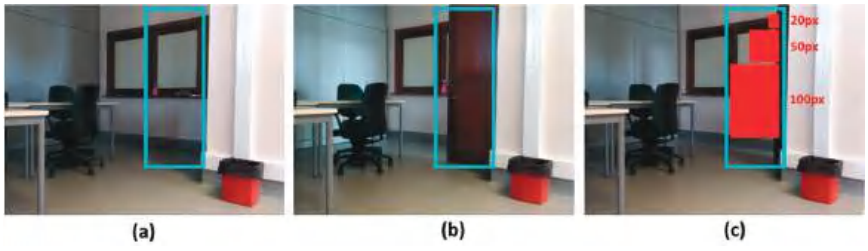


Figure 21. Movement detection of a closed (**a**) and open door (**b**) and the used movement area sizes (**c**).

The graph in Figure 22 shows the results obtained from the tests carried out. As can be seen, when using a smaller motion detection area, the system can recognize the same motion/movement in that subzone more often than when the area is larger.

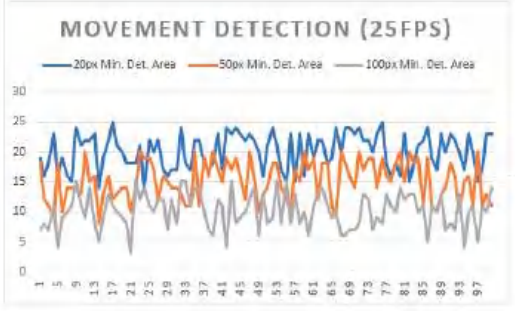


Figure 22. Movement detection with multiple minimum area sizes.

This result is expected and is reflected in previous tests, since the detection of the changed pixels is sometimes very fragmented, creating several small detection areas that are discarded if the minimum size of these to be considered valid is too high.

6. Conclusions and Future Work

In this work, an effective and low-cost indoor monitoring system was proposed to help caregivers take care of the elderly by monitoring their daily lives from a distance. This system brings the advantages of knowing where the elderly person is and the activity dynamics in the house, while fully respecting the elderly person's privacy, thus creating a daily movement record of the elderly person.

The test results of the prototype show that it is possible to use low-price and low-performance IoT equipment, namely a Raspberry Pi Zero W, to build a system that performs monitoring in a specific zone in the house and its associated subzones. In addition, the tests also indicate that the usage of the timeline event feed model is very effective to display the activity inside a home and that it is very simple to interact with.

As future work, an optimization that can be made in terms of processing on the Raspberry Pi is the implementation of gray areas, i.e., areas that will never have points of interest

for motion detection. The detections that happen in these subzones (e.g., reflections) are not sent to the server, leading to less false positives and a smaller amount of information to be sent to the server. Other improvements in this type of system include the implementation of automatic alerts, which could be of two types: a "Non-movement alert", which would inform the caregiver when something abnormal happens in the elderly's routine; or a "Too long alert", which would be used to inform the caregiver that, after movement was detected in an specific area, it suddenly stopped, informing the caregiver that something may be happening.

Author Contributions: Conceptualization, D.F., L.C. and A.P.; data curation, D.F. and L.C.; formal analysis, A.R., C.R., J.B. and A.P.; funding acquisition, A.R., J.B., C.R. and A.P.; investigation, D.F. and L.C.; methodology, A.P.; resources, A.R., C.R., J.B., J.R., N.C. and A.P.; software, D.F. and L.C.; supervision, A.R., J.B., N.C. and A.P.; validation, N.C., A.R., J.R., C.R., J.B. and A.P.; writing—original draft, D.F., L.C., N.C., A.R. and A.P.; writing—review and editing, D.F., L.C., J.R., A.R., C.R., J.B. and A.P. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Project "Digitalization of end-of-line distributed testers for antennas ("D-EoL-TA)", operation number: POCI-01-0247-FEDER-049698, financed by the Program COMPETE 2020, Portugal 2020, by National Funds through the Portuguese funding agency, FCT-Fundação para a Ciência e a Tecnologia, within project UIDB/04524/2020, and was partially supported by Portuguese National funds through FITEC-Programa Interface, with reference CIT "INOV-INESC Inovação-Financiamento Base" and by Portuguese Fundação para a Ciência e a Tecnologia-FCT, I.P., under the project UIDB/50014/2020.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors acknowledge the Computer Science and Communication Research Center for the facilities granted in the implementation of part of this work, in the context of the Smart IoT Ecosystems research line and the Mobile Computing Laboratory of the School of Technology and Management of the Polytechnic of Leiria. The authors also acknowledge the authorship of some of the images used in some of the visual content created using the tool "diagrams.net, accessed on 20 January 2021" and the free content available in "iconfinder.com, accessed on 20 January 2021", " pixabay.com, accessed on 20 January 2021", and "flaticon.com, accessed on 20 January 2021".

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Instituto Nacional de Estatísticas. Projeções de População Residente em Portugal. Portal do INE. Available online: https://www.ine.pt/xpotal/xmain?xpid=INE&xpgid=ine_destaques&DESTAQUESdest_boui=277695619&DESTAQUESmodo=2 (accessed on 17 May 2021).
2. Pereirinha, J.A.; Pereira, E. *Défice Social e Pobreza Relativa: Uma análise da Adequação Do Bem-Estar e da Segurança Económica em Portugal*; ISEG–GHES: Lisboa, Portugal, 2019.
3. Vodjdani, N. The ambient assisted living joint programme. In Proceedings of the 2008 2nd Electronics System-Integration Technology Conference, The University of Greenwich, London, UK, 1–4 September 2008; pp. 1–2.
4. Pinazo-Hernandis, S.; Puente, R.P. Innovación para el envejecimiento activo en la unión europea. Análisis del programa ambient assisted living joint programme (AAL) en el periodo 2008-2015. *Búsqueda* **2015**, *2*, 38–50. [CrossRef]
5. Streitz, N. Beyond 'smart-only'cities: Redefining the 'smart-everything'paradigm. *J. Ambient. Intell. Humaniz. Comput.* **2019**, *10*, 791–812. [CrossRef]
6. Abtoy, A.; Touhafi, A.; Tahiri, A. Ambient Assisted living system's models and architectures: A survey of the state of the art. *J. King Saud Univ. -Comput. Inf. Sci.* **2020**, *32*, 1–10.
7. Alkhomsan, M.N.; Hossain, M.A.; Rahman, S.M.M.; Masud, M. Situation awareness in ambient assisted living for smart healthcare. *IEEE Access* **2017**, *5*, 20716–20725. [CrossRef]
8. Wan, J.; Gu, X.; Chen, L.; Wang, J. Internet of things for ambient assisted living: Challenges and future opportunities. In Proceedings of the 2017 International conference on cyber-enabled distributed computing and knowledge discovery (CyberC), Nanjing, China, 12–14 October 2017; pp. 354–357.
9. Vimarlund, V.; Borycki, E.M.; Kushniruk, A.W.; Avenberg, K. Ambient assisted living: Identifying new challenges and needs for digital technologies and service innovation. *Yearb. Med. Inform.* **2021**, *30*, 141–149. [PubMed]

10. Grguric, A.; Khan, O.; Ortega-Gil, A.; Markakis, E.; Pozdniakov, K.; Kloukinas, C.; Medrano-Gil, A.; Gaeta, E.; Fico, G.; Koloutsou, K. Reference Architectures, Platforms, and Pilots for European Smart and Healthy Living—Analysis and Comparison. *Electronics* **2021**, *10*, 1616. [CrossRef]
11. Maskeliūnas, R.; Damaševičius, R.; Segal, S. A review of internet of things technologies for ambient assisted living environments. *Future Internet* **2019**, *11*, 259. [CrossRef]
12. Correia, L.; Fuentes, D.; Ribeiro, J.; Costa, N.; Reis, A.; Rabadão, C.; Barroso, J.; Pereira, A. Usability of Smartbands by the Elderly Population in the Context of Ambient Assisted Living Applications. *Electronics* **2021**, *10*, 1617. [CrossRef]
13. Bleda, A.L.; Fernández-Luque, F.J.; Rosa, A.; Zapata, J.; Maestre, R. Smart sensory furniture based on WSN for ambient assisted living. *IEEE Sens. J.* **2017**, *17*, 5626–5636. [CrossRef]
14. Malekmohamadi, H.; Moemeni, A.; Orun, A.; Purohit, J.K. Low-cost automatic ambient assisted living system. In Proceedings of the 2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), Athens, Greece, 19–23 March 2018; pp. 693–697.
15. Stefan, I.; Aldea, C.L.; Nechifor, C.S. Web platform architecture for ambient assisted living. *J. Ambient Intell. Smart Environ.* **2018**, *10*, 35–47. [CrossRef]
16. Almeidaa, A.; Costaa, R.; Limaa, L.; Novaisc, P. Non-obstructive authentication in AAL environments. In *Workshop Proceedings of the 7th International Conference on Intelligent Environments*; IOS Press: Nottingham, UK, 2011; pp. 63–73.
17. Navarro, J.; Vidaña-Vila, E.; Alsina-Pagès, R.M.; Hervás, M. Real-time distributed architecture for remote acoustic elderly monitoring in residential-scale ambient assisted living scenarios. *Sensors* **2018**, *18*, 2492. [CrossRef]
18. Colantonio, S.; Coppini, G.; Giorgi, D.; Morales, M.A.; Pascali, M.A. Computer vision for ambient assisted living: Monitoring systems for personalized healthcare and wellness that are robust in the real world and accepted by users, carers, and society. In *Computer Vision for Assistive Healthcare*; Academic Press—Elsevier Ltd.: Amsterdam, The Netherlands, 2018; pp. 147–182.
19. Nikoloudakis, Y.; Panagiotakis, S.; Markakis, E.; Pallis, E.; Mastorakis, G.; Mavromoustakis, C.X.; Dobre, C. A fog-based emergency system for smart enhanced living environments. *IEEE Cloud Comput.* **2016**, *3*, 54–62. [CrossRef]
20. OpenCV Team. Available online: https://opencv.org/ (accessed on 20 January 2021).
21. The Raspberry Pi Foundation. Raspberry Pi. Available online: https://www.raspberrypi.org/ (accessed on 20 January 2021).
22. Arduino. Available online: https://www.arduino.cc/ (accessed on 20 January 2021).
23. Khan, M.B.; Mishra, K.; Qadeer, M.A. Gesture recognition using Open-CV. In Proceedings of the 2017 7th International Conference on Communication Systems and Network Technologies (CSNT), Nagpur, India, 11–13 November 2017; pp. 167–171.
24. Arva, G.; Fryza, T. Embedded video processing on Raspberry Pi. In Proceedings of the 2017 27th International Conference Radioelektronika (RADIOELEKTRONIKA), Brno, Czech Republic, 19–20 April 2017; pp. 1–4.
25. Okabe, R.K.; Carro, S.A. Reconhecimento Facial Em Imagens Capturadas Por Câmeras Digitais De Rede. *Colloq. Exactarum* **2015**, *7*, 106–119. [CrossRef]
26. Audet, S. JavaCV. Available online: https://github.com/bytedeco/javacv (accessed on 20 January 2021).
27. Suryatali, A.; Dharmadhikari, V.B. Computer vision based vehicle detection for toll collection system using embedded Linux. In Proceedings of the 2015 International Conference on Circuits, Power and Computing Technologies [ICCPCT-2015], Nagercoil, India, 19–20 March 2015; pp. 1–7.
28. Patil, N.; Ambatkar, S.; Kakde, S. IoT based smart surveillance security system using raspberry Pi. In Proceedings of the 2017 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 6–8 April 2017; pp. 344–348.
29. Hutabarat, D.P.; Hendry, H.; Pranoto, J.A.; Kurniawan, A. Human tracking in certain indoor and outdoor area by combining the use of RFID and GPS. In Proceedings of the 2016 IEEE Asia Pacific Conference on Wireless and Mobile (APWiMob), Wireless and Mobile (APWiMob), Bandung, Indonesia, 13–15 September 2016; pp. 59–62.
30. Perra, C.; Kumar, A.; Losito, M.; Pirino, P.; Moradpour, M.; Gatto, G. Monitoring Indoor People Presence in Buildings Using Low-Cost Infrared Sensor Array in Doorways. *Sensors* **2021**, *21*, 4062. [CrossRef]
31. Kumar, K.N.K.; Natraj, H.; Jacob, T.P. Motion activated security camera using raspberry Pi. In Proceedings of the 2017 International Conference on Communication and Signal Processing (ICCSP), Chennai, India, 6–8 April 2017; pp. 1598–1601.
32. El-Sayed, H.; Sankar, S.; Prasad, M.; Puthal, D.; Gupta, A.; Mohanty, M.; Lin, C.T. Edge of things: The big picture on the integration of edge, IoT and the cloud in a distributed computing environment. *IEEE Access* **2017**, *6*, 1706–1717. [CrossRef]
33. Al-Humairi, S.N.S.; Kamal, A.A.A. Opportunities and challenges for the building monitoring systems in the age-pandemic of COVID-19: Review and prospects. *Innov. Infrastruct. Solut.* **2021**, *6*, 1–10. [CrossRef]
34. Khan, W.Z.; Ahmed, E.; Hakak, S.; Yaqoob, I.; Ahmed, A. Edge computing: A survey. *Future Gener. Comput. Syst.* **2019**, *97*, 219–235. [CrossRef]
35. Shin, M.K.; Nam, K.H.; Kim, H.J. Software-defined networking (SDN): A reference architecture and open APIs. In Proceedings of the 2012 International Conference on ICT Convergence (ICTC), Jeju, Korea, 15–17 October 2012; pp. 360–361.
36. Cwalina, K.J.; Abrams, B.M.; Moore, A.J.; Anderson, C.L.; Pizzo, M.; Brigham, I.R.A. Design of Application Programming Interfaces (APIs). U.S. Patent No. 7,430,732, 30 September 2008.
37. Naylor, D.; Finamore, A.; Leontiadis, I.; Grunenberger, Y.; Mellia, M.; Munafò, M.; Papagiannaki, K.; Steenkiste, P. The cost of the "s" in https. In Proceedings of the 10th ACM International on Conference on emerging Networking Experiments and Technologies, Sidney, Australia, 2 December 2014; pp. 133–140.

38. Dizdarević, J.; Carpio, F.; Jukan, A.; Masip-Bruin, X. A survey of communication protocols for internet of things and related challenges of fog and cloud computing integration. *ACM Comput. Surv. (CSUR)* **2019**, *51*, 1–29. [CrossRef]
39. The Raspberry Pi Foundation. Raspberry Pi Products. Available online: https://www.raspberrypi.org/products/ (accessed on 21 January 2021).
40. The Raspberry Pi Foundation. Raspberry Pi Zero W. Available online: https://www.raspberrypi.org/products/raspberry-pi-zero-w/ (accessed on 21 January 2021).
41. The Raspberry Pi Foundation. Raspberry Pi Camera Module V2. Available online: https://www.raspberrypi.org/products/camera-module-v2/ (accessed on 21 January 2021).
42. The Raspberry Pi Foundation. Raspberry Pi 3 Model B. Available online: https://www.raspberrypi.org/products/raspberry-pi-3-model-b/ (accessed on 21 January 2021).
43. SIA Mikrotīkls. MikroTik hAP. Available online: https://mikrotik.com/product/RB951Ui-2nD (accessed on 21 January 2021).
44. Venkateswaran, R. Virtual private networks. *IEEE Potentials* **2001**, *20*, 11–15. [CrossRef]
45. Lozoya-de-Diego, A.; Villalba-de-Benito, M.T.; Arias-Pou, M. Taxonomía de información personal de salud para garantizar la privacidad de los individuos. *Prof. De La Inf.* **2017**, *26*, 293–302. [CrossRef]
46. Rosenbrock, A. ImUtils Library. Available online: https://github.com/jrosebr1/imutils (accessed on 25 January 2021).
47. Singh, K.J.; Kapoor, D.S. Create Your Own Internet of Things: A survey of IoT platforms. *IEEE Consum. Electron. Mag.* **2017**, *6*, 57–68. [CrossRef]
48. OpenCV Team. OpenCV: Smoothing Images. Available online: https://docs.opencv.org/3.4/dc/dd3/tutorial_gausian_median_blur_bilateral_filter.html (accessed on 25 January 2021).
49. Karam, G.M. Visualization using timelines. In Proceedings of the 1994 ACM SIGSOFT International Symposium on Software Testing and Analysis; Association for Computing Machinery, Seattle, WA, USA, 17 August 1994; pp. 125–137.
50. Nurseitov, N.; Paulson, M.; Reynolds, R.; Izurieta, C. Comparison of JSON and XML data interchange formats: A case study. *Caine* **2009**, *9*, 157–162.
51. Huwer, S.; Niemann, H. Adaptive change detection for real-time surveillance applications. In Proceedings of the Third IEEE International Workshop on Visual Surveillance, Dublin, Ireland, 1 July 2000; pp. 37–46.

sensors

MDPI

Article

SAR.IoT: Secured Augmented Reality for IoT Devices Management

Daniel Fuentes [1], Luís Correia [1], Nuno Costa [1], Arsénio Reis [2], João Barroso [2] and António Pereira [1,3,*]

[1] Computer Science and Communication Research Centre, School of Technology and Management, Polytechnic Institute of Leiria, 2411-901 Leiria, Portugal; daniel.fuentes@ipleiria.pt (D.F.); luis.correia@ipleiria.pt (L.C.); nuno.costa@ipleiria.pt (N.C.)

[2] INESC TEC, University of Trás-os-Montes e Alto Douro, Quinta de Prados, 5001-801 Vila Real, Portugal; ars@utad.pt (A.R.); jbarroso@utad.pt (J.B.)

[3] INOV INESC Inovação, Institute of New Technologies, Leiria Office, Campus 2, Morro do Lena-Alto do Vieiro, Apartado 4163, 2411-901 Leiria, Portugal

* Correspondence: apereira@ipleiria.pt

Abstract: Currently, solutions based on the Internet of Things (IoT) concept are increasingly being adopted in several fields, namely, industry, agriculture, and home automation. The costs associated with this type of equipment is reasonably small, as IoT devices usually do not have output peripherals to display information about their status (e.g., a screen or a printer), although they may have informative LEDs, which is sometimes insufficient. For most IoT devices, the price of a minimalist display, to output and display the device's running status (i.e., what the device is doing), might cost much more than the actual IoT device. Occasionally, it might become necessary to visualize the IoT device output, making it necessary to find solutions to show the hardware output information in real time, without requiring extra equipment, only what the administrator usually has with them. In order to solve the above, a technological solution that allows for the visualization of IoT device information in actual time, using augmented reality and a simple smartphone, was developed and analyzed. In addition, the system created integrates a security layer, at the level of AR, to secure the shown data from unwanted eyes. The results of the tests carried out allowed us to validate the operation of the solution when accessing the information of the IoT devices, verify the operation of the security layer in AR, analyze the interaction between smartphones, the platform, and the devices, and check which AR markers are most optimized for this use case. This work results in a secure augmented reality solution, which can be used with a simple smartphone, to monitor/manage IoT devices in industrial, laboratory or research environments.

Keywords: augmented reality; internet of things; IoT devices monitoring; IoT security; low-cost solution

Citation: Fuentes, D.; Correia, L.; Costa, N.; Reis, A.; Barroso, J.; Pereira, A. SAR.IoT: Secured Augmented Reality for IoT Devices Management. *Sensors* **2021**, *21*, 6001. https://doi.org/10.3390/21186001

Academic Editor: Robert S. Allison

Received: 31 July 2021
Accepted: 31 August 2021
Published: 7 September 2021

1. Introduction and Motivation

Currently, Internet of Things (IoT) solutions are becoming increasingly common in several areas (e.g., industry, agriculture, human location, and home automation) [1–3]. A key factor for their ease in adoption is the reasonable low cost of this type of equipment, which by not having relevant output peripherals such as an LCD displays can keep the costs low [4,5]. Considering that occasionally it is necessary to visualize the IoT device output or access real-time configurations, and that a simple LCD display device might cost much more than the IoT device itself, it becomes necessary to research solutions to visualize the IoT device data without the use of specific or additional equipment, only that already available to the administrator, namely a smartphone.

A solution with these characteristics can be used in different contexts such as configure information systems, support systems for the elderly to take medication, visualize the state of objects in a home, or even monitor industrial machines with IoT devices integrated. It is intended that the access to the information, of the various IoT devices, is done in a simple way, through a simple smartphone. Considering that the data to be accessed may be confidential, it is a main requirement to guarantee information security, guaranteeing

that the data will only be available to those who have permission, and not to third parties. Considering that, in the scope of IoT systems, many of the equipment are ubiquitous with no information output interfaces, it is useful to use augmented reality to show this information in real-time near the device itself.

A possible scenario of this is a set of IoT devices scattered in a factory/farm that are controlling various sensors, namely temperature, pressure, and CO_2, among others. For an administrator who needs to access real-time information from one of the IoT devices, assuming a universe of 1000 devices, it is necessary for the user to identify where the device is, see the identifier tag associated with it, access it and see the data that it is collecting, while assuming that the device has a web access interface or similar.

Using an augmented reality solution, where it is only necessary to point with an ordinary smartphone to visualize all the information on that device, and even access and configure it, makes the stated work in this article an interesting idea for this specific purpose. Although a solution with a dashboard that aggregates all the information from all devices is also useful and more common, using augmented reality enables the possibility of seeing the information of the device we are pointing to, in real-time, as if it had an output LCD display per example.

In this scope, the current work researches and presents a low-cost solution to monitor the status of IoT devices, in a secure way, using a simple smartphone and augmented reality.

2. Related Work

This section presents some works related to the theme of augmented reality (AR) associated with the Internet of Things. Some articles explained how augmented reality works and others have shown some solutions already implemented in the world of IoT.

Augmented reality combines information and virtual elements with real world imagery acquired through a camera. AR is becoming increasingly popular in common application for general public entertainment (e.g., gaming, video, and photo filters in social media mobile applications). In specific fields, there are other works, focused on marker detection, information security, platforms for interaction with devices, and IoT. Since this is one of the objectives of this document, works related to this theme will be addressed [6]. The implementation of the concept of augmented reality includes several types of technologies [7]: marker-based, marker less, projection-based, and overlay-based. Benefiting from a lower complexity in the interpretation of information, the type most widely implemented and used is augmented reality using a marker. In this approach, a camera and some type of marker is used, and the visual information is only shown when the marker is detected by a device using image or pattern recognition [8]. Ensuring that markers are detected with minimal latency time is a major challenge, and factors such as brightness and distance can affect marker recognition time [9,10].

Regarding previous research, one in particular has motivated a lot of interest—on which this solution was inspired—where the authors in [11], managed to use a smartphone and augmented reality to obtain the status of an IoT device, presenting some real examples. Additionally, a scalable AR framework called ARIoT was presented in [12], where the authors showed how a much friendlier environment makes use of AR to interact with the home IoT appliances. In [13], the focus was the benefits that augmented reality brings to public transport in smart cities and why it should always be used. Another interesting work is shown in [14], where augmented reality and a set of data information provided by IoT devices are used to locate the real position of various wireless transmitters. In the case of the platform presented in [15], it aims to make users aware of the energy consumption of the various electronic equipment in their home. For this purpose, the authors developed an interactive system that can display the energy consumption, measured by several IoT devices. This platform allows the user to visualize the energy consumption in real time and to interact with the device through AR. In the field of agriculture, there are also low-cost IoT solutions that provide real-time monitoring of crops [16], making the data visually available through AR. This work introduces the use of augmented reality as a support for

IoT data visualization, also called AR-IoT. This concept superimposes the data collected from IoT devices directly to real-world objects and enhances the interaction with them. Regarding interior design, some applications that use AR technology have been developed, for example, in pre-sales, the customers can place and visualize furniture pieces inside their homes before purchasing them [17]. In assistive solutions, there are systems that use AR to assist people, for example, the authors of the work proposed in [18] developed a prototype that aimed to assist visually impaired people to read visual signs. The prototype consists of an augmented reality device, installed on top of the user's head, which identifies real-world text (e.g., signs, room numbers, amongst others), highlights the location of the text, converts it into high-contrast letters through AR, and reads the content aloud through text-to-speech conversion.

Most AR applications provide immersive virtual experiences by capturing information from the user's environment and superimposing the virtual output to augment the user's perception of the real world. The immersive interface and the user's perception shift create serious safety and privacy concerns, mainly in situations where the AR information accuracy is crucial for the user (e.g., while driving a car). Because of this, it becomes essential to implement mechanisms to ensure that the information provided through AR is not affected by malicious applications or bugs [19].

The work proposed in the following sections of this paper was developed according to the concepts presented in [11,12,15,16] to create an information visualization system for IoT devices, in real-time, using augmented reality and adding a security layer to the AR. The work presented in [11] demonstrates how augmented reality can be used to expose information from IoT devices to the users, and in this case, using a simple smartphone to achieve that. The solutions shown in [12,15] confirm that the usage of AR to interact and monitor IoT devices is a valid option. In [16], although the focus is to use IoT devices and computer vision, it is not a solution designed to present information from IoT devices to the user, but to show information about something that the IoT devices are acquiring and processing from plantations. Even so, the assumptions exposed in this work and the information processing techniques are in accordance with what is necessary to the development of the work created in this article.

3. Conceptual Architecture

In this section, the conceptual architecture of the Secure Augmented Reality for Internet of Things (SAR.IoT) solution, oriented to the industry and research areas, is pressented. The main objective is to allow an augmented reality interaction between the user and the IoT devices, all through a web solution, and that guarantees the security of the information. The different modules are specified in detail below, namely the Client, Server and the IoT device.

The SAR.IoT solution has a distributed and multi-agent architecture, as presented in Figure 1, which is mainly divided in three major roles (Client, Server, and IoT devices), in a total of four modules.

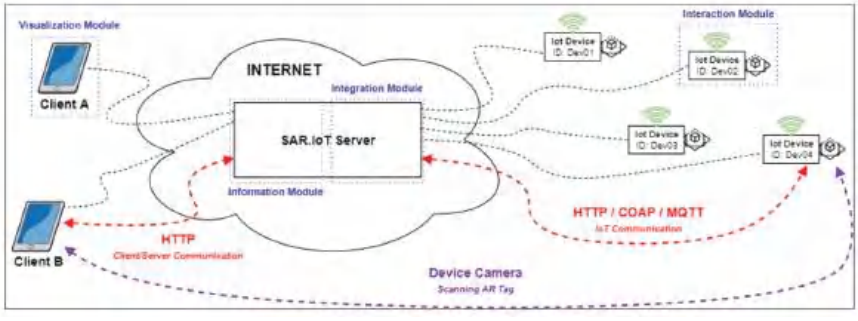


Figure 1. SAR.IoT conceptual architecture.

A distributed architecture [20] is composed of several modules that interact with each other in which each one is responsible for performing a specific task, and where the correct functioning of the entire system depends on the correct interaction of all the dispersed modules. A multi-tenant architecture allows having several customers/entities to interact with the system in general, while this interaction is carried out through the use of credentials that guarantee the privacy of the data.

This proposed architecture was specified considering the communication architecture most widely used in the IoT universe, the Client/Server architecture [21], having been properly modified to incorporate all the requirements necessary for the smooth functioning of the solution. The architecture comprises four modules:

- The visualization module, acting in the client role, acquires and processes images to identify possible markers. In the case of a positive identification, it queries the information module for the data related to the identified marker;
- The information module, acting in the server role, stores the data related to the IoT devices and their associated AR markers. It replies to requests from the display module and assures the security of the information;
- The integration module, also acting in the server role, provides communication between the IoT devices and the information module; and
- The interaction module, acting in the IoT role, provides interaction between the integration module and the IoT device and is located on the device itself.

Each module has an agent, a software-based entity, which is responsible for performing various tasks and ensuring the efficient operation of the overall system.

3.1. Client

The interconnection between the Client, Server, and IoT device can be seen in Figure 2. Note that although there is a direct interaction between the Client and IoT device, this only occurs for the AR marker reading associated with the device, all the communications were performed using the server.

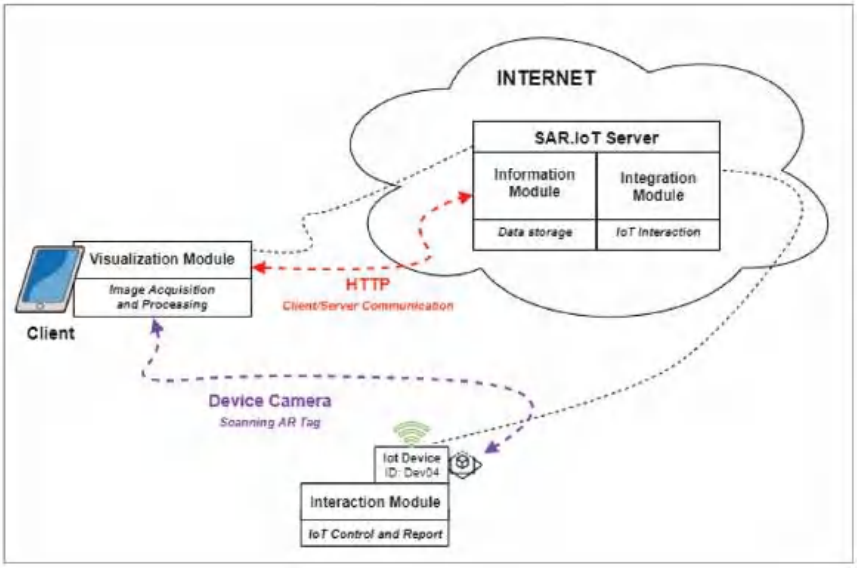


Figure 2. Client role.

The client is a user device such as a smartphone and hosts the visualization module installed and its software agent. This agent acquires and analyzes images to extract AR markers using the resources available in the client device and the augmented reality

framework implemented in the solution, as displayed in Figure 3, allowing the interaction using the smartphone screen.

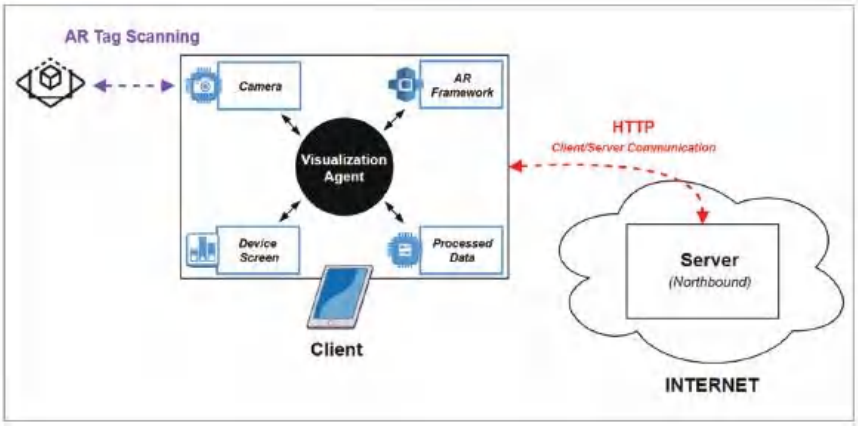


Figure 3. Visualization module architecture.

After identifying the marker, the agent queries the information module present on the server via the server's northbound interface to obtain the related data and to display it on the screen of the client device.

3.2. Server

Figure 4 shows part of the proposed architecture for the server, where it is possible to see the two modules within it, the information module, and the integration module, each performing their respective tasks, interacting with each other and with the other modules via northbound and southbound, respectively.

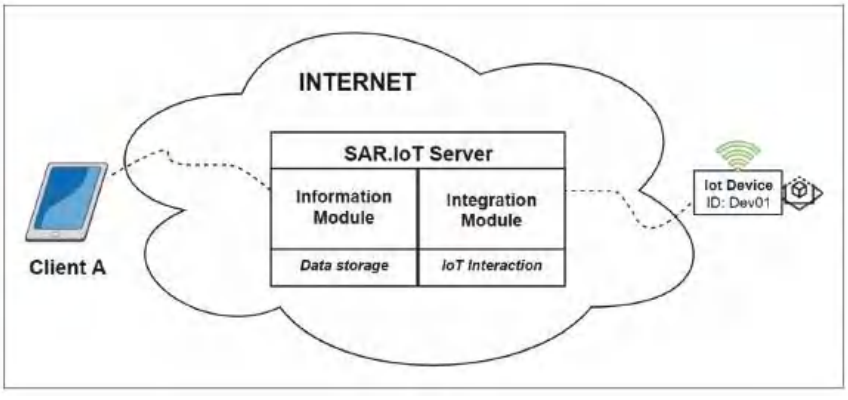


Figure 4. Server role.

The server hosts the information agent and the integration agent, as shown in Figure 5. The server authenticates and replies to the requests from the clients and authenticates and receives information from the IoT devices. The information agent receives, processes, and replies to the requests made to the server. The integration agent receives information from the IoT devices and forwards it to the information agent for data generation and storing. This last agent also performs the actions on the IoT devices such as enabling output, etc.

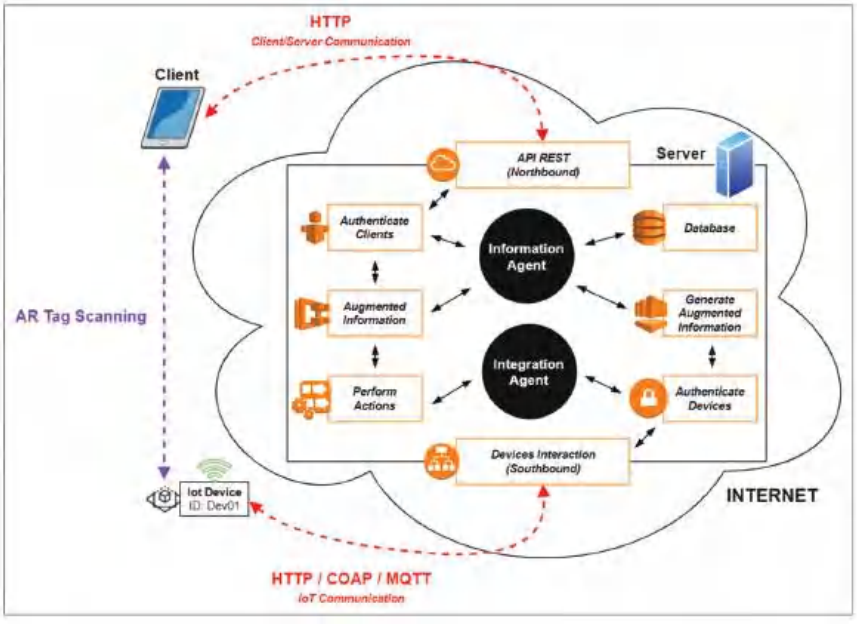


Figure 5. Visualization module and integration module architecture.

The server stores the configuration information as well as all the information sent by the IoT devices. It manages the identification, authentication, and access of IoT devices and users, making the access to the data secure.

3.3. IoT Device

In order to be able to interact with an IoT device, it is necessary to be able to communicate with it, either to obtain information about the equipment itself or to perform actions. For this to be possible, it is necessary to provide this IoT device with a software-based agent that returns the needed information and performs the desired actions. Figure 6 shows the zoom at this point in the general architecture, where it is possible to visualize the connection between the server's integration module and the IoT devices of the interaction module.

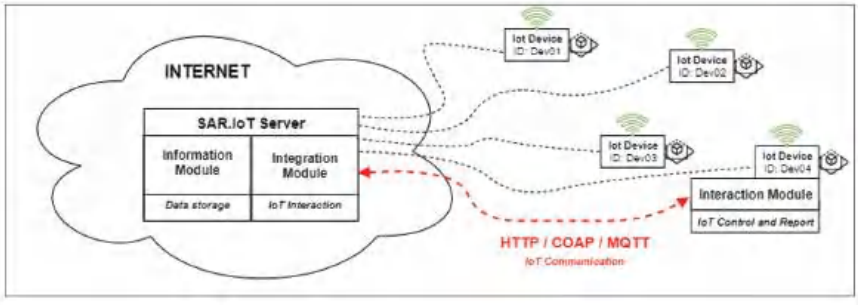


Figure 6. IoT device role.

The agent present in the interaction module, shown in Figure 7, allows for the collection of all the information within the sensors associated with the IoT device, performing actions on the outputs, accessing information about the device itself, among others. The in-

teraction module is responsible for communicating with the server through the integration module and sending/receiving all the data necessary for the system to function.

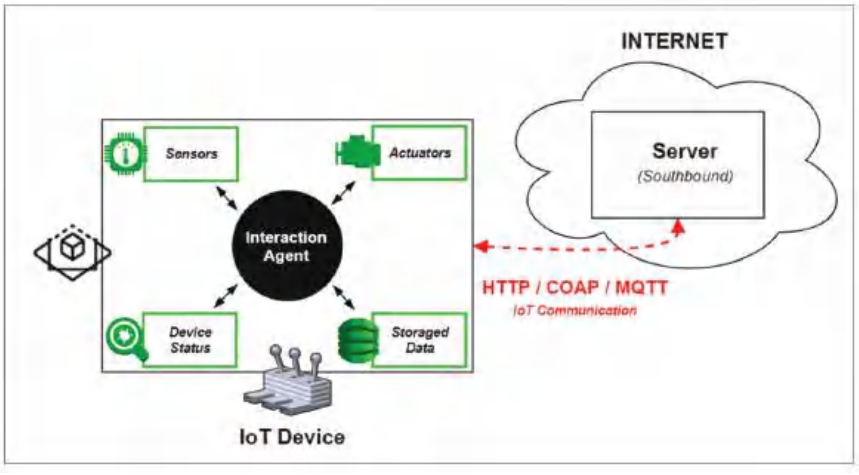


Figure 7. Interaction module architecture.

3.4. Communication

In Figure 8, part of the proposed architecture is presented, focused on the different communication protocols incorporated in it.

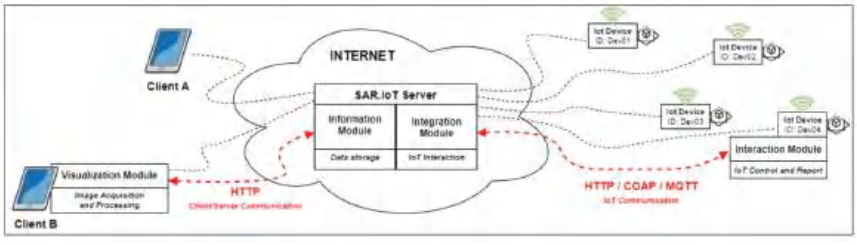


Figure 8. Communication architecture.

In the communication between the different modules in the architecture, and since it is an IoT environment, it makes sense to use communication protocols suitable for this purpose. In the previous figure, there are two different types of communication: northbound communication between clients and the server, accessing the information module; and southbound communication between the integration module on the server and the interaction module on IoT devices.

In the northbound communication, since this typically occurs in a web environment over the Internet, the communication protocol to be used will undoubtedly be HTTP (Hypertext Transfer Protocol)—in its secure version, HTTPS—to ensure data security. This is one of the most widely used protocols [21] for accessing online platforms and is widely used in the IoT environment for the same purpose.

In the southbound communication, since it is mostly communication between IoT devices and the server, several protocols focused on the IoT environment can be used:

- HTTP (Hypertext Transfer Protocol): The most used Client/Server communication protocol on the web, which is also widely used in the IoT world due to its simplicity and efficiency in the delivery of information;

- COAP (Constrained Application Protocol): A communication protocol designed for devices that have limited processing capabilities, very similar to HTTP, but uses much less data when sending messages; and
- MQTT (Message Queuing Telemetry Transport): One of the lightest communication protocols, uses the Publisher/Subscriber model to exchange messages and is widely used in scenarios where network connectivity is not ideal.

These are just some examples of some of the communication protocols most widely used by programmers that can be applied to this architecture. In southbound communication, the use of HTTPS is recommended, the secured version of HTTP. The protocols COAP and MQTT can also be used, but only with an implementation of the protocols that ensure the data security, namely Lithe [22] and SMQTT [23].

4. Prototype Implementation

This section presents the prototype developed to test and validate the proposed architecture, describing the analyzed frameworks, the used equipment, and the operation of the entire solution. For this project, a solution was developed, mostly focused on web technologies, called the Secured Augmented Reality for IoT, shortly named SAR.IoT.

4.1. Frameworks, SDKs, and Augmented Reality Libraries

Below is presented a review of the four most widely used and currently available SDKs. For this project, we selected the ARToolkit SDK, a choice justified at the end of the subsection.

4.1.1. Vuforia

The Vuforia SDK [24] is one of the most popular augmented reality SDKs to develop AR solutions for Android, iOS, UWP, and Unity. It can recognize images, objects, and text. It uses simultaneous localization and mapping (SLAM) technology, which makes it possible for applications to recognize 3D scenes and objects. Regarding the licensing, Vuforia is free for development.

4.1.2. Apple ARKit

The Apple ARKit framework [25] was introduced in iOS11 and allows for the creation of augmented reality applications for iPhone and iPad. It can recognize images, objects, and text. It also uses SLAM technology in conjunction with the device's built-in sensors. Regarding the licensing, the platform is free, but it only works on Apple devices running iOS11+ and with A9, A10, and A11 processors.

4.1.3. Google ARCore

Google's ARCore SDK [26] was designed to support the creation of AR applications for Android 7.0+ devices. It can also recognize images, objects, and text. It also uses SLAM technology in conjunction with the device's built-in sensors. Regarding the licensing, the platform is free, but only works on Android and iOS devices along with ARKit.

4.1.4. ARToolkit

ARToolKit [27] is a free open-source library, from version 5.2 onward (GPLv3), which can be used to create cross-platform AR applications including Android, iOS, UWP, Unity, and Web solutions. It can recognize images, text, and NFT (Natural Feature Tracking) and was used in many previous works [28] with success. Since one of the defined goals is the use of a web platform, it was selected to implement the presented solution.

4.2. Equipment Used in Prototype

The economic value is an important issue for the device to be used in this project, so we opted to use ordinary equipment, readily available to most users. For the client device

(i.e., to capture images and process the augmented reality), we selected an Android [29] smartphone, displayed in Figure 9, priced around €150.

Figure 9. Smartphone Android Xiaomi A2 Lite.

For the server equipment (i.e., which receives and replies to requests, stores information from the IoT devices, and authenticates users and IoT devices), we selected a Raspberry Pi 3 B+ [30], displayed in Figure 10, priced at €50 including a 5 V 2.5 A power supply and a Micro SD card. The characteristics of this equipment, despite being an IoT device, are suitable for the intended purpose, since it includes a quad-core 1.4 GHz processor and 1 GB of RAM, which assures sufficient server performance.

Figure 10. Raspberry Pi 3 B+.

4.3. Operation

To present augmented information in a device, it is necessary to analyze and process the image captured by the device itself, usually using a proper software library. In this case, we used the ARToolKit library, the Javascript version, JSARToolKit5 [31], to work in a web environment, allowing the system to be used by any device with an updated browser (e.g., smartphones, tablets, computers, etc.). To create the augmented information in Javascript, we used the Three.js library, commonly used for WebGL 3D development. Together with the ARToolKit, we. used the Threex.ARToolKit [32], which is also used by the AR.js library [33].

For this solution, a web platform was developed using HTML, PHP, MySQL, and Javascript, and is accessible by the smartphone devices, though the HTTPS protocol (the secured version of HTTP). The client-side data processing is executed in Javascript on the smartphones and includes the AR marker detection and the augmented content rendering systems, as displayed in Figure 11.

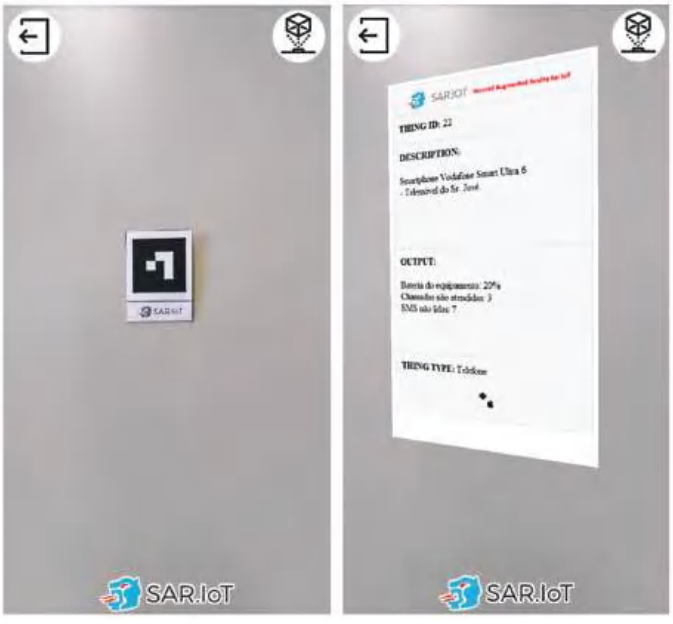


Figure 11. AR marker (**left**) being detected and processed (**right**).

Each client can only see the information of the markers to which it has been granted access by the system administrator. Figure 12 presents an example of this security feature in operation. In the left side is displayed a marker with augmented content to which the user has access rights, and in the right side is displayed a marker to which the user does not have access rights, together with a access restriction notification.

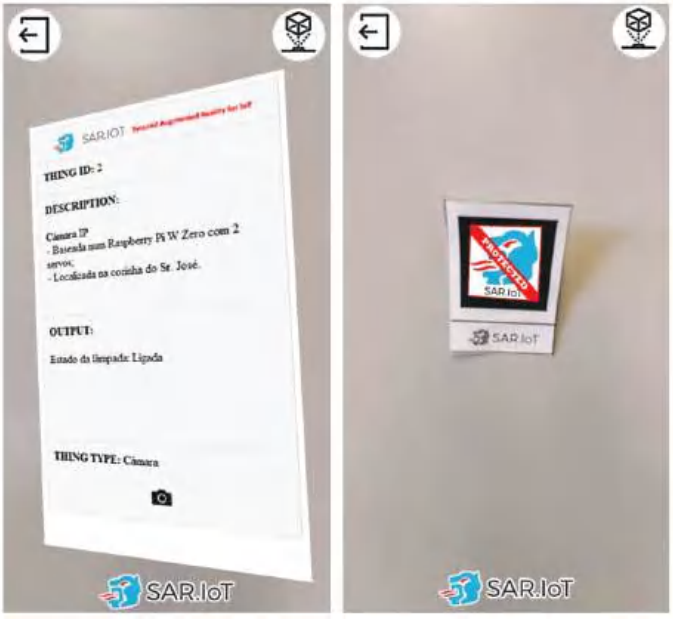


Figure 12. AR markers with (**left**) and without (**right**) access granted.

By doing this, it is possible to have a secure system where only authorized persons can access the information of the devices. The advantage in this method is that the user is able, in real-time, to have a sense of what is happening with the devices, all of this without the need to read codes manually, and subsequently accessing a URL with the device information, in the case of using QR-Codes. It should also be noted that the ease of use of an augmented reality system, in an IoT scenario with these characteristics, is an asset for any multi-user implementation because there is no need to physically interact with the IoT devices.

The diagram in Figure 13 presents a process view of the client operation, detailing the actions performed by each activity of the process.

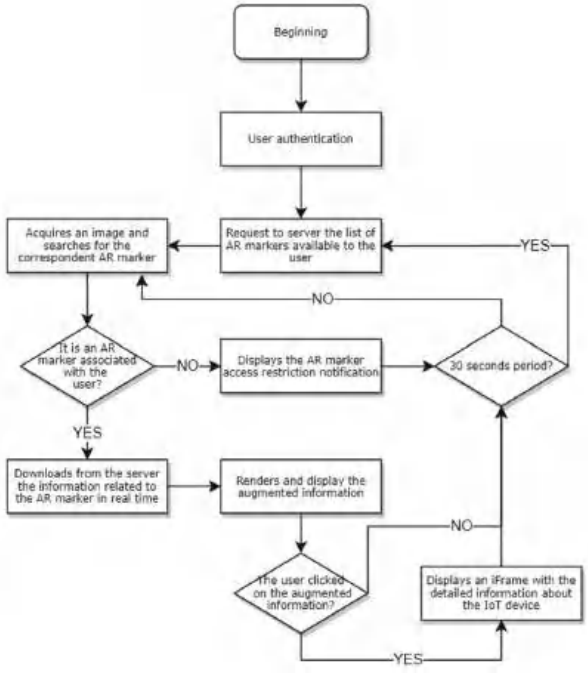


Figure 13. Process view of the client's operation.

That same operation occurs in the following order:

- The process begins when the client logs into the platform and the marker detection system is started;
- The client (client-side Javascript) sends a request to the server for the bookmarks associated with the current user;
- The mobile device captures images and the client searches for AR markers in the images; and
- When a marker is detected, the system checks whether it can be displayed to the current user:
  - If yes, the system downloads the information about the marker from the server and renders and displays it on the device's screen (augmented reality). If the user clicks on the augmented information, an embedded webpage is displayed with the full information regarding the IoT device.
  - If not, the system displays a marker access restriction notification.

The system refreshes the information from the markers and devices every 30 s and the authentication activity is mandatory, as shown in Figure 14. If the user has an adminis-

tration profile, it is redirected to the platform management portal. Otherwise, the client process starts as previously described.

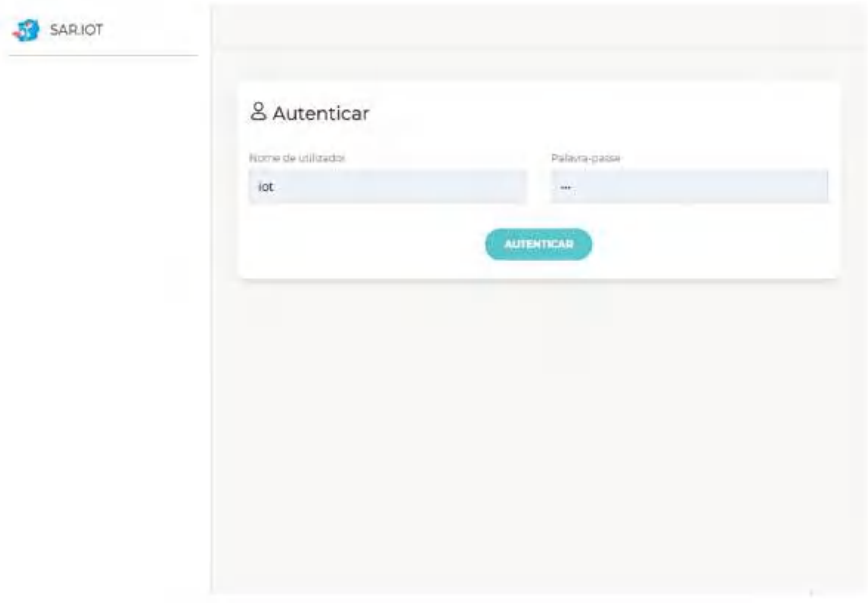


Figure 14. Authentication page.

The SAR.IoT platform includes the management features (i.e., user profiles, IoT devices, and bookmarks and associations). Figure 15 shows the users' management page used for Create, Read, Update, and Delete (CRUD) operations, where there can exist normal users or administrators.

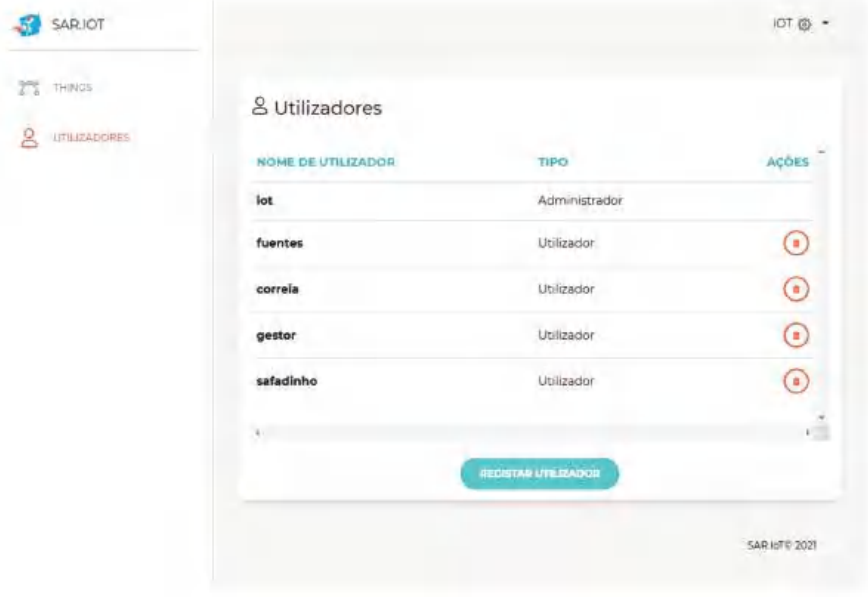


Figure 15. User listing.

Figure 16 shows the IoT device management page with CRUD (Create, Read, Update, and Delete) operations. The IoT devices can specify the device types, the AR markers, and users that can interact with them.

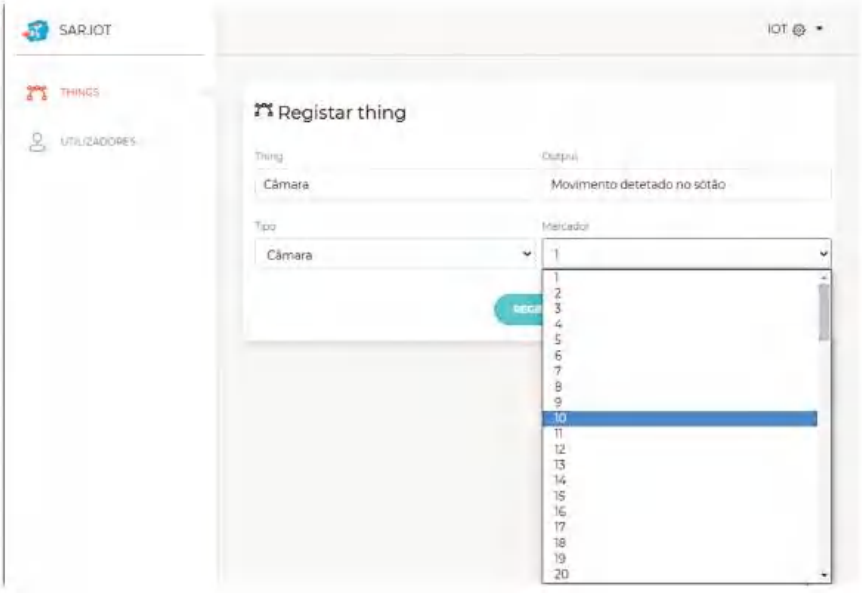


Figure 16. Thing parameters.

Figure 17 shows the page to manage the association between IoT devices and users for access purposes.

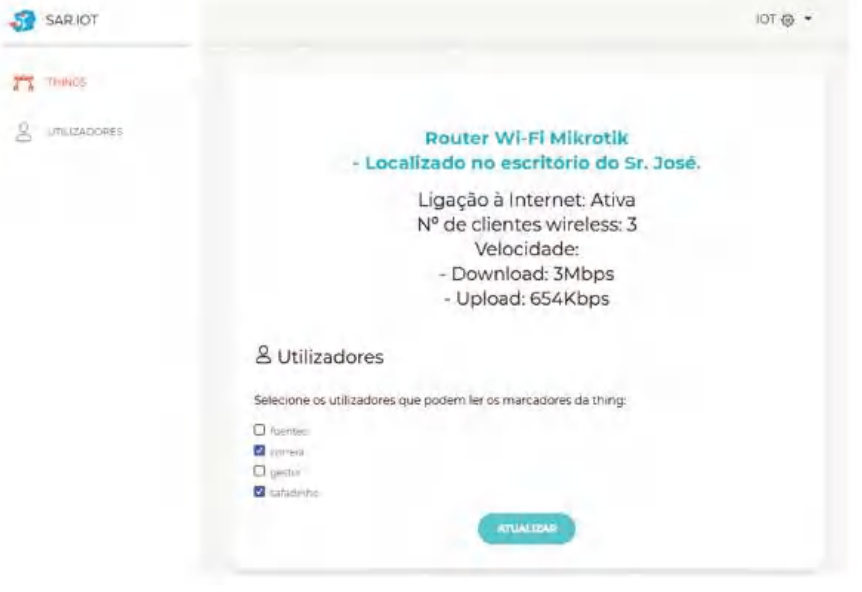


Figure 17. Association between an IoT device and users.

4.4. Visualization Modes

In the visualization module, two visualization modes were implemented: a normal mode and a debug mode, as displayed in Figure 18. The normal mode presents the IoT device's augmented information, while the debug mode adds graphical pins to signal the presence of augmented information that is not available or configured in the system. To switch between modes, the user can click the button on the upper right screen corner.

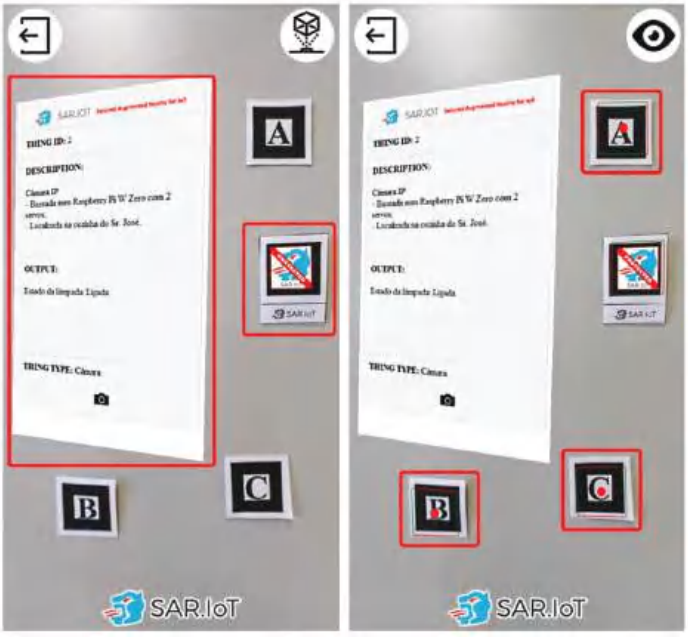


Figure 18. Normal mode (**left**) and debug mode (**right**).

The debug mode is very useful in situations where we want to confirm whether the mobile device (smartphone) is able to recognize the AR marker or not. This is implemented because there may be situations in which the equipment does not present any information and may not be able to read the tags due to some defect on them. With this, it is possible to know if the system is working correctly or not, if it just cannot obtain the information from the server, or if there is another problem with the client module, namely reading the markers.

4.5. Interaction with Augmented Reality

When an AR marker augmented reality information in clicked/pressed on the smartphone's screen, an embedded web page with content regarding the IoT device is produced, as displayed in Figure 19. The contents are updated in real time and can allow interaction with the IoT device.

This available information is acquired by the interaction module and sent to the server through the integration module. It was decided to implement this interaction system to allow the user to consult the information of the IoT devices more comfortably, without the need to be pointing to the marker. The idea in this approach is for the user to be able to verify in real-time the information of all the IoT devices that they can see, but if they want to interact or analyze the information in detail, when clicking on the information generated in AR, a new window appears with all the information that was in the AR information and other further details.

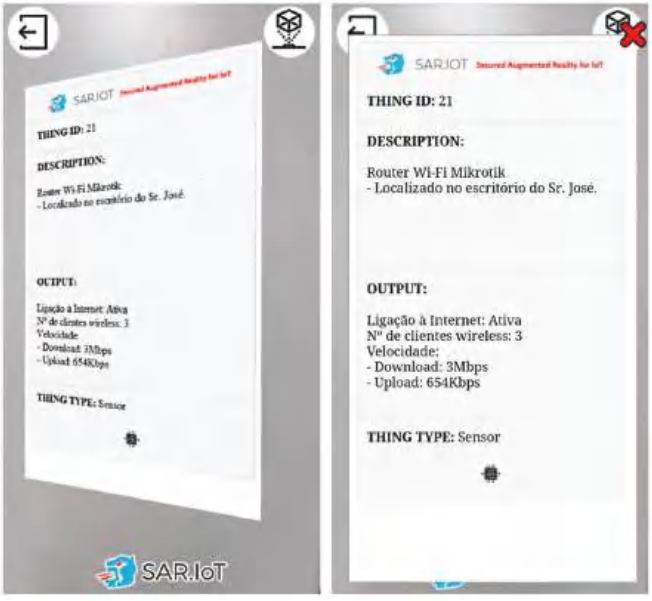


Figure 19. Processed AR marker (**left**) and the correspondent interaction dashboard (**right**).

4.6. An Augmented Monitoring Tool

The great advantage of an augmented reality system incorporated in a smartphone, is its ability to easily allow the detection of equipment that is malfunctioning by simply pointing the smartphone's camera to the IoT devices and checking if any of them have warnings. In the example shown in Figure 20, it is possible to observe a scenario where multiple IoT devices have an AR marker attached (a), when using the proposed system, one of the multiple AR information windows visible in the smartphone is drawing attention (b), that is, there is a problem with that specific device that needs to be checked, doing that by simply closing in the smartphone and seeing what is happening (c). When using a simple dashboard on a smartphone, it indicates that there is a device with problems, but the user must search for the physical equipment and, typically, for written tags with their identification until they find it.

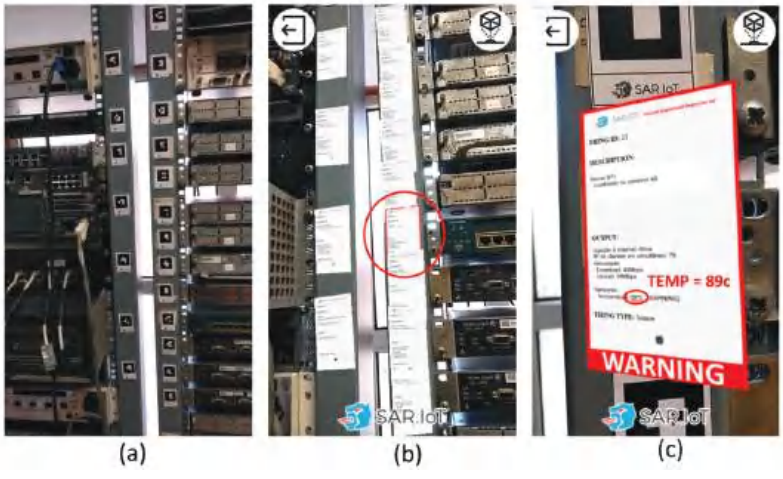


Figure 20. AR markers (**a**) being detected (**b**) and showing an anomaly in a device (**c**).

The main benefit of the given solution is that it transforms any smartphone in a real-time monitoring tool, for example, allowing better monitorization of all the IoT devices in an industry or a laboratory.

5. Tests and Optimizations

In this section, all the tests and optimizations performed are exposed. Various types of AR markers were tested and different degrees of confidence in the recognition system were analyzed. The performance of the solution was also analyzed, with multiple AR markers being shown in the screen at the same time. The security layer of the AR markers was also tested, namely, if the system could block access to the information of the markers to unauthorized users.

5.1. Types of AR Markers

During the development, we tested PATTERN, 2D BARCODE, and NFT (Natural Feature Tracking) markers, the AR markers used in the tests had a size of 2.5 cm $\times$ 2.5 cm, as the ones in Figure 21 that is showing some pattern markers, and the distance of the readings taken with the smartphone varied between 5 cm and 50 cm.

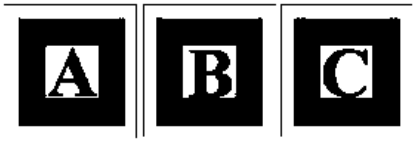


Figure 21. PATTERN markers.

To use PATTERN type markers with this framework, they must be defined in PATT files. These files contain a mapping between the graphic content of the marker's image and numeric values in the 0 to 255 range. Each value represents a color, in a gray scale, from 0 white to 255 black, as presented in Figure 22. An important indicator for the AR markers' recognition is the confidence degree, which is a percentage number that defines the certainty of the recognition or the certain probability of a correct recognition.

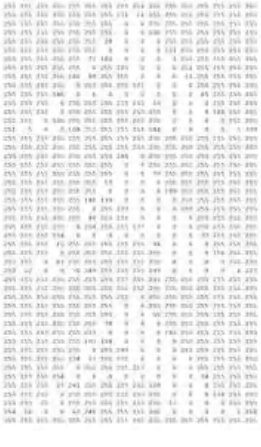


Figure 22. Extract from a.PATT file with an image pattern.

To test the effects of the confidence degree in the AR marker recognition, we used PATTERN type markers (characters) with a 50% confidence degree, and it was verified that the system mismatched the "B" and the "F" markers. In another test, the confidence degree was lowered to 25% and the results were predictably worse. The "B" marker was now also

mismatched with the "D" marker. In a third test, we used an 85% confidence degree, and all the markers (characters) were correctly recognized.

When using 2D BARCODE type markers, visible in Figure 23, it is not necessary to include a definition file, which renders a much lighter processing. In the tests, like the ones previously described, the system always recognized the marker correctly with a confidence degree between 95% and 100%.

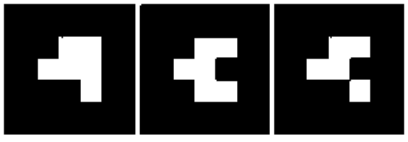


Figure 23. 2D BARCODE markers.

An additional conclusion is that the software library (the Threex.ARToolkit) had a software bug on the calculus of the confidence degree for 2D BARCODE type markers. It would always return a 100% confidence degree. The bug was corrected in the library and the calculation is now accurate.

Figure 24 illustrates an example of some Natural Feature Tracking (NFT) markers that can also be used with the ARToolkit framework. These types of markers allow for the usage of any image to create a customized marker, where it is also necessary to generate from the image for each tag a file with the unique keypoints to correctly identify the marker.

Figure 24. NFT markers.

The results of the tests carried out with the different types of markers are shown in Figure 25. These tests were executed 10 times each, with different distances (50, 25, and 5 cm) and different numbers of markers simultaneously (1, 2, 3, 4, 5, and 10). The values presented in the figure are the average values from the 10 iterations of each one.

Figure 25. Marker detection time comparison (0 = not available).

The 2D BARCODE markers were the fastest to be detected by the system, while the NFT markers were the most time consuming and presented several problems in their detection, namely in terms of distance, where they were only detected at 5 cm from the camera and with more than two markers at the same time, where the system (on the smartphone) could not detect any AR marker.

For the AR markers to be used in the implementation, we chose the 2D BARCODE type as the detection time was lower, the recognition confidence was higher, the detection of multiple markers simultaneously was faster, and it did not need additional files to work.

5.2. AR Performance

The solution exposed in this article was intended to work using a simple browser in a low-medium range smartphone, allowing the largest possible visualization of AR markers at the same tie, being the 60 markers simultaneously on the screen, the maximum allowed by the framework. In order to validate the capabilities of the system, it was necessary to verify the performance of the solution and whether it could handle a large number of markers captured by the system at the same time. For this purpose, several tests were carried out with the AR solution using 4×4 2D barcode markers (allowing a total of 8181 different markers), and in Figure 26, it is possible to see one of those tests, where 60 augmented information windows were generated at the same time on the same smartphone screen.

Figure 26. Performance test with multiple AR markers.

Overall, the system managed to always generate all the AR information windows needed, with only a slight drag in the animation when the number of AR markers recognized by the system was very high (as the picture above), but when zooming on a specific AR marker, this drag completely disappears.

5.3. Marker Security Protection

To assure the security while retrieving the information from the IoT devices, several tests were conducted with a regular smartphone accessing the system, and although the system detects all the markers and processes them all, it only returns and displays the real-time information of the devices available to the current user, so every other device will have access denied and the AR marker will show that, as shown in Figure 27.

In the example above, we can see that the markers unlocked to users U1 (a) and U2 (b) were completely different, the two users had access to different devices, and the system only showed the information of the IoT devices that each user had permission for. Regarding user U3 (c), it could view and manage a set of devices to which users (a)

and (b) also had access. This behavior is very useful, mainly in industrial or laboratorial situations, where different employees/users can only view or manage the IoT devices they have permissions for, which adds a security feature to who can view or not the information in real-time of each device using augmented reality.

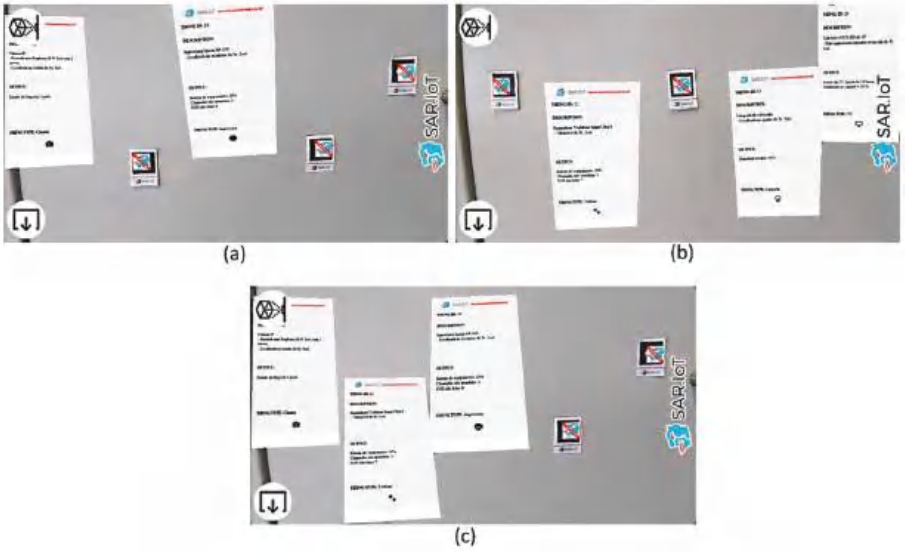


Figure 27. Available and not available AR markers for users U1 (**a**), U2 (**b**) and U3 (**c**).

6. Conclusions and Future Work

In this work, we proposed a real-time visualization system for IoT devices using consumer grade mobile phone devices and augmented reality. The principal objective was to be able to extract information in real-time from the IoT devices and present it using augmented reality without the need for additional specific hardware, keeping the solution low cost. The main contributions in this article were the creation of an architecture that allows the simplified use of augmented reality to visualize information in real-time from IoT devices with a security layer added to the AR, and the development of a functional prototype that demonstrates the operation of the proposed solution and validates the architecture.

The tests results concluded which type of AR marker was best to use and validated the security model used to protect the access to the information on the IoT devices.

In future work, an interesting approach to research would be the implementation of an AR marker generator, creating unique markers optimized for AR usage, mainly using a random quantity of triangles and rectangles, composing a unique pattern.

Author Contributions: Conceptualization, D.F., L.C. and A.P.; data curation, D.F. and L.C.; formal analysis, A.R., N.C., J.B. and A.P.; funding acquisition, A.R., J.B. and A.P.; investigation, D.F. and L.C.; methodology, A.P.; resources, A.R., J.B., N.C. and A.P.; software, D.F. and L.C.; supervision, A.R., J.B., N.C. and A.P.; validation, N.C., A.R., J.B. and A.P.; writing—original draft, D.F., L.C., N.C., A.R. and A.P.; writing—review and editing, D.F., L.C., N.C., A.R., J.B. and A.P. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Project "Digitalization of end-of-line distributed testers for antennas ("D-EoL-TA)", operation number: POCI-01-0247-FEDER-049698, financed by the Program COMPETE 2020, Portugal 2020, by National Funds through the Portuguese funding agency, FCT-Fundação para a Ciência e a Tecnologia, within project UIDB/04524/2020, and was partially supported by Portuguese National funds through FITEC-Programa Interface, with reference CIT

"INOV-INESC Inovação-Financiamento Base" and by Portuguese Fundação para a Ciência e a Tecnologia-FCT, I.P., under the project UIDB/50014/2020.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors acknowledge the Computer Science and Communication Research Center for the facilities granted in the implementation of part of this work, in the context of the Smart IoT Ecosystems research line, and the Mobile Computing Laboratory of the School of Technology and Management of the Polytechnic of Leiria. The authors also acknowledge the authorship of some of the images used in some of the visual content created using the tool "diagrams.net" and the free content available in "iconfinder.com", "pixabay.com", and "flaticon.com".

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Farooq, M.; Riaz, S.; Abid, A.; Abid, K.; Naeem, M. A Survey on the Role of IoT in Agriculture for the Implementation of Smart Farming. *IEEE Access* **2019**, *7*, 156237–156271. [CrossRef]
2. Hutabarat, D.; Hendry, H.; Pranoto, J.; Kurniawan, A. Human tracking in certain indoor and outdoor area by combining the use of RFID and GPS. In Proceedings of the 2016 IEEE Asia Pacific Conference on Wireless and Mobile (APWiMob), Bandung, Indonesia, 13–15 September 2016.
3. Zemrane, H.; Baddi, Y.; Hasbi, A. IoT Smart Home Ecosystem: Architecture and Communication Protocols. In Proceedings of the 2019 International Conference of Computer Science and Renewable Energies (ICCSRE), Agadir, Morocco, 22–24 July 2019.
4. Ferreira, G.; Penicheiro, P.; Bernardo, R.; Mendes, L.; Barroso, J.; Pereira, A. Low Cost Smart Homes for Elders. In *Universal Access in Human–Computer Interaction. Human and Technological Environments, Proceedings of the International Conference on Universal Access in Human-Computer Interaction, Vancouver, BC, Canada, 9–14 July 2017*; Springer International Publishing: Springfield, IL, USA, 2017; pp. 507–517.
5. Ferreira, G.; Penicheiro, P.; Bernardo, R.; Neves, Á.; Mendes, L.; Barroso, J.; Pereira, A. Security monitoring in a low cost smart home for the elderly. In Proceedings of the International Conference on Universal Access in Human-Computer Interaction, Las Vegas, NV, USA, 15–20 July 2018; pp. 262–273.
6. Nenna, F.; Zorzi, M.; Gamberini, L. Augmented Reality as a research tool: Investigating cognitive-motor dual-task during outdoor navigation. *Int. J. Hum. Comput. Stud.* **2021**, *152*, 102644. [CrossRef]
7. Brito, P.Q.; Stoyanova, J. Marker versus Markerless Augmented Reality. Which Has More Impact on Users? *Int. J. Hum. Comput. Interact.* **2018**, *34*, 819–833. [CrossRef]
8. Patrao, B.; Cruz, L.; Goncalves, N. Large scale information marker coding for augmented reality using graphic code. In Proceedings of the 2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR), Taichung, Taiwan, 10–12 December 2018; pp. 132–135.
9. Bhakar, S.; Bhatt, D.P. Product Application to Recognize the Marker Through Augmented Reality. In Proceedings of the 2019 6th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 7–8 March 2019; pp. 594–601.
10. Xavier, R.S.; da Silva, B.M.F.; Gonalves, L.M.G. Accuracy analysis of augmented reality markers for visual mapping and localization. In Proceedings of the 13th 2017 Workshop of Computer Vision (WVC), Natal, Brazil, 30 October–1 November 2017; pp. 73–77.
11. White, G.; Cabrera, C.; Palade, A.; Clarke, S. Augmented reality in iot. In *Proceedings of the International Conference on Service-Oriented Computing, Hangzhou, China, 12–15 November 2018*; Springer: Cham, Switzerland, 2018; pp. 149–160.
12. Jo, D.; Kim, G.J. ARIoT: Scalable augmented reality framework for interacting with Internet of Things appliances everywhere. *IEEE Trans. Consum. Electron.* **2016**, *62*, 334–340. [CrossRef]
13. Pokrić, B.; Krco, S.; Pokrić, M. Augmented reality based smart city services using secure iot infrastructure. In *Proceedings of the 2014 28th International Conference on Advanced Information Networking and Applications Workshops, Victoria, BC, Canada, 13–16 May 2014*; IEEE: New York, NY, USA, 2014; pp. 803–808.
14. Park, Y.; Yun, S.; Kim, K.H. When IoT met augmented reality: Visualizing the source of the wireless signal in AR view. In Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services, Seoul, Korea, 17–21 June 2019; pp. 117–129.
15. Purmaissur, J.A.; Towakel, P.; Guness, S.P.; Seeam, A.; Bellekens, X.A. Augmented-reality computer-vision assisted disaggregated energy monitoring and IoT control platform. In Proceedings of the 2018 International Conference on Intelligent and Innovative Computing Applications (ICONIC), Mon Tresor, Mauritius, 6–7 December 2018; pp. 1–6.
16. Phupattanasilp, P.; Tong, S.-R. Augmented Reality in the Integrative Internet of Things (AR-IoT): Application for Precision Farming. *Sustainability* **2019**, *11*, 2658. [CrossRef]
17. Sandu, M.; Scarlat, I.S. Augmented Reality Uses in Interior Design. *Inform. Econ.* **2018**, *22*, 5–13. [CrossRef]

18. Yang, X.-D.; Huang, J.; Jarosz, W.; Dunn, M.J.; Cooper, E.A.; Kinateder, M. An augmented reality sign-reading assistant for users with reduced vision. *PLoS ONE* **2019**, *14*, e0210630.
19. Lebeck, K.; Ruth, K.; Kohno, T.; Roesner, F. Securing Augmented Reality Output. In Proceedings of the 2017 IEEE Symposium on Security and Privacy (SP), San Jose, CA, USA, 22–26 May 2017; pp. 320–337.
20. El-Sayed, H.; Sankar, S.; Prasad, M.; Puthal, D.; Gupta, A.; Mohanty, M.; Lin, C.T. Edge of things: The big picture on the integration of edge, IoT and the cloud in a distributed computing environment. *IEEE Access* **2017**, *6*, 1706–1717. [CrossRef]
21. Dizdarević, J.; Carpio, F.; Jukan, A.; Masip-Bruin, X. A survey of communication protocols for internet of things and related challenges of fog and cloud computing integration. *ACM Comput. Surv. (CSUR)* **2019**, *51*, 1–29. [CrossRef]
22. Raza, S.; Shafagh, H.; Hewage, K.; Hummen, R.; Voigt, T. Lithe: Lightweight secure CoAP for the internet of things. *IEEE Sens. J.* **2013**, *13*, 3711–3720. [CrossRef]
23. Singh, M.; Rajan, M.A.; Shivraj, V.L.; Balamuralidhar, P. Secure mqtt for internet of things (iot). In *Proceedings of the 2015 Fifth International Conference on Communication Systems and Network Technologies, Gwalior, India, 4–6 April 2015*; IEEE: New York, NY, USA, 2015; pp. 746–751.
24. PTC. Vuforia: Market-Leading Enterprise AR. Vuforia. Available online: https://www.ptc.com/en/products/vuforia (accessed on 5 April 2021).
25. Apple Inc. ARKit—Augmented Reality. ARKit. Available online: https://developer.apple.com/augmented-reality/arkit/ (accessed on 5 April 2021).
26. Google Developers. ARCore. Available online: https://developers.google.com/ar/discover (accessed on 5 April 2021).
27. Lamb, P. Ianyyin/Artoolkit5. ARToolkit. Available online: https://github.com/ianyyin/artoolkit5 (accessed on 5 April 2021).
28. Shepiliev, D.S.; Semerikov, S.O.; Yechkalo, Y.V.; Tkachuk, V.V.; Markova, O.M.; Modlo, Y.O.; Kiv, A.E. Development of career guidance quests using WebAR. *J. Phys. Conf. Ser.* **2021**, *1840*, 012028. [CrossRef]
29. Google Inc. Android | The Platform Pushing What's Possible. Android OS. Available online: https://www.android.com/ (accessed on 6 April 2021).
30. The Raspberry Pi Foundation. Raspberry Pi 3 Model B. Available online: https://www.raspberrypi.org/products/raspberry-pi-3-model-b/ (accessed on 6 April 2021).
31. Bux, T. Artoolkitx/Jsartoolkit5. JSARToolKit5. Available online: https://github.com/artoolkitx/jsartoolkit5 (accessed on 5 April 2021).
32. Etienne, J. Threex-artoolkit. Threex.ARToolKit. Available online: https://jeromeetienne.github.io/AR.js/three.js/ (accessed on 5 April 2021).
33. Etienne, J. AR-js-org/AR.js. AR.Js. Available online: https://github.com/AR-js-org/AR.js (accessed on 5 April 2021).

sensors

MDPI

Article

SFPD: Simultaneous Face and Person Detection in Real-Time for Human–Robot Interaction

Marc-André Fiedler *, Philipp Werner, Aly Khalifa and Ayoub Al-Hamadi

Neuro-Information Technology Group, Otto von Guericke University Magdeburg, 39106 Magdeburg, Germany; philipp.werner@ovgu.de (P.W.); aly.khalifa@ovgu.de (A.K.); ayoub.al-hamadi@ovgu.de (A.A.-H.)
* Correspondence: marc-andre.fiedler@ovgu.de

Abstract: Face and person detection are important tasks in computer vision, as they represent the first component in many recognition systems, such as face recognition, facial expression analysis, body pose estimation, face attribute detection, or human action recognition. Thereby, their detection rate and runtime are crucial for the performance of the overall system. In this paper, we combine both face and person detection in one framework with the goal of reaching a detection performance that is competitive to the state of the art of lightweight object-specific networks while maintaining real-time processing speed for both detection tasks together. In order to combine face and person detection in one network, we applied multi-task learning. The difficulty lies in the fact that no datasets are available that contain both face as well as person annotations. Since we did not have the resources to manually annotate the datasets, as it is very time-consuming and automatic generation of ground truths results in annotations of poor quality, we solve this issue algorithmically by applying a special training procedure and network architecture without the need of creating new labels. Our newly developed method called Simultaneous Face and Person Detection (SFPD) is able to detect persons and faces with 40 frames per second. Because of this good trade-off between detection performance and inference time, SFPD represents a useful and valuable real-time framework especially for a multitude of real-world applications such as, e.g., human–robot interaction.

Keywords: face detection; person detection; multi-task learning; real-time detection

Citation: Fiedler, M.-A.; Werner, P.; Khalifa, A.; Al-Hamadi A. SFPD: Simultaneous Face and Person Detection in Real-Time for Human–Robot Interaction. *Sensors* **2021**, *21*, 5918. https://doi.org/10.3390/s21175918

Academic Editor: Manuel José Cabral dos Santos Reis

Received: 3 August 2021
Accepted: 31 August 2021
Published: 2 September 2021

1. Introduction

The detection of face and person bounding boxes from images is very important for a variety of applications. For example, they can be used in the field of human–computer interaction (HCI) to detect possible interaction partners, in autonomous driving to perceive road users such as pedestrians, or in mobile robot navigation to identify moving obstacles. Furthermore, they are the first component for a large number of recognition systems in many applications, such as face recognition [1], facial expression analysis [2,3], body pose estimation [4], face attribute detection [5], human action recognition [6] and others. In such systems, face and/or person detection are often a prerequisite for the following processing steps; so, their detection rate is crucial for the performance of the overall system. Through deep learning, the results in the area of object detection have been greatly improved. However, many state-of-the-art approaches that use deep neural networks require very heavy computation so that inference does not run in real-time on a conventional graphics processing unit (GPU), which severely limits their suitability for many real-world applications that require high framerates.

Our application, for which we combined face and person detection, lies in the area of autonomous robotic systems. The robot must be able to detect persons with their faces in real-time, especially, in close range to the system with only limited computational capacity in order to perform HCI. However, the use of our framework is not limited to this field of application and is useful for many more real-world applications.

A major difficulty for the integration of the two tasks, face and person detection, in a single neural network is the fact that publicly available databases contain only ground truths for one of the two tasks. To the best of our knowledge, there is no extensive dataset containing coordinates of face as well as person bounding boxes. To perform the two tasks simultaneously within the same convolutional neural network (CNN), it is trained using multi-task learning (MTL). The distinctive characteristic of our training procedure lies in the fact that we train our network in a single continuous process simultaneously on both databases for the tasks of face and person detection, although ground truths are missing for one of the two classes in each database. Thereby, we are able to handle this circumstance without the need of generating new labels, since the manual generation of annotations is very time-consuming and the automatic generation only results in annotations of poor quality. To our knowledge, such a training process has not been presented in the research community so far.

In this work, we propose an MTL framework for simultaneous detection of faces and persons, which is able to process 40 frames per second (fps) and is therefore more than real-time capable. This makes it possible to add further downstream recognition tasks to the framework and still maintain its real-time runtime. Thus, the algorithm is very interesting for real-world applications. The results achieved on the WIDER Face [7] and Pascal VOC [8,9] datasets can compete with other lightweight state-of-the-art methods. In addition, our framework is completely end-to-end trainable, without pre-training individual network parts, splitting up the training process, freezing single network layers or creating additional annotations for one database, as it is mostly the case with other MTL networks.

The main contributions of our work can be summarized as follows:

1. We propose a new CNN for Simultaneous Face and Person Detection (SFPD) in real-time, which is completely end-to-end trainable using MTL with two datasets, each containing the ground truths for one of the two detection tasks;
2. A new network architecture was developed which consists of a joint backbone with shared feature maps and separate detection layers for each task;
3. A multi-task loss was designed which allows to generate loss values throughout the whole training process despite missing ground truth labels in the training datasets;
4. Comprehensive experimental validation was performed by comparing the detection performance and inference runtime of multiple algorithms.

Our paper is structured in the following way: In Section 2, related work on general object detection, face detection, and multi-task learning is reviewed. In Section 3, our method is presented in detail with regard to the used network architecture and loss function. In Section 4, the experiments and their results are reported providing details on the training procedure and the datasets used. Finally, in Section 5, conclusions are drawn.

2. Related Work

There are three major research areas related to our work: general object detection, face detection and multi-task learning. This section gives a brief summary about these areas.

2.1. Object Detection

The general goal of object detection is to localize the borders of a wide range of objects inside an image. These object boundaries are described using bounding boxes and are intended to fit as closely as possible to the object shapes. Additionally, a class label is predicted as output for each detected object. It is possible that the image contains multiple objects. The difference to image classification lies in the fact that in classification there is only one object in the image whose class label is predicted as output, but the bounding box is not localized.

Especially due to the developments in the field of deep CNNs, the performance of detection tasks could be increased significantly in recent years. This can be attributed to the large amount of annotated training data, as well as to the availability of more powerful GPUs, enabling the training of increasingly deeper and more complex network

architectures. However, still the most accurate modern neural networks do not operate in real-time and require large number of GPUs for training with a large mini-batch size [10]. Thus, these methods often cannot be applied for real-world applications with specific requirements regarding the runtime, hardware, energy consumption, etc.

Modern detection frameworks usually consist of two parts: A backbone for obtaining the features, which is often pre-trained on ImageNet [11], and a head for predicting the object classes and bounding box coordinates. Thereby, the head parts can be categorized into single-stage and two-stage detectors.

Two-stage detectors initially generate a large amount of generic object proposals. For this purpose, they use external algorithms, such as Selective Search [12], Egde Boxes [13] or Adobe Boxes [14]. In more recent approaches, the generation of object proposals is integrated into the network structure by using a region proposal network making the framework end-to-end trainable. In the next step, each region proposal is classified, whether it contains an object or not using a CNN. The first two-stage object detection algorithm was R-CNN [15], upon which newer variants, such as Fast R-CNN [16], Faster R-CNN [17], R-FCN [18], Mask R-CNN [19] and Libra R-CNN [20] are based on. Although the two-stage detectors have the capability to achieve the best detection accuracy, they are rarely used in practice because of their limited suitability for real-time systems. This is primarily due to the generation of region proposals, which is a computationally intensive process and the main bottleneck for reaching a real-time detection framework.

Single-stage detectors, often also called single-shot detectors, directly compute object confidence scores and bounding box coordinates for a given input image without generating region proposals. For this purpose, a fixed set of anchor boxes with different aspect ratios and scales is applied to all image components in order to be able to immediately predict the confidence scores. This greatly improves the detection speed and enables real-time detection, while reducing the detection accuracy [21]. Due to the better processing speed, the single-stage detectors are used in practice much more often. To ensure detection of differently scaled objects in a single forward pass through the network, they utilize the built-in pyramid structure of CNNs. Feature maps from different stages of layers with various sizes are collected and pooled, allowing the network to perform direct object classification and regression of bounding boxes for several scales of objects. The most representative models for single-stage object detectors are the versions of YOLO [10,22–24], SSD [25] and RetinaNet [26]. In recent years, more approaches have been introduced: EfficientDet [27] is a scalable object detection framework where it is easily possible to change the backbone in order to optimize accuracy and efficiency of the network. With FCOS [28] and FoveaBox [29], two anchor-free frameworks have been introduced. Their advantage lies in the fact that complicated computations related to anchor boxes such as overlaps during training are avoided by eliminating the predefined set of anchors. Instead, pixel-wise classification is applied to the feature map outputs of the backbone, similar to semantic segmentation, for detecting the objects.

The recognition task of person detection is mainly handled within the general object detection, because most object recognition datasets have persons annotated as one of their object categories. Therefore, most general object detection frameworks perform the detection of persons besides further object classes.

2.2. Face Detection

Face detection is a specialization of general object detection, which focuses on the detection of human faces. Many algorithms for face detection have been derived from methods for general object detection.

Before deep learning became the standard in object and face detection, manually acquired features were used to accomplish the detection tasks. One of the most popular algorithms for face detection was developed by Viola and Jones [30]. It utilizes Haar-Like features and AdaBoost [31] learning to train cascaded classifiers, which achieve good performance in real-time speed. Besides Viola and Jones, the deformable parts model

(DPM) [32] has been proposed in the literature [33–35] for face detection using histogram of oriented gradient (HOG) [36] features, which are robust to partial occlusion and define a face as a collection of its parts. The main problem for the usage of Haar-Like and HOG features in unconstrained face detection lies in their inability to capture facial information at different resolution, viewpoint, illumination, expression, skin color, occlusions and cosmetic conditions [37].

To overcome these limitations, various deep learning-based face detection models have been introduced in the literature. One of the first CNN-based face detection algorithms is Cascade-CNN [38]. It uses an image pyramid to detect differently scaled faces. Then, it merges the individual faces detected from pyramid structure for the whole image using non-maximum suppression (NMS) [39], discarding strongly overlapping bounding boxes. A similar cascade is used by Multi-scale Cascade CNN [7] and by MTCNN [40], while MTCNN additionally captures five facial landmarks for improved face detection.

In recent years, many more algorithms have been introduced: Face R-FCN [41] is built on the R-FCN [18] framework and is optimized for face detection. To improve detection accuracy, they exploit position-sensitive average pooling, multi-scale training and testing as well as on-line hard example mining. S^3FD [42] consists of a scale-invariant network with a new anchor matching strategy for improved recall rate on tiny faces. In order to increase performance in particular for partially occluded faces, the specially developed approach FAN [43] uses anchor-level attention maps. In PyramidBox [44], the authors applied context modules on feature pyramids to enlarge the receptive field for better observation of context information. ScaleFace [45] is able to handle an extremely wide range of scales using a specialized set of deep CNNs with different structures. The challenging problem of simultaneous dense localization and alignment of faces of arbitrary scales in images is addressed in RetinaFace [46] through adding a self-supervised mesh decoder branch for additional prediction of pixel-wise 3D shape information. DSFD [47] proposes a novel feature enhance module and an enhanced anchor matching strategy for obtaining more discriminability and better initialization for the regressor. DBCFace [48] is an anchor-free face detector that generates binary segmentation masks indicating for each pixel whether it belongs to a face or not.

Due to this multitude of developments, the performance in the field of face detection has been enhanced significantly. However, the performance of the algorithms is also strongly correlated to the required computation time, which is the reason why almost none of the previous mentioned deep learning approaches are able to run in real-time on a conventional GPU, e.g., PyramidBox [44] only achieves 3 fps on an NVIDIA Titan RTX (Nvidia Corporation, Santa Clara, CA, USA) and ScaleFace [45] only 4 fps on an NVIDIA Titan X. One approach that combines good results with real-time runtime is YOLO-face [49]. The method was developed based on YOLOv3 [24] and reaches 38 fps on an NVIDIA GeForce GTX 1080 Ti.

2.3. Multi-Task Learning (MTL)

MTL describes the simultaneous learning of multiple tasks at the same time, whereby several output targets are generated for one input target [50]. MTL for machine learning was first introduced by Caruana [51] in 1998. However, before deep learning algorithms were extensively deployed, it was highly limited to just a few use cases as the required features strongly differed. With the upcoming trend of using CNNs for computer vision tasks and the rejection of hand-crafted features, the fields of application for MTL could be extended considerably.

Several MTL frameworks were presented such as: DAGER [52] for age, gender and emotion recognition; HyperFace [53] for face detection, pose estimation, landmark localization and gender recognition; or All-In-One [54] for face detection, landmark localization, face recognition, 3D head pose estimation, smile detection, facial age estimation and gender classification. Additionally, Levi and Hassner [55] proposed a CNN for age and gender estimation, Zhang et al. [56] optimized facial landmark localization with facial attribute

inference and head pose estimation, and Gkioxari et al. [57] trained a CNN for person pose estimation and action detection.

Chen et al. [58] proposed to combine face detection and alignment in one framework, because they observed that aligned face shapes provide better features for face detection. Furthermore, Saxen et al. [59] proved that a CNN can detect faces more easily by adding face orientation as a training target. Inspired by these approaches, various methods for face detection were developed, which incorporated the prediction of additional facial features into the network for improved performance: MTCNN [40] and RetinaFace [46] predict five ancillary face landmarks, He et al. [60] predict plenty facial attributes and Wu et al. [61] predict the head pose.

The advantage of having an MTL network, instead of constructing independent CNNs for each task, is to profit from the inherent correlation between the related tasks and thereby to enhance each others performance [61]. By sharing the feature maps for the different detection layers, the generalization capability of the features improves and they can adapt more effectively to the complete set of recognition domains. This enhances both learning efficiency and prediction accuracy [62]. In addition, the shared use of several CNN layers reduces the computational time, which helps realizing a real-time system for simultaneous execution of multiple tasks.

3. Method

This section introduces our new method for simultaneous face and person detection, called SFPD, in detail. The basic design of our SFPD algorithm is inspired by the SSD [25] framework. The layout of the network architecture and the applied loss function are explained in the following subsections.

The novelty of our training procedure and network architecture lies in the fact that it is trained end-to-end on two datasets which are both only partially annotated and therefore only contain labels for one of the two target object classes (faces and persons). We solve this problem algorithmically without an additional generation of new ground truths, since we do not have the resources to generate new labels manually as it is very time-consuming and automatically generated labels are of worse quality. For this reason, the training process alternates between batches with face and batches with person annotations. Details about the training data can be found in Section 4.1, details about the training procedure in Section 4.2.

3.1. Network Architecture

Our SFPD algorithm belongs to the group of single-stage object detectors (see Section 2.1) and is a feed-forward CNN which uses predefined anchor boxes to output bounding box coordinates and confidence scores for the respectively targeted class. The network architecture of SFPD consists of two parts: A joint backbone with shared feature maps and separate detection layers for the single detection tasks. The detailed structure is illustrated in Figure 1.

The backbone generates base features and is shared by the two detection layer branches for faces and persons. This becomes possible because the first layers extract very rough features such as contours and edges. The middle and back layers of the backbone already exhibit specific task-related features, which however gain better generalization ability through training on two related detection tasks. The first part of our backbone consists of the VGG-16 [63] network. Each of these convolutional blocks (conv1–conv5) consists of a series connection of one or more convolutional layers with rectified linear unit (ReLU) activation function and a kernel size of 3×3 followed by a max pooling layer with 2×2 kernel. All weights were pre-initialized with values trained on ImageNet [11]. The ReLU activation is able to increase the overall non-linear fitting ability of the CNN. Similar as in SSD [25], the fully connected layers fc6 and fc7 are replaced by convolutional layers (conv6 and conv7) with 1024 filters each, fully connected layer fc8 is removed and four additional convolutional layer blocks (conv8 to conv11) with two convolutional layers each

and successive kernel sizes 1 × 1 and 3 × 3 are added at the end of the VGG-16 [63] network. The layers of the first additional block have 256 and 512 filters, those of the following three ones first 128 and then 256 filters. This results in a feature map size of 1 × 1 at the end of the backbone for input images with 300 × 300 pixels. The advantage of the 1 × 1 convolution lies in the fact that it performs the dimension reduction of the feature map without significantly increasing the number of parameters. The newly added convolutional layers are initialized by the Xavier [64] method.

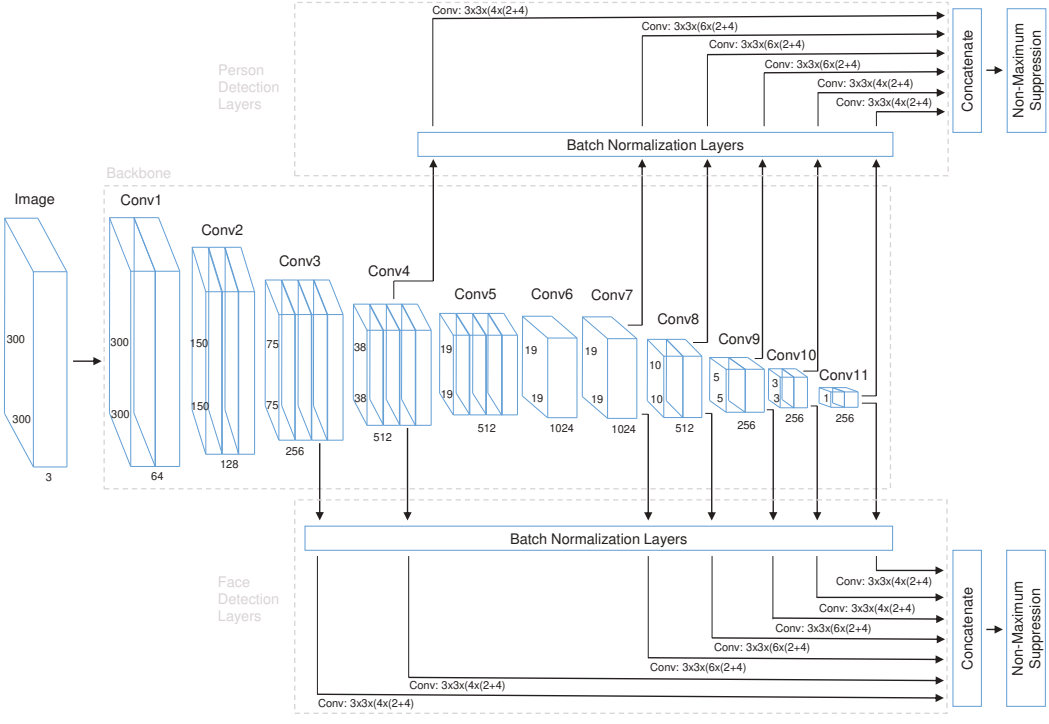


Figure 1. Network architecture of SFPD. It consists of a shared backbone and separate detection layers for face and person detection.

The detection layers generate as output the bounding box coordinates and percentage class confidence scores for each detection task. Therefore, each detection layer consists of two head layers, one for bounding box regression and one for class prediction. In order to be able to detect persons and faces of different scales in one pass through the CNN without generating image pyramids, features must be tapped at different levels of the backbone. This is possible because the layers of the backbone are progressively decreasing in size. The taps for the detection layers are located after layer conv4-3 and conv7 (formerly fc7) as well as at the end of each newly added block following the VGG-16 [63] network. In order to be able to detect more smaller faces, a seventh tap after conv3-3 is added to the branch for the face detection layers. The detection layer for each tap consists of a batch normalization layer followed by two parallel convolutional layers corresponding to the two heads. Afterwards, all detection layer feature maps of a branch are concatenated in order to aggregate the multi-scale detections. The entire CNN is composed of 24, 453, 160 parameters in total from which 24, 451, 112 are trainable.

The anchor boxes are very important hyper-parameters and crucial for the later detection performance. A set of anchor boxes with different sizes and aspect ratios is assigned to each detection layer feature map allowing to cover suitable boxes for a large range of faces and persons that may appear in the images. Usually, the height of faces

and persons in images is greater than the width. Therefore, besides square anchor boxes, additional ones with aspect ratios of one half and one third are applied. However, the test data showed exceptions to this assumption. For that reason the flipped anchor boxes with aspect ratios two and three were also added. The anchor box sizes were adopted from the original SSD300 [25] implementation.

The SFPD network outputs a fixed-sized set of bounding boxes and their respective confidence scores for the presence of a face or person. During inference, the final detections must be generated out of these. Most boxes can already be sorted out by the confidence threshold. The confidence threshold plays an important role, because if it is set too high, correct detections are rejected and if it is set too low, many false positives remain in the results. Depending on the layer where the bounding boxes are tapped, we use different confidence scores because it has been observed that especially for small objects, it is often difficult to achieve a sufficiently high score. Therefore, the bounding boxes from the first two person and the first three face detection layers receive a confidence threshold of 0.1, the next two of 0.2 and the last two of 0.3. To avoid multiplicate detection of the same object, NMS is used. Boxes with an intersection over union (IoU) of more than 0.5 are rejected and a maximum of 300 detections is kept per image. The decision is based on the highest confidence score.

3.2. Loss Function

During the training of our SFPD network a loss function consisting of multiple parts is optimized. For each detection branch, a loss is calculated consisting of a confidence loss (L_{conf}) for the confidence scores and a regression loss (L_{reg}) for the bounding box coordinates.

Since the two detection layers decide for each anchor box, if it contains a person or a face (depending on the branch) or if the box is classified as background, these are both binary decision problems. For this reason, we use the binary cross-entropy loss for L_{conf}:

$$L_{conf} = -\frac{1}{N}\sum_{i=1}^{N} y_i \, log(\hat{y}_i) + (1 - y_i) \, log(1 - \hat{y}_i) \tag{1}$$

where $\hat{y}_i$ is the model output for the i-th anchor box, y_i is the corresponding target value and N is the number of anchor boxes.

We use the generalized intersection over union (GIoU) [65] loss for L_{reg}:

$$L_{reg} = \frac{1}{M}\sum_{j=1}^{M} 1 - IoU_j + \frac{A_j^c - U_j}{A_j^c} \tag{2}$$

where IoU_j is the intersection over union between the predicted and ground truth bounding box for the j-th anchor box remaining after hard negative mining, A_j^c the smallest enclosing area and U_j the union area between the two bounding boxes. M is the total number of remaining anchor boxes after hard negative mining. GIoU was chosen for the regression loss because it is superior to other loss functions in the regression of 2D bounding boxes [65].

The losses for face (L_{face}) and person (L_{person}) detection are calculated from the respective L_{conf} and L_{reg}:

$$L_{face} = L_{conf\_face} + 2 \times L_{reg\_face} \tag{3}$$

$$L_{person} = L_{conf\_person} + 2 \times L_{reg\_person} \tag{4}$$

Thereby, the regression loss is weighted twice as high as the confidence loss. The weight was chosen empirically and resulted in an improved optimization during training. Learning the binary classification proved to be uncritical even with weaker weighting.

To realize a complete end-to-end trainable framework for both detection tasks, a total loss function is required. This total loss L is composed of the loss functions for faces and persons:

$$L = 3\alpha \times L_{face} + \beta \times L_{person} \tag{5}$$

Our training process alternates between batches of face and batches of person samples, which come from different databases. During training a batch with face annotations $\alpha = 1$ and $\beta = 0$ are set, during a batch with person annotations $\alpha = 0$ and $\beta = 1$ are set. This ensures a steady calculation of the loss during the whole training process, despite the fact that one of the two ground truths is missing for the input images. The face loss is triple-weighted compared to the person loss because it has been observed that otherwise the network optimizes itself strongly in the direction of person detection and neglects face detection to a large extent.

By applying this loss function, a network could be designed which is able to detect faces and persons simultaneously. The framework is completely end-to-end trainable, although the available datasets have either face or person labels, but no dataset has both. Details about the exact training procedure can be found in Section 4.2.

4. Experiments and Results

This section describes the experiments and their results in detail. First, the datasets used for training and testing our SFPD network are introduced and, then, the training procedure is precisely specified. Afterwards, the achieved results are presented and discussed. Finally, the limitations of our new algorithm are pointed out.

4.1. Datasets

Training a CNN for simultaneous detection of faces and persons in images is not a straightforward task, as extensive and publicly available datasets, which contain face as well as person bounding box annotations, do not exist in the research community. In order to train such a network, partially annotated datasets have to be used.

For training and testing the face detection task, we utilize the WIDER Face [7] dataset. It is currently the most popular and commonly used dataset in face detection. Besides, it is very challenging due to the high variability in scale, pose, expression and occlusion of the faces pictured in its images. For training, we apply the WIDER train set with 12,880 images and, for testing, the WIDER validation set with 3226 images. The sets are divided into the three categories “easy”, “medium” and “hard” according to their level of difficulty for detection.

The task of person detection is trained and tested using the Pascal VOC datasets [8,9] from 2007 and 2012. The two datasets contain annotations for 20 different object classes, however, we are only interested in the person annotations. For this reason, all images without person annotations are sorted out. In addition, annotations of other object classes are ignored during training. This results in 2095 remaining images for the VOC 2007 trainval set and 9583 for the VOC 2012 trainval set, which are used for training the SFPD network. In total, this results in 11,678 training images with person annotations, which leads to a relatively balanced number of training images between the two detection tasks compared to 12,880 for WIDER train. No negative samples (without faces and without persons) were used in the training as the test performance showed no need for this, since no false positives were detected on images without objects. The same procedure for rejecting images is applied to the person test sets. This leaves 2097 images in the VOC 2007 test set and 5138 in the VOC 2012 test set for testing the person detection of our SFPD network.

4.2. Training Procedure

The SFPD algorithm has been trained on partially annotated databases because there is a lack of datasets with person as well as face annotations. Therefore, the training procedure is slightly more complex compared to other CNNs.

First, the input images are loaded and scaled to a size of 300 × 300 pixels. Thereby, a batch size of 32 is used. Each batch contains only images with either face or person annotations. Batches with mixed images from both detection tasks do not occur in our training process. The images within the batches are randomly selected from the dataset. Whether a face or person batch is loaded, is determined by the probability calculated as

the ratio of the total number of face to person batches. The training epoch ends once all batches of the three training datasets have been loaded.

To increase the generalization capability of the network, various data augmentation techniques are applied to the input images. The images are flipped horizontally with a probability of 0.5 and vertically with 0.1. Furthermore, every third image is rotated in the range of -30 to 30 degrees. Since it is difficult for the network to detect small objects, additional training data are generated. Therefore, the images are effectively downscaled to create smaller faces and persons. For this purpose, every third image is expanded by a black area, which extends the original image size by a random factor between one and four. The aspect ratio remains unchanged. Additionally, some photometric distortions are applied on the input images, such as adjusting the brightness, contrast, saturation and hue.

During training, the anchor boxes have to be matched to the ground truth coordinates. Each anchor box above an IoU threshold of 0.5 is classified as positive. This simplifies the learning problem because the network should not only find the one anchor box with the highest IoU overlap, but should also predict high confidence scores for multiple appropriate anchor boxes. During inference, these multiple detections are sorted out using NMS. Since the number of negative anchor boxes greatly exceeds the number of positive ones at training time, hard negative mining is performed to compensate for this imbalance. The negative classified anchor boxes with the highest confidence scores are selected to obtain a ratio of 3:1 between negative and positive training samples.

All training is performed on an NVIDIA GeForce RTX 2080 Ti GPU. The total number of training epochs is 130. We start with a learning rate of 10^{-4} which increases by factor 10 after the first ten epochs. By starting the training directly with a higher learning rate, an unstable behavior could be observed. Therefore, it is increased after the weights of the network have reached a more stable state. After 80 and 100 epochs, the learning rate is then reduced by a factor of 0.1 each time. As optimizer, we utilize stochastic gradient descent (SGD) with a momentum of 0.9.

4.3. Evaluation Results and Discussion

The evaluation of our SFPD network, which is able to detect faces and persons simultaneously, was conducted on task-specific datasets for each detection target.

To evaluate the person detection, the Pascal VOC [8,9] "person" subsets of 2007 and 2012 were chosen. The results obtained with our SFPD method and other algorithms are presented in Table 1. Sample images from the databases with SFPD detections are shown in Figure 2. Our SFPD method outperforms the comparison algorithms Fast R-CNN [16], Faster R-CNN [17], SSD [25] and the first two versions of YOLO [22,23], which are among the most commonly used object detection frameworks. SFPD has one of the fastest computation times considering that both faces and persons are detected in 40 fps and the Titan X, Titan V and RTX 2080 Ti are GPUs with comparable technical specifications. The average precision score was improved by about two percent compared to SSD [25] with unchanged input image size of 300×300 on both datasets. Compared to EfficientDet-D2 [27], SFPD shows similar performance results but detects faces additionally to persons. However, the comparison is not quite fair since EfficientDet-D2 [27] was trained on the significantly larger MS COCO dataset. The same applies to EfficientDet-D3 [27], which achieves improved detection results but can only process 27 fps. SSD512 [25], RetinaNet [26] and FoveaBox [29] show slightly higher results of less than two percent, however, they are not even half as fast as SFPD and only manage to generate person bounding boxes in this amount of time.

Face detection was tested on the three WIDER Face [7] validation subsets. The results for several of these detection algorithms are listed in Table 2. Furthermore, images with sample detections of SFPD are shown in Figure 3. Corresponding precision–recall curves are outlined in Figure 4. SFPD was compared with a variety of algorithms. The results show that there is only a small number of algorithms that achieve satisfying performance combined with real-time runtime on this dataset. All approaches with average precision

values above 90 percent are not able to be executed in real-time. DSFD [47] with a ResNet50 architecture represents an exception and is capable of running almost in real-time with 22 fps on a high-end Tesla P40 GPU. All other methods at the top of the results list are far below this runtime. This shows that face detection is a complex and computing intensive computer vision task. The two implementations of YOLO-face [49] indicate the best trade-off between performance and runtime achieving 89.9 percent at 38 fps and 82.5 percent at 45 fps on the "easy" subset. Our SFPD ranks just below them in terms of performance. The average precision score is between one and four percent worse on each of the three subsets than YOLO-face [49] with darknet-53 architecture. The framerates are in similar range, but it has to be mentioned that SFPD additionally detects persons in the same amount of time and no additional CNN is needed for this purpose.

Table 1. Results of our SFPD network and other detectors on the Pascal VOC test "person" subsets 2007 and 2012.

	VOC Test Set			
Method	**2007**	**2012**	**fps**	**GPU**
Fast R-CNN [16]	69.9	72.0	1	Tesla K40
Faster R-CNN [17]	76.7	79.6	5	Tesla K40
			7	Titan X
SSD300 [25]	76.2	79.4	46	Titan X
SSD512 [25]	79.7	83.3	19	Titan X
YOLO [22]	-	63.5	45	Titan X
YOLOv2 [23]	-	81.3	40	Titan X
EfficientDet-D2 [27] †	78.8	81.9	43	Titan V
EfficientDet-D3 [27] †	81.1	85.6	27	Titan V
RetinaNet [26]	78.3	-	14	Tesla V100
FoveaBox [29]	79.5	-	16	Tesla V100
SFPD [ours]	78.1	81.5	40 *	RTX 2080 Ti

All frameworks (except † denoted) were trained exclusively with person annotations from the Pascal VOC trainval sets of 2007 and 2012; the inference time was determined with a batch size of one; * at our SFPD method denotes that both faces and persons are detected within this inference time; † denotes that the network is trained on MS COCO [66] and not on Pascal VOC datasets.

Figure 2. Example detections of SFPD on the Pascal VOC [8,9] test sets 2007 and 2012: Red bounding boxes indicate detected faces; green bounding boxes detected persons.

Table 2. Results of our SFPD network and other detectors on the WIDER Face validation set.

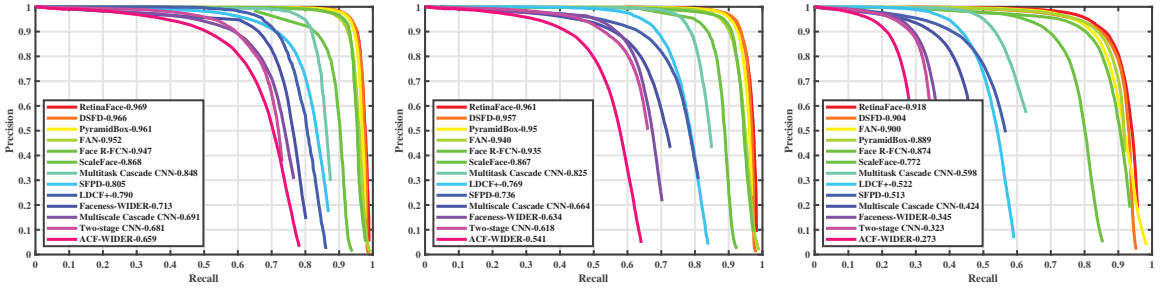
	WIDER Validation Set				
Method	**Easy**	**Medium**	**Hard**	**fps**	**GPU**
YOLOv2 [23] (from [49])	33.1	29.3	13.8	40	Titan X
ACF-WIDER [67]	65.9	54.1	27.3	20	CPU
Two-stage CNN [7]	68.1	61.8	32.3	-	-
YOLOv3 [24] (from [49])	68.3	69.2	51.1	35	Titan X
Multi-scale Cascade CNN [7]	69.1	66.4	42.4	-	-
Faceness-WIDER [68]	71.3	63.4	45.6	-	-
LDCF+ [69]	79.0	76.9	52.2	3	CPU
YOLO-face (darknet-53) [49]	82.5	77.8	52.5	45	GTX 1080 Ti
Multitask Cascade CNN [40]	84.8	82.5	59.8	16	Titan Black
ScaleFace [45]	86.8	86.7	77.2	4	Titan X
YOLO-face (deeper darknet) [49]	89.9	87.2	69.3	38	GTX 1080 Ti
DSFD (ResNet50) [47]	93.7	92.2	81.8	22	Tesla P40
Face R-FCN [41]	94.7	93.5	87.4	3	Tesla K80
FCOS [28] (from [48])	95.0	90.6	55.0	-	-
FAN [43]	95.2	94.0	90.0	11	Titan Xp
FoveaBox [29] (from [48])	95.6	93.5	67.8	11	Tesla V100
DBCFace [48]	95.8	95.0	90.3	7	GTX 1080 Ti
FDNet [70]	95.9	94.5	87.9	-	-
PyramidBox [44]	96.1	95.0	88.9	3	Titan RTX
DSFD (ResNet152) [47]	96.6	95.7	90.4	-	-
RetinaFace [46]	96.9	96.1	91.8	13	Tesla P40
SFPD [ours]	80.5	73.6	51.3	40 *	RTX 2080 Ti

All frameworks were trained exclusively with face annotations from the WIDER Face train set; the inference time was determined with a batch size of one; * at our SFPD method denotes that both faces and persons are detected within this inference time.

Figure 3. Example detections of SFPD on the WIDER Face [7] validation set: Red bounding boxes indicate detected faces; green bounding boxes detected persons.

While the performance gap between the "easy" and "medium" subset is not very big, SFPD has to deal with a drop of more than 22 percent between "medium" and "hard". This can be explained by the fact that the "hard" subset mainly consists of small faces and these cause difficulties for SFPD. However, this high drop in performance can be observed for almost all algorithms with double-digit frame rates. Especially, the detection of very small objects is difficult to implement with only few runtime losses. However, it was not the goal of our SFPD approach to perform very well with small faces, because it mainly targets

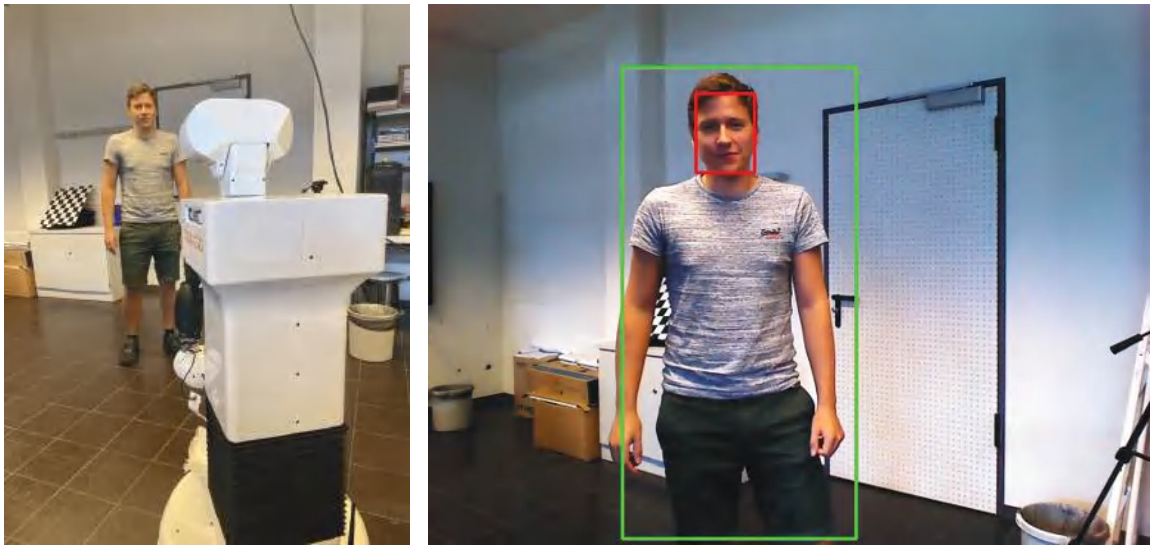
close-range human–robot interaction scenarios. This was achieved by implementing the proposed network on a mobile robot and it is successfully used for real-time human–robot interaction in a demo application (see Figure 5).



(**a**) easy (**b**) medium (**c**) hard

Figure 4. Precision–recall curves of our SFPD network and other detectors on the WIDER Face validation set: (**a**) easy, (**b**) medium and (**c**) hard.

In conclusion, SFPD achieves good results in both person and face detection. Furthermore, it was important to us that the two detection tasks are executed with high frame rate, so that additional modules such as face recognition can be integrated into the pipeline and still real-time processing of the entire system is guaranteed. This goal could be achieved with a frame rate of 40 fps for the detection of faces and persons. The main advantages of our SFPD are that it detects both faces and persons simultaneously and reaches high framerates with good detection performance for both tasks. Compared to all other models, SFPD is either faster or more reliable in terms of detection performance. Thus, SFPD represents the optimal network for real-time human–robot interaction applications.



(**a**) exterior view (**b**) interior view

Figure 5. Our proposed SFPD network implemented on a mobile robot system for human–robot interaction in a demo application: (**a**) exterior view and (**b**) interior view of the robot.

4.4. Limitations

Although our SFPD network shows good results on the test datasets, especially in relation to the required computational time, there are still some limitations regarding the recognition performance.

SFPD has difficulties to detect very small objects. This can be explained by the fact that they often consist of only a few pixels and it is therefore difficult to extract meaningful features which are necessary for correct detection. This is particularly evident in the results on the WIDER Face validation "hard" subset which consists mainly of small faces. A possible solution approach would be to scale the input images for the network to a larger format so that the individual objects would comprise more pixels. However, this would have negative effects on the runtime of the CNN, which was no option due to the real-time requirements of the targeted human–robot interaction application.

Another difficulty of SFPD is that it has problems to separate several objects that are located close to each other, especially, in crowded scenes. For highly overlapping objects, this effect will be further intensified. An approach to solve this problem would be to lower the NMS threshold, so that, e.g., two strongly overlapping objects can be recognized as two objects and none is rejected because of a too high IoU between the boxes. However, this would result in multiple detections of the same objects.

Despite these limitations, we believe that the SFPD algorithm offers a good trade-off between detection performance and inference time making it a good detection framework for many real-world applications. In particular, it is suitable for human–robot interaction, which requires real-time processing and does not suffer from the limitations detecting very small faces and handling crowded scenes (due to close-range interaction with a quite small number of people).

Future work may address improving performance with low-quality images that may occur, e.g., due to bad lighting or low-cost camera hardware. This may be done by collecting additional low-quality training data or applying data augmentation that degrades the image quality.

5. Conclusions

Our newly developed SFPD approach is able to detect faces and persons simultaneously in real-time. For this purpose, it employs a joint CNN backbone with shared feature maps and separate detection layers for each task. The difficulty for training this network was the fact that available datasets only contain annotations of bounding box coordinates for one of the two detection tasks. By applying a special training procedure and by designing a custom multi-task loss function, this problem could be addressed during training and a completely end-to-end trainable framework was created. Thereby, SFPD does not need any auxiliary steps during training, such as pre-training individual network parts, splitting up the training process, freezing single network layers or creating additional annotations for datasets, as it is mostly the case with other multi-task learning networks. SFPD performs well against other algorithms. Person detection was evaluated on the Pascal VOC datasets and face detection on the WIDER Face dataset. Moreover, our approach is capable of processing 40 fps. It is superior to all other algorithms in at least one of processing speed, detection performance or providing both face and person detections. Because of the good trade-off between detection performance for both detection tasks and inference time, SFPD represents a useful framework especially for close-range human–robot interaction scenarios and many more real-world applications.

Author Contributions: Conceptualization, M.-A.F. and P.W.; methodology, M.-A.F. and P.W.; software, M.-A.F., P.W. and A.K.; validation, M.-A.F., P.W. and A.K.; formal analysis, M.-A.F. and P.W.; investigation, M.-A.F., P.W. and A.K.; resources, M.-A.F., P.W. and A.K.; data curation, M.-A.F., P.W. and A.K.; writing—original draft preparation, M.-A.F. and P.W.; writing—review and editing, M.-A.F., P.W., A.K. and A.A.-H.; visualization, M.-A.F. and P.W.; supervision, A.A.-H.; project administration, A.A.-H.; funding acquisition, A.A.-H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the German Federal Ministry of Education and Research (BMBF) under Grant Nos. 03ZZ0448L (RoboAssist), 03ZZ0470 (HuBA), and 03ZZ04X02B (RoboLab) within the Zwanzig20 Alliance 3Dsensation. We acknowledge support for the Article Processing Charge by the Open Access Publication Fund of Magdeburg University. The responsibility for the content lies solely with the authors.

Institutional Review Board Statement: Ethical review and approval were waived for this study, due to the use of public databases, which were conducted according to the guidelines of the Declaration of Helsinki and approved by the relevant review boards. We complied with the terms of use of the databases regarding the publication of data.

Informed Consent Statement: According to the documentation of the used public databases, informed consent was obtained from all subjects involved.

Data Availability Statement: The WIDER Face dataset can be obtained at http://shuoyang1213.me/WIDERFACE/ (accessed on 3 August 2021). The Pascal VOC datasets can be obtained at http://host.robots.ox.ac.uk (accessed on 3 August 2021).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CNN	convolutional neural network
DPM	deformable parts model
fps	frames per second
GIoU	generalized intersection over union
GPU	graphics processing unit
HCI	human-computer interaction
HOG	histogram of oriented gradient
IoU	intersection over union
L	loss
L_{conf}	confidence loss
L_{reg}	regression loss
MTL	multi-task learning
NMS	non-maximum suppression
ReLU	rectified linear unit
SFPD	simultaneous face and person detection
SGD	stochastic gradient descent

References

1. Wang, M.; Deng, W. Deep face recognition: A survey. *arXiv* **2018**, arXiv:1804.06655.
2. Werner, P.; Saxen, F.; Al-Hamadi, A.; Yu, H. Generalizing to unseen head poses in facial expression recognition and action unit intensity estimation. In Proceedings of the IEEE International Conference on Automatic Face & Gesture Recognition (FG), Lille, France, 14–18 May 2019. [CrossRef]
3. Werner, P.; Saxen, F.; Al-Hamadi, A. Facial action unit recognition in the wild with multi-task CNN self-training for the EmotioNet challenge. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 14–19 June 2020; pp. 1649–1652. [CrossRef]
4. Handrich, S.; Waxweiler, P.; Werner, P.; Al-Hamadi, A. 3D human pose estimation using stochastic optimization in real time. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 555–559. [CrossRef]
5. Saxen, F.; Werner, P.; Handrich, S.; Othman, E.; Dinges, L.; Al-Hamadi, A. Face attribute detection with MobileNetV2 and NasNet-Mobile. In Proceedings of the International Symposium on Image and Signal Processing and Analysis (ISPA), Dubrovnik, Croatia, 23–25 September 2019; pp. 176–180. [CrossRef]
6. Zhang, H.B.; Zhang, Y.X.; Zhong, B.; Lei, Q.; Yang, L.; Du, J.X.; Chen, D.S. A comprehensive survey of vision-based human action recognition methods. *Sensors* **2019**, *19*, 1005. [CrossRef] [PubMed]
7. Yang, S.; Luo, P.; Loy, C.C.; Tang, X. WIDER Face: A face detection benchmark. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 5525–5533.
8. Everingham, M.; Van Gool, L.; Williams, C.K.I.; Winn, J.; Zisserman, A. The Pascal visual object classes (VOC) challenge. *Int. J. Comput. Vis.* **2010**, *88*, 303–338. [CrossRef]

9. Everingham, M.; Eslami, S.M.A.; Van Gool, L.; Williams, C.K.I.; Winn, J.; Zisserman, A. The Pascal visual object classes challenge: A retrospective. *Int. J. Comput. Vis.* **2015**, *111*, 98–136. [CrossRef]
10. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. YOLOv4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
11. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Kai Li.; Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Miami, FL, USA, 20–25 June 2009; pp. 248–255. [CrossRef]
12. Uijlings, J.R.R.; van de Sande, K.E.A.; Gevers, T.; Smeulders, A.W.M. Selective search for object recognition. *Int. J. Comput. Vis.* **2013**, *104*, 154–171. [CrossRef]
13. Zitnick, C.L.; Dollár, P. Edge Boxes: Locating object proposals from edges. In Proceedings of the European Conference on Computer Vision (ECCV), Zurich, Switzerland, 6–12 September 2014; pp. 391–405.
14. Fang, Z.; Cao, Z.; Xiao, Y.; Zhu, L.; Yuan, J. Adobe Boxes: Locating object proposals using object adobes. *IEEE Trans. Image Process.* **2016**, *25*, 4116–4128. [CrossRef]
15. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014; pp. 580–587. [CrossRef]
16. Girshick, R. Fast R-CNN. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015.
17. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards real-time object detection with region proposal Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *4*, 1137–1149. [CrossRef]
18. Dai, J.; Li, Y.; He, K.; Sun, J. R-FCN: Object detection via region-based fully convolutional networks. *arXiv* **2016**, arXiv:1605.06409.
19. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask R-CNN. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017.
20. Pang, J.; Chen, K.; Shi, J.; Feng, H.; Ouyang, W.; Lin, D. Libra R-CNN: Towards balanced learning for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 821–830. [CrossRef]
21. Zhang, H.; Hu, Z.; Hao, R. Joint information fusion and multi-scale network model for pedestrian detection. *Vis. Comput.* **2020**, 1–10. [CrossRef]
22. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; [CrossRef]
23. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017. [CrossRef]
24. Redmon, J.; Farhadi, A. YOLOv3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
25. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. SSD: Single Shot MultiBox Detector. In Proceedings of the European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 11–14 October 2016; pp. 21–37. [CrossRef]
26. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal loss for dense object detection. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017. [CrossRef]
27. Tan, M.; Pang, R.; Le, Q.V. EfficientDet: Scalable and efficient object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020. [CrossRef]
28. Tian, Z.; Shen, C.; Chen, H.; He, T. FCOS: Fully convolutional one-stage object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019. [CrossRef]
29. Kong, T.; Sun, F.; Liu, H.; Jiang, Y.; Li, L.; Shi, J. FoveaBox: Beyound anchor-based object detection. *IEEE Trans. Image Process.* **2020**, *29*, 7389–7398. [CrossRef]
30. Viola, P.; Jones, M.J. Robust real-time face detection. *Int. J. Comput. Vis.* **2004**, *57*, 137–154. [CrossRef]
31. Freund, Y.; Schapire, R.E. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **1997**, *55*, 119–139. [CrossRef]
32. Felzenszwalb, P.F.; Girshick, R.B.; McAllester, D.; Ramanan, D. Object detection with discriminatively trained part-based models. *IEEE Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 1627–1645. [CrossRef]
33. Zhu, X.; Ramanan, D. Face detection, pose estimation, and landmark localization in the wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, USA, 16–21 June 2012; pp. 2879–2886.
34. Yan, J.; Lei, Z.; Wen, L.; Li, S.Z. The fastest deformable part model for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014; pp. 2497–2504. [CrossRef]
35. Mathias, M.; Benenson, R.; Pedersoli, M.; Gool, L. Face detection without bells and whistles. In Proceedings of the European Conference on Computer Vision (ECCV), Zurich, Switzerland, 6–12 September 2014.
36. Dalal, N.; Triggs, B. Histograms of oriented gradients for human detection. *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.* **2005**, *1*, 886–893. [CrossRef]
37. Ranjan, R.; Sankaranarayanan, S.; Bansal, A.; Bodla, N.; Chen, J.C.; Patel, V.M.; Castillo, C.D.; Chellappa, R. Deep learning for understanding faces: Machines may be just as good, or better, than humans. *IEEE Signal Process. Mag.* **2018**, *35*, 66–83. [CrossRef]
38. Li, H.; Lin, Z.; Shen, X.; Brandt, J.; Hua, G. A convolutional neural network cascade for face detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 5325–5334. [CrossRef]

39. Rothe, R.; Guillaumin, M.; Gool, L. Non-maximum suppression for object detection by passing messages between windows. In Proceedings of the Asian Conference on Computer Vision (ACCV), Singapore, 1–5 November 2014.
40. Zhang, K.; Zhang, Z.; Li, Z.; Qiao, Y. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Process. Lett.* **2016**, *23*, 1499–1503. [CrossRef]
41. Wang, Y.; Ji, X.; Zhou, Z.; Wang, H.; Li, Z. Detecting faces using region-based fully convolutional networks. *arXiv* **2017**, arXiv:1709.05256.
42. Zhang, S.; Zhu, X.; Lei, Z.; Shi, H.; Wang, X.; Li, S. S^3FD: Single shot scale-invariant face detector. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 192–201.
43. Wang, J.; Yuan, Y.; Yu, G. Face Attention Network: An effective face detector for the occluded faces. *arXiv* **2017**, arXiv:1711.07246.
44. Tang, X.; Du, D.K.; He, Z.; Liu, J. PyramidBox: A context-assisted Single Shot Face Detector. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
45. Yang, S.; Xiong, Y.; Loy, C.C.; Tang, X. Face detection through scale-friendly deep convolutional networks. *arXiv* **2017**, arXiv:1706.02863.
46. Deng, J.; Guo, J.; Ververas, E.; Kotsia, I.; Zafeiriou, S. RetinaFace: Single-shot multi-level face localisation in the wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 5202–5211. [CrossRef]
47. Li, J.; Wang, Y.; Wang, C.; Tai, Y.; Qian, J.; Yang, J.; Wang, C.; Li, J.; Huang, F. DSFD: Dual shot face detector. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 5055–5064. [CrossRef]
48. Li, X.; Lai, S.; Qian, X. DBCFace: Towards PURE convolutional neural network face detection. *IEEE Trans. Circuits Syst. Video Technol.* **2021**. [CrossRef]
49. Chen, W.; Huang, H.; Peng, S.; Zhou, C.; Zhang, C. YOLO-face: A real-time face detector. *Vis. Comput.* **2020**, *37*, 805–813. [CrossRef]
50. Thung, K.H.; Wee, C.Y. A brief review on multi-task learning. *Multimed. Tools Appl.* **2018**, *77*, 29705–29725. [CrossRef]
51. Caruana, R. Multitask Learning. *Encycl. Mach. Learn. Data Min.* **1998**, *28*, 41–75. [CrossRef]
52. Dehghan, A.; Ortiz, E.G.; Shu, G.; Masood, S.Z. DAGER: Deep age, gender and emotion recognition using convolutional neural network. *arXiv* **2017**, arXiv:1702.04280.
53. Ranjan, R.; Patel, V.M.; Chellappa, R. HyperFace: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 121–135. [CrossRef]
54. Ranjan, R.; Sankaranarayanan, S.; Castillo, C.D.; Chellappa, R. An All-In-One convolutional neural network for face analysis. In Proceedings of the IEEE International Conference on Automatic Face & Gesture Recognition (FG), Washington, DC, USA, 30 May–3 June 2017; pp. 17–24. [CrossRef]
55. Levi, G.; Hassner, T. Age and gender classification using convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Boston, MA, USA, 7–12 June 2015; pp. 34–42.
56. Zhang, Z.; Luo, P.; Loy, C.C.; Tang, X. Facial landmark detection by deep multi-task learning. In Proceedings of the European Conference on Computer Vision (ECCV), Zurich, Switzerland, 6–12 September 2014.
57. Gkioxari, G.; Hariharan, B.; Girshick, R.B.; Malik, J. R-CNNs for pose estimation and action detection. *arXiv* **2014**, arXiv:1406.5212.
58. Chen, D.; Ren, S.; Wei, Y.; Cao, X.; Sun, J. Joint cascade face detection and alignment. In Proceedings of the European Conference on Computer Vision (ECCV), Zurich, Switzerland, 6–12 September 2014.
59. Saxen, F.; Handrich, S.; Werner, P.; Othman, E.; Al-Hamadi, A. Detecting arbitrarily rotated faces for face analysis. In Proceedings of the IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 22–25 September 2019; pp. 3945–3949. [CrossRef]
60. He, K.; Fu, Y.; Xue, X. A jointly learned deep architecture for facial attribute analysis and face detection in the wild. *arXiv* **2017**, arXiv:1707.08705.
61. Wu, H.; Zhang, K.; Tian, G. Simultaneous face detection and pose estimation using convolutional neural network cascade. *IEEE Access* **2018**, *6*, 49563–49575. [CrossRef]
62. Cipolla, R.; Gal, Y.; Kendall, A. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 7482–7491. [CrossRef]
63. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2015**, arXiv:1409.1556.
64. Glorot, X.; Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS), Sardinia, Italy, 13–15 May 2010.
65. Rezatofighi, H.; Tsoi, N.; Gwak, J.; Sadeghian, A.; Reid, I.; Savarese, S. Generalized intersection over union: A metric and a loss for bounding box regression. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 658–666. [CrossRef]
66. Lin, T.Y.; Maire, M.; Belongie, S.J.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft COCO: Common objects in context. In Proceedings of the European Conference on Computer Vision (ECCV), Zurich, Switzerland, 6–12 September 2014.
67. Yang, B.; Yan, J.; Lei, Z.; Li, S.Z. Aggregate channel features for multi-view face detection. In Proceedings of the IEEE International Joint Conference on Biometrics, Clearwater, FL, USA, 29 September–2 October 2014; pp. 1–8. [CrossRef]

68. Yang, S.; Luo, P.; Loy, C.C.; Tang, X. From facial parts responses to face detection: A deep learning approach. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 3676–3684. [CrossRef]
69. Ohn-Bar, E.; Trivedi, M.M. To boost or not to boost? On the limits of boosted trees for object detection. In Proceedings of the International Conference on Pattern Recognition (ICPR), Cancun, Mexico, 4–8 December 2016; pp. 3350–3355. [CrossRef]
70. Zhang, C.; Xu, X.; Tu, D. Face detection using improved Faster RCNN. *arXiv* **2018**, arXiv:1802.02142.

sensors

MDPI

Communication

Stochastic Memristive Interface for Neural Signal Processing

Svetlana A. Gerasimova [1,2], Alexey I. Belov [2], Dmitry S. Korolev [2], Davud V. Guseinov [2], Albina V. Lebedeva [1], Maria N. Koryazhkina [3], Alexey N. Mikhaylov [2], Victor B. Kazantsev [1,4,5] and Alexander N. Pisarchik [3,4,6,*]

1 Institute of Biology and Biomedicine, National Research Lobachevsky State University of Nizhny Novgorod, 603950 Nizhny Novgorod, Russia; gerasimova@neuro.nnov.ru (S.A.G.); lebedeva@neuro.nnov.ru (A.V.L.); kazantsev@neuro.nnov.ru (V.B.K.)
2 Research Institute and Technology, National Research Lobachevsky State University of Nizhny Novgorod, 603950 Nizhny Novgorod, Russia; belov@nifti.unn.ru (A.I.B.); dmkorolev@phys.unn.ru (D.S.K.); guseinov@phys.unn.ru (D.V.G.); mian@nifti.unn.ru (A.N.M.)
3 Research and Educational Center "Physics of Solid State Nanostructures", National Research Lobachevsky State University of Nizhny Novgorod, 603950 Nizhny Novgorod, Russia; mahavenok@mail.ru
4 Laboratory of Neuroscience and Cognitive Technology, Innopolis University, 420500 Innopolis, Russia
5 Center for Neurotechnology and Machine Learning, Immanuel Kant Baltic Federal University, 236016 Kaliningrad, Russia
6 Center for Biomedical Technology, Universidad Politécnica de Madrid, Pozuelo de Alarcón, 28223 Madrid, Spain
* Correspondence: alexander.pisarchik@ctb.upm.es

Abstract: We propose a memristive interface consisting of two FitzHugh–Nagumo electronic neurons connected via a metal–oxide ($Au/Zr/ZrO_2(Y)/TiN/Ti$) memristive synaptic device. We create a hardware–software complex based on a commercial data acquisition system, which records a signal generated by a presynaptic electronic neuron and transmits it to a postsynaptic neuron through the memristive device. We demonstrate, numerically and experimentally, complex dynamics, including chaos and different types of neural synchronization. The main advantages of our system over similar devices are its simplicity and real-time performance. A change in the amplitude of the presynaptic neurogenerator leads to the potentiation of the memristive device due to the self-tuning of its parameters. This provides an adaptive modulation of the postsynaptic neuron output. The developed memristive interface, due to its stochastic nature, simulates a real synaptic connection, which is very promising for neuroprosthetic applications.

Keywords: memristive device; neuron-like oscillator; stochastic dynamics; synchronization; neuromorphic circuit; FitzHugh–Nagumo neuron

Citation: Gerasimova, S.A.; Belov, A.I.; Korolev, D.S.; Guseinov, D.V.; Lebedeva, A.V.; Koryazhkina, M.N.; Mikhaylov, A.N.; Kazantsev, V.B.; Pisarchik, A.N. Stochastic Memristive Interface for Neural Signal Processing. *Sensors* **2021**, *21*, 5587. https://doi.org/10.3390/s21165587

Academic Editor: Manuel José Cabral dos Santos Reis

Received: 2 July 2021
Accepted: 16 August 2021
Published: 19 August 2021

1. Introduction

The design of compact neuromorphic systems, including micro- and nanochips, capable of reproducing information and computational functions of brain cells is a great challenge of modern science and technology. Such systems are of interest for both fundamental research in the field of nonlinear dynamics and the synchronization of complex systems [1–7], as well as medical applications in the devices for monitoring and stimulating brain activity in the framework of neuroprosthetic tasks [8–10]. Due to their importance, memristive devices have recently become the subject of intense research, especially in the area of neuromorphic and neurohybrid applications [11–17]. Neuromorphic technologies are especially relevant for intelligent adaptive automatic control systems—biorobots. It is also worth noting that the construction and creation of electronic neurons and synapses (connections between neurons) based on thin-film memristive nanostructures is a fast-growing area of interdisciplinary research in the development of neuromorphic systems [18–20].

The history of neuromorphic technologies began in the late 1980s with the emergence of computation machines, and since then, significant advances have been achieved in elec-

tronics, physics of micro- and nanostructures, and solid-state nanoelectronics. The careful development of neuron-like electrical circuits made it possible to reproduce basic neural behaviors, such as resting, spiking, and bursting dynamics, as well as more sophisticated regimes, including chaos and multistability [21–25].

A memristive device is usually based on the Chua's model [19], which is an element of an electrical circuit capable of changing resistance depending on an electrical signal entering its input. In recent decades, various thin-film memristive nanostructures have been created. They are capable of changing their conductivity under the action of a pulsed signal [26,27], which makes the memristor an almost ideal electronic analogue of a synapse [13]. A synapse is known to be a communication channel between neurons that provides unidirectional signal transmission from a transmitting (presynaptic) neuron to a receiving (postsynaptic) neuron. This communication channel ensures the propagation of a nerve impulse along the axon of the transmitting cell.

The synaptic communication results in synchronization of postsynaptic and presynaptic neurons. Neural synchronization was extensively studied using various mathematical models and described in terms of periodic solutions [3,6,28–35]. Such artificial synapses were implemented as electronic circuits that convert pulses of presynaptic voltage into postsynaptic currents with some synaptic amplification. Different strategies were used for the hardware implementation of synaptic circuits, e.g., an optical interface between electronic neurons [4,5,7].

Recent advances in nanotechnology allowed for miniaturization of artificial synapses by creating memristive nanostructures that mimic dynamics of real synapses. Among various candidates for the role of electronic synapses, memristive devices have a great potential for implementing massive parallelism and three-dimensional integration in order to achieve good efficiency per unit volume [36–38]. In this regard, it is important to create a memristor-based neuromorphic system capable of processing neuron-like signals.

Recently, the interaction between electronic neurons through a metal-oxide memristive device was successfully implemented in hardware [39]. The prerequisite for such a device was the study of the interaction of Van der Pol generators via a memristor [40]. Later, a significant effort was invested in theoretical research to study synchronization between neuron-like generators connected through a memristive device [14,41]. However, to the best of our knowledge, experimental studies of the dynamics of FitzHugh–Nagumo (FHN) neurons connected by a memristive synapse have not yet been carried out. We believe that the creation of neuromorphic memristive systems will lead to the production of simple and compact neuroelements based on memristive devices capable of imitating the electrophysiological behavior of real neurons.

At the same time, a memristive device made of metal oxides is of interest not only for experimental research, but also for theoretical studies. Neuromemristive models were found to exhibit complex dynamics, including chaos and chimeras [42,43], the study of which can contribute to the fundamental theory. On the other hand, many theoretical "memristive" neural models reported in the literature have nothing to do with the concept of memristive elements [44]. Therefore, the development of adequate mathematical models that can simulate real laboratory neuromemristive experiments is an actual problem.

Summarizing all the above, significant theoretical investigations of memristors and the possibility of their use as a part of neuromorphic systems were performed. In particular, not only dynamical were effects simulated, but also the simplest learning rules were implemented [45–52]. Currently, technologies are being developed to improve the characteristics of memristive devices in order to create reliable memristive networks capable of solving some mathematical tasks [53], classifying images [54–58], etc. [59–61]. Despite impressive theoretical results in the development of neuromorphic memristive systems, the experimental research of laboratory memristive devices, rather than their substitutes based on transistors or resistors as parts of dynamical systems, was not carried out because of high complexity of this task, which requires the cooperation of nanotechnologists, physicists, and neuroscientists.

In this work, we experimentally implement a memristive interface based on the metal–oxide nanostructure that acts as a synaptic interface connecting two electronic FHN neural generators. The interface allows for the analog simulation of the adaptive behavior and neural timing effects, which can be associated with synaptic plasticity. We also investigate the stochastic properties of the memristive device. For the first time, to the best of our knowledge, we perform an experimental study on such a memristive neural system and compare experimental results with numerical simulations.

2. Materials and Methods

In order to simulate neural dynamics, we explored two FHN neuron generators with cubic nonlinearity constructed using diodes [7,22]. The dynamics of the presynaptic FHN neuron was modeled by the normalized equations obtained with the Kirchhoff law [21] as follows:

$$\begin{array}{l} \frac{du_1}{dt} = f(u_1) - v_1 \\ \frac{dv_1}{dt} = \varepsilon(g(u_1) - v_1) - I_1 \end{array} \quad (1)$$

where u_1 is the membrane potential of the presynaptic neuron, ν_1 is the "recovery" variable related to the ion current, $f(u_1) = u_1 - {u_1}^3/3$ is the cubic nonlinearity, I_1 is the depolarization parameter characterizing the excitation threshold, and ε is a small coefficient. If $u_1 < 0$, the function $g(u_1) = \alpha u_1$, and if $u_1 \geq 0$, $g(u_1) = \beta u_1$ (α, β being the parameters that control, respectively, the shape and location of the ν-nullcline [22]).

The memristive device model was developed based on a standard approach to reflect the dynamical response of a memristor to electrical stimulation. The model describes a change in resistance, similar to potentiation and depression, based on physical laws identified in experiments [62]. The memristor model is given by the complex function:

$$\begin{array}{l} j = wj_{lin} + (1-w)j_{nonlin} \\ j_{lin} = u_1/\rho \\ j_{nonlin} = u_1 \exp(b\sqrt{u_1} - E_b) \\ w(u_1) = A \exp(-\frac{E_m - \alpha_1 u_1}{kT}) \end{array} \quad (2)$$

This approach supposes the introduction of internal state variable w, which is determined by the fraction of the insulator region occupied by filaments. The change in this state is associated with the processes of migration of oxygen ions (vacancies) with the height of the effective migration barrier E_m. In turn, the migration is provided by the Joule heating kT and applied electric voltage u_1. The total current density j through the memristor is the sum of the linear j_{lin} and nonlinear j_{nonlin} components. The former corresponds to ohmic conductivity with resistivity ρ, whereas the latter is determined by the transport of charge carriers through defects in the regions of the insulator not occupied by filaments (including those in the filament rupture region). It was previously found that, in the insulating state of the studied ZrO_2-based memristive devices, the current transport is implemented by the Poole–Frenkel mechanism with an effective barrier E_b [62]. The smooth transition between high- and low-resistance states (HRS and LRS, correspondingly) is determined by the dynamic contribution to the total current of the conductive filaments and, therefore, the state variable. In Equation (2), b, α_1, and A are coefficients derived from experimental data. In our numerical simulations we used the Runge–Kutta integration methods for stochastic differential equations in Matlab [63–65].

In order to compare the experimentally observed dynamics of the memristive device with the results of numerical simulations, we needed to take into account stochasticity of microscopic processes leading to a change in the internal state w of the dynamical system. Random fluctuations of the normal distribution were added to energy barrier E_m for ion hopping (dispersion 10%), energy barrier E_b for electron jumps in the Poole–Frenkel conduction mechanism in the HRS (dispersion 1%), and ohmic resistance ρ of the structure in the LRS (dispersion 10%). This led to the scattering of the experimental current–voltage characteristics. The finite spread of the switching voltages is mainly related

to the stochasticity of the energy barrier for ions, whereas the change in the resistive states from cycle to cycle is associated with the electron transport stochasticity.

One-way communication between two neurons through the memristive device was modeled by the following equations:

$$\begin{array}{l} \frac{du_1}{dt} = f(u_1) - v_1 \\ \frac{dv_1}{dt} = \varepsilon_1(g(u_1) - v_1) - I_1 \\ \frac{du_2}{dt} = f(u_2) - v_2 + j(u_1)Sd \\ \frac{dv_2}{dt} = \varepsilon_2(g(u_2) - v_2) - I_2 \end{array} \quad (3)$$

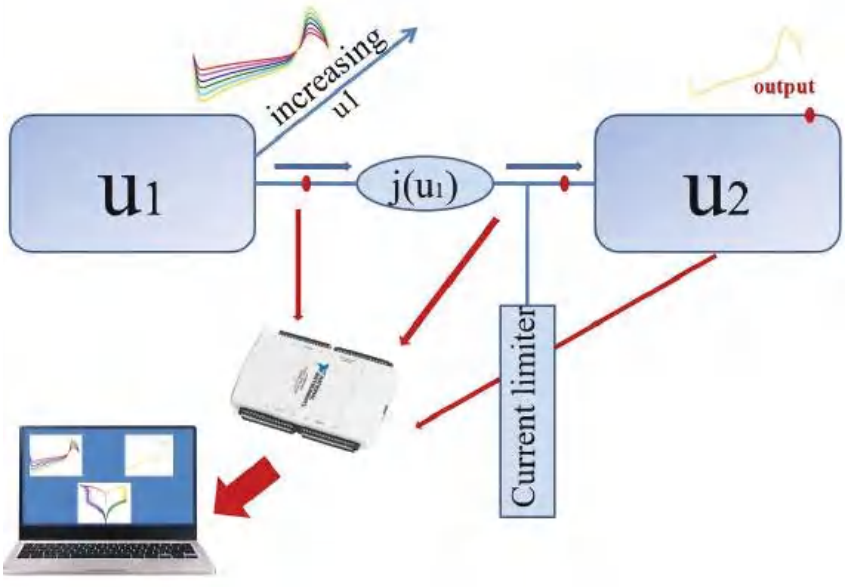
where d is the equivalent load resistance, $j(u_1)$ is the current density through the memristive device, S is the area of conductive filaments obtained from the experiment, and ε is a small recovery parameter. The signal from the presynaptic neural generator (u_1) was sent to the postsynaptic neural generator (u_2) through the memristive device.

Thus, the two neurogenerators were connected in such a way that part of the current $j(u_1)$ generated by the presynaptic neuron passed through the load resistor, which was connected in series with the memristive device, before reaching the postsynaptic neuron. The initial conditions and model parameters corresponded to the experimental conditions. In particular, both neural oscillators were initially in a self-oscillatory regime.

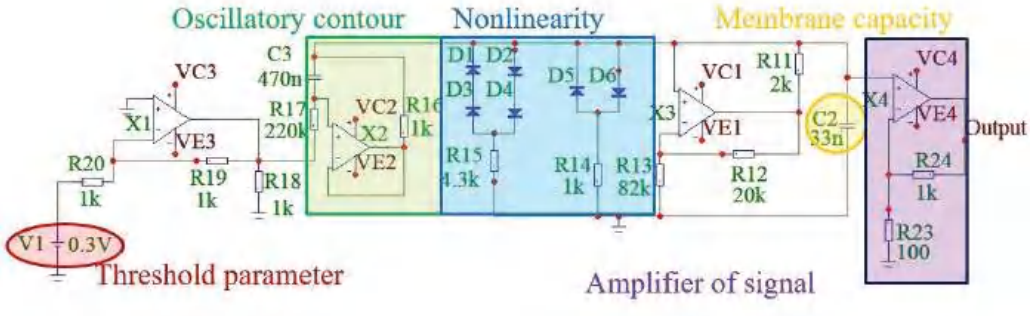
The designed neuromorphic circuit consisted of an FHN electronic circuit, a memristive device formed by the thin-film metal–oxide–metal nanostructure based on yttria-stabilized zirconia (Au/Zr/ZrO_2(Y)/TiN/Ti) [66], and a load resistor (Figure 1a). This memristive interface operated as follows. The electronic FHN neuron generated a pulse signal that affects the memristive device and thus modulates the oxidation and recovery of conductive filaments in the oxide film of the memristive device. The analog electronic FHN neuron consisted of the following blocks: an oscillatory contour unit, a nonlinearity unit, and an amplifier unit (see Figure 1b). The detailed design of this device is described in [7,22]. The FHN neural generator demonstrates the main qualitative features of neurodynamics: the presence of an excitability threshold and the existence of resting and spiking regimes. These regimes were controlled using a potentiometer. The spiking frequency was varied in the range of 10–150 Hz, the spike duration in the range of 10–25 ms, and the spike amplitude u_1 in the range of 1–6 V.

In this work, we used the National Instruments USB-6212 data acquisition system, which consists of a digital-to-analog converter (DAC) and two analog-to-digital converters (ADC). The data acquisition system was controlled using LabVIEW software. The prerecorded neuron-like signal was applied to a memristive device with a sampling frequency of 5 kHz via the DAC. The ADCs recorded the voltage drop across the memristive device and the load resistor, which made it possible to calculate the memristive device resistance in real time. The potential difference across the memristive device (R_m) and the load resistor (R_2) was digitized at a sampling frequency of 10 kHz. Matlab was used to analyze the results.

After testing and tuning, the neuron-like oscillators were connected through the memristive device. Both analog neurogenerators were turned in the oscillatory regime. Under the neuron-like signal action, the memristive device changed its state from high resistive to low resistive. The amplitude of the presynaptic neuron was adjusted by the potentiometer in order to obtain a frequency-locking regime between two oscillators.



(**a**)



(**b**)

Figure 1. The system description: (**a**) block diagram of the interaction between presynaptic (u_1) and postsynaptic (u_2) electronic neurons through a memristive device. The neurons are initially in an oscillatory regime. The output of the presynaptic neuron is increased during the experiment; (**b**) analog electrical circuit of the FitzHugh–Nagumo neuron. The inductance is implemented by the circuit with operational amplifier, cubic nonlinearity is set using diodes D1–D6, capacitor C2 is related to the capacitance of the neuron membrane, and potential V1 is associated with an equilibrium controlled by the power source.

3. Results and Discussion

The output signal of the presynaptic electronic neuron is shown in Figure 2a. This signal is applied to the memristive device. The used neuron-like signal (u_1) is asymmetric (the minimum voltage is -5 V and the maximum voltage is 4 V) due to the asymmetry of the current–voltage characteristic (*I*–*V* curves) of the memristive device. For a more detailed study of the effect of the neuron-like signal on the memristive device, the curve in Figure 2a is visually divided into four intervals with different colors. Each interval corresponds to a specific fragment of the *I*–*V* curves in Figure 2b. The *I*–*V* curves in Figure 2b display the switching between LRS and HRS. The RESET process (switching from LRS to HRS) occurs with a positive voltage and SET (switching from HRS to LRS) with a negative voltage. The

scattering of the *I–V* curves in Figure 2b results from random fluctuations applied to the memristor parameters E_m, E_b, and ρ. Figure 2c demonstrates the increase in the amplitude of the presynaptic neuron from 1.558 V to 4 V. Figure 2d shows that, even when exposed to a small amplitude signal of 2 V (purple curve), the memristive device can switch from HRS to LRS.

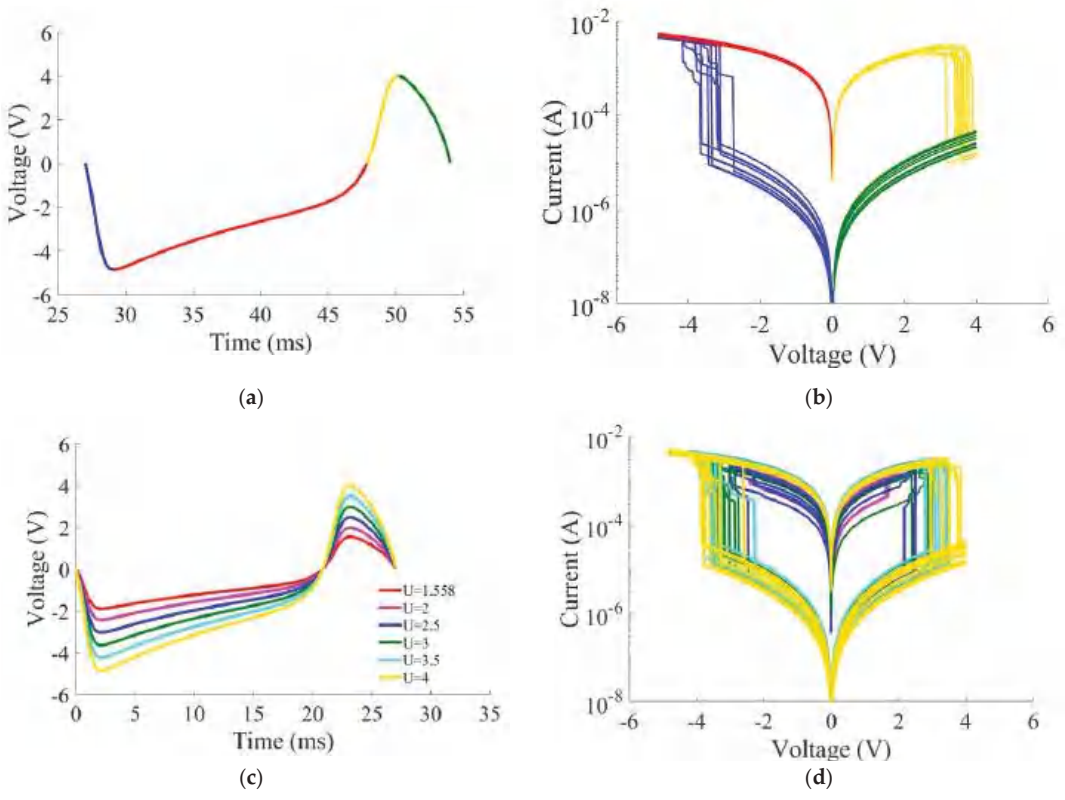


Figure 2. Experimental resistive switching in the response to a neuron-like signal: (**a**) neural-like pulse. The blue, red, yellow, and green colors show, respectively, the curve segments increasing from 0 to −5 V, from −5 to 0 V, from 0 to 4 V, and 4 to 0; (**b**) *I–V* curves. Each colored *I-V* section corresponds to a colored section of the input signal to memristor; (**c**) increasing amplitude of the neuron-like signal. The red, purple, blue, green, light blue, and yellow curve corresponds, respectively, to the peak amplitude of 1.558 V, 2 V, 2.5 V, 3V, 3.5 V, and 4 V; (**d**) resistive switching of the memristive device under the action of corresponding neuron-like signals on the I-V curves. Each colored *I–V* curve corresponds to a colored curve of the input signal to memristor.

The laboratory memristor demonstrates different responses to an input signal with a small stochastic spread. Figure 2d shows that, for one curve in Figure 2c, with the yellow curve used as an example, the numerical memristor model yields 10 possible curves (also a yellow color) with a small spread. The *I–V* curves in Figure 2d illustrate the effect of stochastic switching in the memristor response to the voltage signals of the corresponding amplitudes. Since memristor conductivity is adaptively changed according to the input signal, the memristive device demonstrates the property of plasticity.

There is a threshold value of the amplitude (u_1) of the neuron-like signal at which the memristor state switches at each spike. At high amplitudes of the input signal (u_1), the system enters a state of extreme resistance and does not respond to each spike anymore. The memristive device remains in this state. The switching degree strongly depends on the internal changes in the memristive device related to the interrelated transport phenom-

ena in oxide dielectrics, due to electric potential gradients, ion concentration, and local heating [67,68]. These reasons result in the partial recovery and oxidation of conducting filaments in the oxide film. The corresponding dynamical change in conductivity is limited by the applied voltage and leads to the modulation of the strength of neuron coupling and different types of synchronization. In the course of the study, the optimal coupling strength is $z = j(u_1)$; SR = (0.02–0.06) for 1:1 frequency-locking (Figure 3c) and z = (0.06–0.095) for intermittent synchronization (Figure 3d).

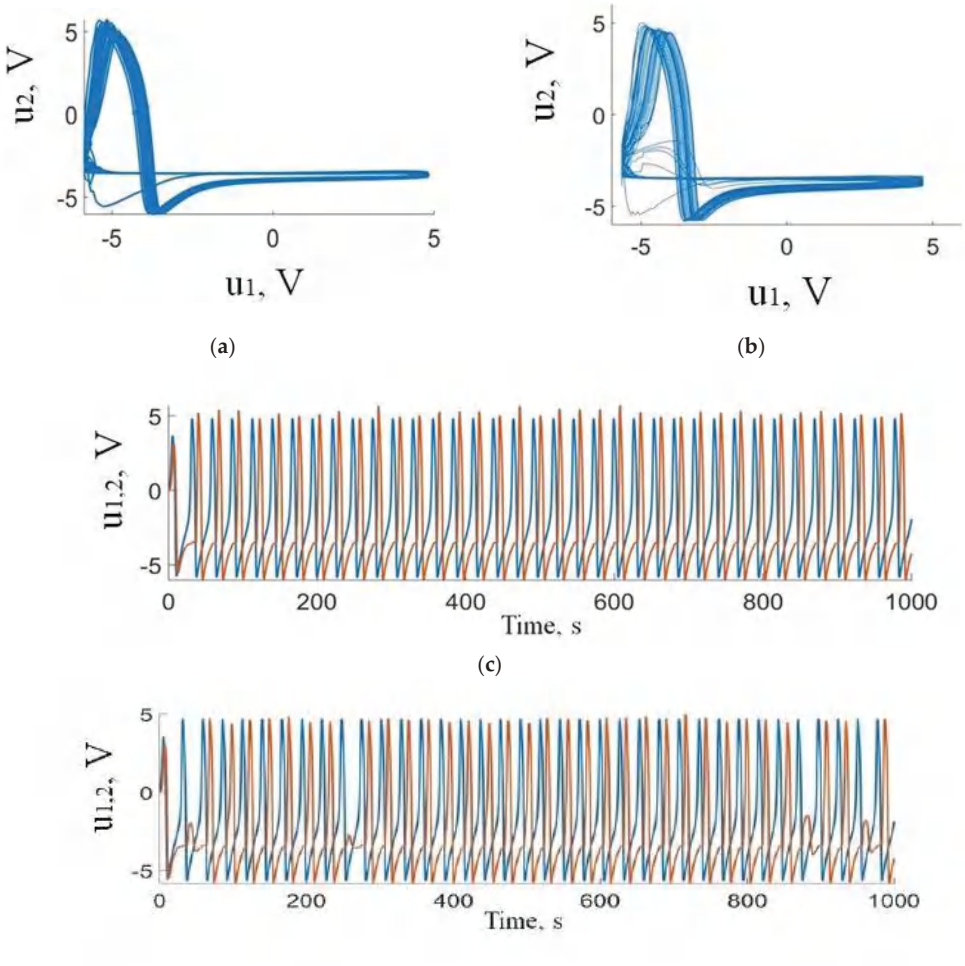


Figure 3. Results of numerical simulations of the dynamics of FHN neuron generators with memristive coupling: (**a**,**b**) phase portraits and (**c**,**d**) time series representing (**a**,**c**) 1:1 and (**b**,**d**) intermittent frequency-locking regimes. Blue and red curves show action potentials of presynaptic (u_1) and postsynaptic (u_2) neurons, respectively.

The experiments show that, when the amplitude of the presynaptic neuron u_1 is varied from 1.6 to 2 V, the oscillation frequencies of the coupled neurons are locked either as 2:1 (Figure 4a) or 3:1 (Figure 4b), i.e., the presynaptic neuron u_1 fires the postsynaptic neuron u_2 twice or thrice. This ratio can be randomly changed when chaotic synchronization is reached at higher voltage amplitudes. Although the phase portraits obtained numerically and experimentally do not completely match, the experiment confirms the diversity of

phase-locking regimes predicted by the model. Moreover, our model demonstrates dynamics close to the experimentally observed one, despite of to the first-order memristor model, if the stochasticity of switching is accounted for.

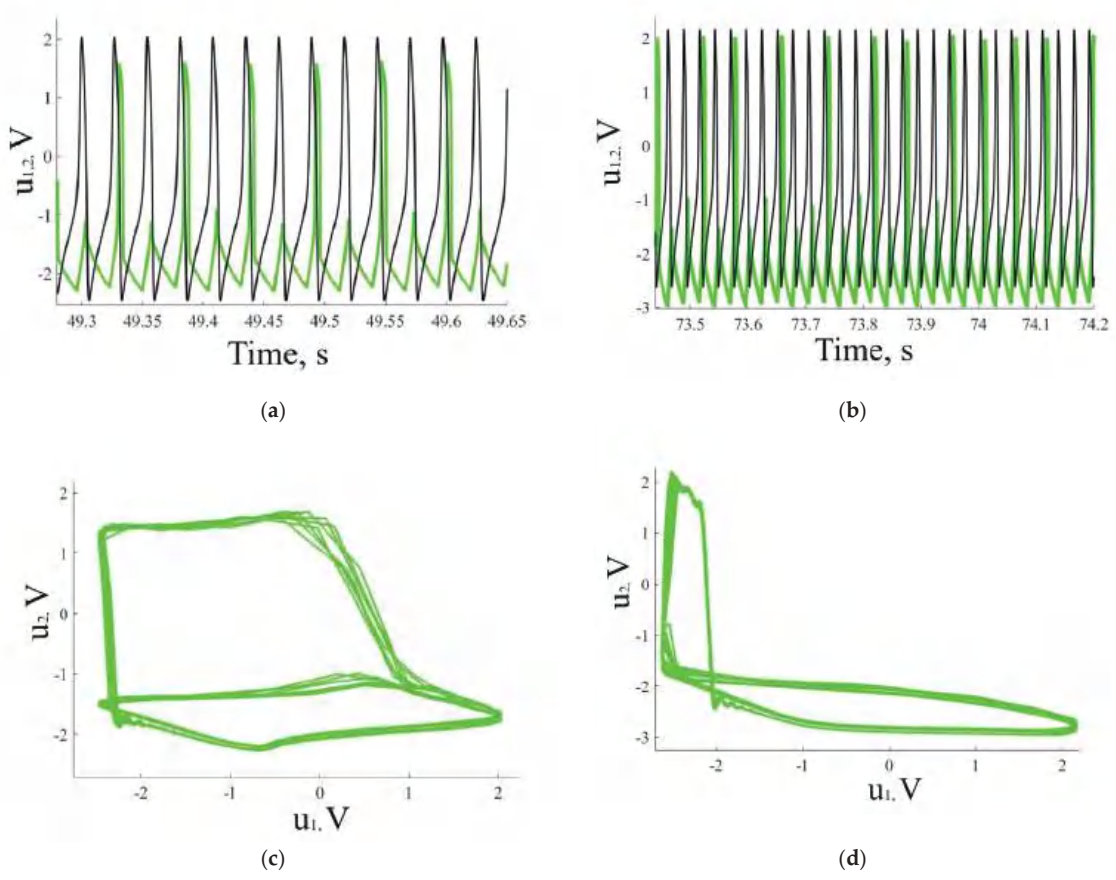


Figure 4. Experimental results demonstrating frequency-locking of FHN electronic neurons coupled by the memristive device: (**a**,**c**) time series and (**b**,**d**) phase portraits representing (**a**,**b**) 2:1 and (**c**,**d**) intermittent frequency-locking regimes. Black and green curves show action potential of presynaptic (u_1) and postsynaptic (u_2) neurons, respectively.

The stochasticity is an inalienable property of resistive-switching devices, enabling the so-called stochastic plasticity used to mimic neural synchrony in a simple electronic cognitive system [69]. To the best of our knowledge, the present work is the first attempt to study this important phenomenon both numerically and experimentally. In our case, the stochasticity is modeled through the introduction of fluctuations in the model parameters in a way similar to [70]. Recently, Agudov et al. [71] developed a more generic stochastic model of a memristive device that can be further used to adequately describe the observed complex dynamics of the proposed memristive interface. Another option is to use the deterministic, but at the same time higher-order memristor models based on two or more state variables in order to simulate the experimentally observed intermittency route to chaos [72].

4. Conclusions

In this work, we have studied the dynamics of two coupled FitzHugh–Nagumo neuron generators coupled through a memristive device of a metal–oxide type that adapts

the synaptic connection according to the amplitude of the presynaptic neuron oscillations. The stochastic switching of the memristive device from a high-resistance state to a low-resistance state is achieved by the variation of the internal parameters. Therefore, the memristive synaptic device demonstrates the property of stochastic plasticity. Different synchronous regimes were observed, including 1:1, 2:1, and 3:1 frequency-locking, intermittent synchronization, and more complex dynamics. Its relative compactness and high sensitivity make the proposed neuromemristive device very promising for biorobotics and other bioengineering applications [73].

Author Contributions: Conceptualization, S.A.G. and A.N.P.; methodology, S.A.G. and D.S.K.; software, S.A.G.; validation, S.A.G., A.N.P. and A.N.M.; formal analysis, A.I.B.; investigation, M.N.K. and D.V.G.; resources, A.N.M.; data curation, S.A.G., A.I.B., A.N.P. and A.N.M.; writing—original draft preparation, S.A.G., A.I.B., A.V.L. and V.B.K.; writing—review and editing, M.N.K., A.N.M., V.B.K. and A.N.P.; visualization, S.A.G.; supervision, V.B.K., A.N.M. and A.N.P.; project administration, A.N.M. and A.N.P.; funding acquisition, A.N.M. and A.N.P. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Russian Science Foundation (Project No. 21-11-00280). A.N.P. acknowledges the Lobachevsky University Competitiveness Program in the frame of the 5-100 Russian Academic Excellence Project.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Alombah, N.H.; Fotsin, H.; Romanic, K. Coexistence of multiple attractors, metastable chaos and bursting oscillations in a multiscroll memristive chaotic circuit. *Int. J. Bifurc. Chaos* **2017**, *27*, 1750067. [CrossRef]
2. Pikovsky, A.; Rosenblum, M.; Kurths, J. *Synchronization: A Universal Concept in Nonlinear Sciences*; Cambridge University Press: New York, NY, USA, 2001; p. 411.
3. Boccaletti, S.; Pisarchik, A.N.; del Genio, C.I.; Amann, A. *Synchronization: From Coupled Systems to Complex Networks*; Cambridge University Press: Cambridge, UK, 2018; p. 255.
4. Pisarchik, A.; Jaimes-Reátegui, R.; Sevilla-Escoboza, J.R.; López, J.H.G.; Kazantsev, V. Optical fiber synaptic sensor. *Opt. Lasers Eng.* **2011**, *49*, 736–742. [CrossRef]
5. Pisarchik, A.N.; Sevilla-Escoboza, R.; Jaimes-Reategui, R.; Huerta-Cuellar, G.; García-Lopez, J.H.; Kazantsev, V.B. Experimental implementation of a biometric laser synaptic sensor. *Sensors* **2013**, *13*, 17322–17331. [CrossRef]
6. Simonov, A.Y.; Gordleeva, S.Y.; Pisarchik, A.; Kazantsev, V.B. Synchronization with an arbitrary phase shift in a pair of synaptically coupled neural oscillators. *JETP Lett.* **2014**, *98*, 632–637. [CrossRef]
7. Gerasimova, S.A.; Gelikonov, G.V.; Pisarchik, A.N.; Kazantsev, V.B. Synchronization of optically coupled neural-like oscillators. *J. Commun. Technol. Electron.* **2015**, *60*, 900–903. [CrossRef]
8. Horch, K.W.; Kipke, D.R. *Neuroprosthetics Theory and Practice*, 2nd ed.; World Scientific: Singapore, 2017; Volume 8, p. 934.
9. Gerasimova, S.; Lebedeva, A.; Fedulina, A.; Koryazhkina, M.; Belov, A.; Mishchenko, M.; Matveeva, M.; Guseinov, D.; Mikhaylov, A.; Kazantsev, V.; et al. A neurohybrid memristive system for adaptive stimulation of hippocampus. *Chaos Solitons Fractals* **2021**, *146*, 110804. [CrossRef]
10. Hramov, A.E.; Maksimenko, V.A.; Pisarchik, A.N. Physical principles of brain-computer interfaces and their applications for rehabilitation, robotics and control of human brain states. *Phys. Rep.* **2021**, *918*, 1–133. [CrossRef]
11. Indiveri, G.; Linares-Barranco, B.; Hamilton, T.J.; van Schaik, A.; Etienne-Cummings, R.; Delbruck, T.; Liu, S.-C.; Dudek, P.; Häfliger, P.; Renaud, S.; et al. Neuromorphic silicon neuron circuits. *Front. Neurosci.* **2011**, *5*, 73. [CrossRef] [PubMed]
12. Kuzum, D.; Yu, S.; Wong, H.-S.P. Synaptic electronics: Materials, devices and applications. *Nanotechnology* **2013**, *24*, 382001. [CrossRef]
13. Bill, J.; Legenstein, R. A compound memristive synapse model for statistical learning through STDP in spiking neural networks. *Front. Neurosci.* **2014**, *8*, 412. [CrossRef] [PubMed]
14. Zhang, T.; Yin, M.; Lu, X.; Cai, Y.; Yang, Y.; Huang, R. Tolerance of intrinsic device variation in fuzzy restricted Boltzmann machine network based on memristive nano-synapses. *Nano Futur.* **2017**, *1*, 015003. [CrossRef]
15. Nair, M.V.; Muller, L.K.; Indiveri, G. A differential memristive synapse circuit for on-line learning in neuromorphic computing systems. *Nano Futur.* **2017**, *1*, 035003. [CrossRef]

16. Strukov, D.B. Tightening grip. *Nat. Mater.* **2018**, *17*, 293–295. [CrossRef]
17. Mikhaylov, A.; Pimashkin, A.; Pigareva, Y.; Gerasimova, S.; Gryaznov, E.; Shchanikov, S.; Zuev, A.; Talanov, M.; Lavrov, I.; Demin, V.; et al. Neurohybrid memristive CMOS-integrated systems for biosensors and neuroprosthetics. *Front. Neurosci.* **2020**, *14*, 358. [CrossRef]
18. Thomas, A. Memristor-based neural networks. *J. Phys. D Appl. Phys.* **2013**, *46*, 093001. [CrossRef]
19. Adamatzky, A.; Chua, L.O. *Memristor Networks*; Springer: Cham, Switzerland, 2014; p. 716.
20. Ge, R.; Wu, X.; Kim, M.; Shi, J.; Sonde, S.; Tao, L.; Zhang, Y.; Lee, J.C.; Akinwande, D. Atomristor: Nonvolatile resistance switching in atomic sheets of transition metal Dichalcogenides. *Nano Lett.* **2017**, *18*, 434–441. [CrossRef]
21. Binczak, S.; Jacquir, S.; Bilbault, J.-M.; Kazantsev, V.B.; Nekorkin, V.I. Experimental study of electrical FitzHugh–Nagumo neurons with modified excitability. *Neural Netw.* **2006**, *19*, 684–693. [CrossRef] [PubMed]
22. Shchapin, D. Dynamics of two neuronlike elements with inhibitory feedback. *J. Commun. Technol. Electron.* **2009**, *54*, 175–184. [CrossRef]
23. Adamchik, D.A.; Matrosov, V.V.; Semyanov, A.V.; Kazantsev, V.B. Model of self-oscillations in a neuron generator under the action of an active medium. *JETP Lett.* **2015**, *102*, 624–627. [CrossRef]
24. Mishchenko, M.A.; Bolshakov, D.I.; Matrosov, V.V. Instrumental implementation of a neuronlike generator with spiking and bursting dynamics based on a phase-locked loop. *Tech. Phys. Lett.* **2017**, *43*, 596–599. [CrossRef]
25. Pisarchik, A.N.; Jaimes-Reátegui, R.; García-Vellisca, M.A. Asymmetry in electrical coupling between neurons alters multistable firing behavior. *Chaos: Interdiscip. J. Nonlinear Sci.* **2018**, *28*, 033605. [CrossRef]
26. Gambuzza, L.V.; Frasca, M.; Fortuna, L.; Ntinas, V.; Vourkas, I.; Sirakoulis, G.C. Memristor crossbar for adaptive synchronization. *IEEE Trans. Circuits Syst. I Regul. Pap.* **2017**, *64*, 2124–2133. [CrossRef]
27. Guseinov, D.; Tetelbaum, D.; Mikhaylov, A.; Belov, A.; Shenina, M.; Korolev, D.; Antonov, I.; Kasatkin, A.; Gorshkov, O.; Okulich, E.; et al. Filamentary model of bipolar resistive switching in capacitor-like memristive nanostructures on the basis of yttria-stabilised zirconia. *Int. J. Nanotechnol.* **2017**, *14*, 604. [CrossRef]
28. Matrosov, V.V.; Kazantsev, V.B. Bifurcation mechanisms of regular and chaotic network signaling in brain astrocytes. *Chaos: Interdiscip. J. Nonlinear Sci.* **2011**, *21*, 023103. [CrossRef] [PubMed]
29. Matrosov, V.V.; Mishchenko, M.A.; Shalfeev, V.D. Neuron-like dynamics of a phase-locked loop. *Eur. Phys. J. Spéc. Top.* **2013**, *222*, 2399–2405. [CrossRef]
30. Selyutskiy, Y.D. On auto-oscillations of a plate in flow. In *AIP Conference Proceedings*; AIP Publishing LLC location: Melville, NY, USA, 2017; Volume 1798, p. 20139. [CrossRef]
31. Sausedo-Solorio, J.; Pisarchik, A. Synchronization of map-based neurons with memory and synaptic delay. *Phys. Lett. A* **2014**, *378*, 2108–2112. [CrossRef]
32. Sausedo-Solorio, J.M.; Pisarchik, A.N. Synchronization in network motifs of delay-coupled map-based neurons. *Eur. Phys. J. Spéc. Top.* **2017**, *226*, 1911–1920. [CrossRef]
33. Andreev, A.V.; Frolov, N.S.; Pisarchik, A.N.; Hramov, A.E. Chimera state in complex networks of bistable Hodgkin-Huxley neurons. *Phys. Rev. E* **2019**, *100*, 022224. [CrossRef] [PubMed]
34. Andreev, A.V.; Maksimenko, V.A.; Pisarchik, A.N.; Hramov, A.E. Synchronization of interacted spiking neuronal networks with inhibitory coupling. *Chaos Solitons Fractals* **2021**, *146*, 110812. [CrossRef]
35. Bashkirtseva, I.A.; Ryashko, L.B.; Pisarchik, A.N. Ring of map-based neural oscillators: From order to chaos and back. *Chaos Solitons Fractals* **2020**, *136*, 109830. [CrossRef]
36. Matveyev, Y.; Egorov, K.V.; Markeev, A.; Zenkevich, A. Resistive switching and synaptic properties of fully atomic layer deposition grown TiN/HfO2/TiN devices. *J. Appl. Phys.* **2015**, *117*, 044901. [CrossRef]
37. Shi, Y.; Fong, S.; Wong, H.-S.P.; Kuzum, D. Synaptic devices based on phase-change memory. In *Neuro-Inspired Computing Using Resistive Synaptic Devices*; Springer: Cham, Switzerland, 2017; pp. 19–51.
38. Choi, S.; Tan, S.H.; Li, Z.; Kim, Y.; Choi, C.; Chen, P.-Y.; Yeon, H.; Yu, S.; Kim, J. SiGe epitaxial memory for neuromorphic computing with reproducible high performance based on engineered dislocations. *Nat. Mater.* **2018**, *17*, 335–340. [CrossRef]
39. Gerasimova, S.A.; Mikhaylov, A.; Belov, A.I.; Korolev, D.; Gorshkov, O.N.; Kazantsev, V.B. Simulation of synaptic coupling of neuron-like generators via a memristive device. *Tech. Phys.* **2017**, *62*, 1259–1265. [CrossRef]
40. Ignatov, M.; Ziegler, M.; Hansen, M.; Petraru, A.; Kohlstedt, H. A memristive spiking neuron with firing rate coding. *Front. Neurosci.* **2015**, *9*, 376. [CrossRef] [PubMed]
41. Korotkov, A.G.; Kazakov, A.; Levanova, T.; Osipov, G.V. The dynamics of ensemble of neuron-like elements with excitatory couplings. *Commun. Nonlinear Sci. Numer. Simul.* **2018**, *71*, 38–49. [CrossRef]
42. Bao, H.; Zhang, Y.; Liu, W.; Bao, B. Memristor synapse-coupled memristive neuron network: Synchronization transition and occurrence of chimera. *Nonlinear Dyn.* **2020**, *100*, 937–950. [CrossRef]
43. Parastesh, F.; Jafari, S.; Azarnoush, H.; Hatef, B.; Namazi, H.; Dudkowski, D. Chimera in a network of memristor-based Hopfield neural network. *Eur. Phys. J. Spéc. Top.* **2019**, *228*, 2023–2033. [CrossRef]
44. Pershin, Y.V.; Di Ventra, M. On the validity of memristor modeling in the neural network literature. *Neural Netw.* **2020**, *121*, 52–56. [CrossRef]
45. Williamson, A.; Schumann, L.; Hiller, L.; Klefenz, F.; Hoerselmann, I.; Husar, P.; Schober, A. Synaptic behavior and STDP of asymmetric nanoscale memristors in biohybrid systems. *Nanoscale* **2013**, *5*, 7297–7303. [CrossRef]

46. Yang, R.; Huang, H.-M.; Hong, Q.-H.; Yin, X.-B.; Tan, Z.-H.; Shi, T.; Zhou, Y.-X.; Miao, X.-S.; Wang, X.-P.; Mi, S.-B.; et al. Synaptic suppression triplet-STDP learning rule realized in second-order memristors. *Adv. Funct. Mater.* **2017**, *28*. [CrossRef]
47. Serrano-Gotarredona, T.; Masquelier, T.; Prodromakis, T.; Indiveri, G.; Linares-Barranco, B. STDP and STDP variations with memristors for spiking neuromorphic learning systems. *Front. Neurosci.* **2013**, *7*, 2. [CrossRef] [PubMed]
48. Nikiruy, K.E.; Surazhevsky, I.A.; Demin, V.A.; Emelyanov, A.V. Spike-timing-dependent and spike-shape-independent plasticities with dopamine-like modulation in nanocomposite memristive synapses. *Phys. Status Solidi A* **2020**, *217*, 1900938. [CrossRef]
49. Prudnikov, N.V.; Lapkin, D.A.; Emelyanov, A.V.; Minnekhanov, A.A.; Malakhova, Y.N.; Chvalun, S.N.; Demin, V.A.; Erokhin, V.V. Associative STDP-like learning of neuromorphic circuits based on polyaniline memristive microdevices. *J. Phys. D Appl. Phys.* **2020**, *53*, 414001. [CrossRef]
50. Demin, V.; Nekhaev, D.; Surazhevsky, I.; Nikiruy, K.; Emelyanov, A.; Nikolaev, S.; Rylkov, V.; Kovalchuk, M. Necessary conditions for STDP-based pattern recognition learning in a memristive spiking neural network. *Neural Netw.* **2020**, *134*, 64–75. [CrossRef]
51. Sarmiento-Reyes, A.; Rodriguez-Velasquez, Y. Maze-solving with a memristive grid of charge-controlled memristors. In Proceedings of the 2018 IEEE 9th Latin American Symposium on Circuits & Systems (LASCAS), Puerto Vallarta, Mexico, 25–28 February 2018; pp. 1–4. [CrossRef]
52. Isah, A.; Nguetcho, A.T.; Binczak, S.; Bilbault, J. Dynamics of a charge-controlled memristor in master–slave coupling. *Electronics* **2020**, *56*, 211–213. [CrossRef]
53. Raj, P.M.P.; Kalita, A.R.; Kundu, S. Memristive computational amplifiers and equation solvers. In *Modelling, Simulation and Intelligent Computing*; Goel, N., Ed.; Springer Nature: Singapore, 2020; pp. 74–82. [CrossRef]
54. Guo, T.; Wang, L.; Zhou, M.; Duan, S. A multi-layer memristive recurrent neural network for solving static and dynamic image associative memory. *Neurocomputing* **2018**, *334*, 35–43. [CrossRef]
55. Tanaka, G.; Nakane, R.; Yamane, T.; Takeda, S.; Nakano, D.; Nakagawa, S.; Hirose, A. Waveform classification by memristive reservoir computing. In *International Conference on Neural Information Processing*; Springer: Cham, Switzerland, 2017; pp. 457–465.
56. Erokhin, V. Memristive Devices for neuromorphic applications: Comparative analysis. *BioNanoScience* **2020**, *10*, 834–847. [CrossRef]
57. Isah, A.; Nguetcho, A.S.T.; Binczak, S.; Bilbault, J. Memristor dynamics involved in cells communication for a 2D non-linear network. *IET Signal Process.* **2020**, *14*, 427–434. [CrossRef]
58. Chakma, G.; Adnan, M.; Wyer, A.R.; Weiss, R.; Schuman, C.D.; Rose, G.S. Memristive mixed-signal neuromorphic systems: Energy-efficient learning at the circuit-level. *IEEE J. Emerg. Sel. Top. Circuits Syst.* **2017**, *8*, 125–136. [CrossRef]
59. Battistoni, S.; Cocuzza, M.; Marasso, S.L.; Verna, A.; Erokhin, V. The role of the internal capacitance in organic memristive device for neuromorphic and sensing applications. *Adv. Electron. Mater.* **2021**, 2100494. [CrossRef]
60. Bian, H.; Goh, Y.Y.; Liu, Y.; Ling, H.; Xie, L.; Liu, X. Stimuli-responsive memristive materials for artificial synapses and neuromorphic nomputing. *Adv. Mater.* **2021**, 2006469. [CrossRef] [PubMed]
61. Alsuwian, T.; Kousar, F.; Rasheed, U.; Imran, M.; Hussain, F.; Khalil, R.A.; Algadi, H.; Batool, N.; Khera, E.A.; Kiran, S.; et al. First principles investigation of physically conductive bridge filament formation of aluminum doped perovskite materials for neuromorphic memristive applications. *Chaos Solitons Fractals* **2021**, *150*, 111111. [CrossRef]
62. Gerasimova, S.A.; Mikhaylov, A.N.; Belov, A.I.; Korolev, D.; Guseinov, D.V.; Lebedeva, A.V.; Gorshkov, O.N.; Kazantsev, V.B. Design of memristive interface between electronic neurons. In *AIP Conference Proceedings*; AIP Publishing LLC location: Melville, NY, USA, 2018; Volume 1959, p. 090005. [CrossRef]
63. Kasdin, N.J. Runge-Kutta Algorithm for the numerical integration of stochastic differential equations. *J. Guid. Control. Dyn.* **1995**, *18*, 114–120. [CrossRef]
64. Kasdin, N.J. Discrete simulation of colored noise and stochastic processes and 1/fα power law noise generation. *Proc. IEEE* **1995**, *83*, 802–827. [CrossRef]
65. Higham, D.J. An Algorithmic Introduction to numerical simulation of stochastic differential equations. *SIAM Rev.* **2001**, *43*, 525–546. [CrossRef]
66. Emelyanov, A.; Nikiruy, K.; Demin, V.; Rylkov, V.; Belov, A.; Korolev, D.; Gryaznov, E.; Pavlov, D.; Gorshkov, O.; Mikhaylov, A.; et al. Yttria-stabilized zirconia cross-point memristive devices for neuromorphic applications. *Microelectron. Eng.* **2019**, *215*, 110988. [CrossRef]
67. Pan, F.; Gao, S.; Chen, C.; Song, C.; Zeng, F. Recent progress in resistive random access memories: Materials, switching mechanisms, and performance. *Mater. Sci. Eng. R Rep.* **2014**, *83*, 1–59. [CrossRef]
68. Lee, J.S.; Lee, S.; Noh, T.W. Resistive switching phenomena: A review of statistical physics approaches. *Appl. Phys. Rev.* **2015**, *2*, 031303. [CrossRef]
69. Ignatov, M.; Ziegler, M.; Hansen, M.; Kohlstedt, H. Memristive stochastic plasticity enables mimicking of neural synchrony: Memristive circuit emulates an optical illusion. *Sci. Adv.* **2017**, *3*, e1700849. [CrossRef]
70. Miranda, E.; Mehonic, A.; Ng, W.H.; Kenyon, A. Simulation of cycle-to-cycle Instabilities in SiOx- based ReRAM devices using a self-correlated process with long-term variation. *IEEE Electron Device Lett.* **2018**, *40*, 28–31. [CrossRef]
71. Agudov, N.; Dubkov, A.; Safonov, A.; Krichigin, A.; Kharcheva, A.; Guseinov, D.; Koryazhkina, M.; Novikov, A.; Shishmakova, V.; Antonov, I.; et al. Stochastic model of memristor based on the length of conductive region. *Chaos Solitons Fractals* **2021**, *150*, 111131. [CrossRef]

72. Guseinov, D.; Matyushkin, I.; Chernyaev, N.; Mikhaylov, A.; Pershin, Y. Capacitive effects can make memristors chaotic. *Chaos Solitons Fractals* **2021**, *144*, 110699. [CrossRef]
73. Parvizi-Fard, A.; Amiri, M.; Kumar, D.; Iskarous, M.M.; Thakor, N.V. A functional spiking neuronal network for tactile sensing pathway to process edge orientation. *Sci. Rep.* **2021**, *11*, 1320. [CrossRef] [PubMed]

sensors

MDPI

Article

A Spectrum Correction Algorithm Based on Beat Signal of FMCW Laser Ranging System

Yi Hao [1], Ping Song [1,*], Xuanquan Wang [1] and Zhikang Pan [2]

[1] Key Laboratory of Biomimetic Robots and Systems (Ministry of Education), Beijing Institute of Technology, Beijing 100081, China; 3120190156@bit.edu.cn (Y.H.); 3120185108@bit.edu.cn (X.W.)
[2] Instrument of Science and Technology, Beijing Information Science and Technology University, Beijing 100192, China; panzhikang@bistu.edu.cn
* Correspondence: sping2002@bit.edu.cn; Tel.: +86-136-6136-5650

Abstract: The accuracy of target distance obtained by a frequency modulated continuous wave (FMCW) laser ranging system is often affected by factors such as white Gaussian noise (WGN), spectrum leakage, and the picket fence effect. There are some traditional spectrum correction algorithms to solve the problem above, but the results are unsatisfactory. In this article, a decomposition filtering-based dual-window correction (DFBDWC) algorithm is proposed to alleviate the problem caused by these factors. This algorithm reduces the influence of these factors by utilizing a decomposition filtering, dual-window in time domain and two phase values of spectral peak in the frequency domain, respectively. With the comparison of DFBDWC and these traditional algorithms in simulation and experiment on a built platform, the results show a superior performance of DFBDWC based on this platform. The maximum absolute error of target distance calculated by this algorithm is reduced from 0.7937 m of discrete Fourier transform (DFT) algorithm to 0.0407 m, which is the best among all mentioned spectrum correction algorithms. A high performance FMCW laser ranging system can be realized with the proposed algorithm, which has attractive potential in a wide scope of applications.

Keywords: FMCW laser ranging; spectrum correction; white Gaussian noise; spectrum leakage; picket fence effect; signal processing

Citation: Hao, Y.; Song, P.; Wang, X.; Pan, Z. A Spectrum Correction Algorithm Based on Beat Signal of FMCW Laser Ranging System. *Sensors* **2021**, *21*, 5057. https://doi.org/10.3390/s21155057

Academic Editor: Manuel José Cabral dos Santos Reis

Received: 23 June 2021
Accepted: 23 July 2021
Published: 26 July 2021

1. Introduction

The frequency modulated continuous wave (FMCW) laser ranging system is a non-contact detecting and distance measurement system, which has a large detection range and high measurement accuracy and has been widely used in high precision ranging. The system utilizes the corresponding relationship between frequency and distance, which means that the accuracy of distance value relies on the frequency resolution of beat signal obtained by a series of processing with the emitted signal and echo signal. Therefore, the key of ranging lies in the frequency calculation of beat signal [1].

The frequency value of the beat signal can be computed by discrete Fourier transform (DFT) after the signal is sampled and digitized. Ideally, the frequency resolution of it is closely related to the number of sampling points. Too few sampling points will decrease the frequency resolution and lead to the picket fence effect, which affects ranging accuracy, while too many will increase the computing time and the complexity of signal processing. An optimization method is to add points whose value are zero after the sampled beat signal [2]. The accuracy of this method for calculating the frequency completely depends on the number of added points. However, this operation is equivalent to utilizing a rectangular window function on the beat signal in the time domain. This not only cannot change the width of the main lobe in the spectrum but also causes spectrum leakage to a certain extent. Consequently, spectrum correction algorithms to improve the accuracy and resolution of the beat signal frequency have become more significant.

The ratio algorithm is the interpolation-based correction method [3–5]. Agrez et al. [6] and Belega et al. [7–9] have conducted a further studies on it and proposed new methods based on it to reduce the influence of spectrum leakage on the accuracy of correction. However, the above methods are all at the expense of noise adaptability. The phase difference (PD) algorithm originally comes from a phase interpolation estimator of a single tone frequency in noise proposed by McMahon et al. [10]. Zhu et al. [11] and Kang et al. [12] have performed further research on it, which indicates that the PD algorithm provides superior accuracy in frequency estimates compared with the ratio algorithm and has good adaptability. Luo et al. [13] proposed a new PD method based on asymmetric windows, which can be used to correct the errors of frequency. The main advantages of this algorithm are its characteristics of simple application and strong anti-noise performance, but its reduction of spectrum leakage is unsatisfactory.

The concept of energy centrobaric correction (ECC) algorithm [14] is originally proposed by Offelli et al. [15]. Many researchers have investigated the interferences from spectral components, wideband noise, and other precision factors related to the estimated parameters [16,17]. This algorithm has fast speed and great accuracy of frequency calculation, so it has been applied to engineering after improvement [18]. However, the correction accuracy is too dependent on the symmetric window function and is easily affected by white Gaussian noise (WGN). The Chirp z-transform (CZT) algorithm is a z-transformation method [19]. Because of the low complexity of calculation and the high correction accuracy, there are a lot of CZT-based related applications [20–25]. Because this algorithm still analyzes the truncated signal, it only reduces the influence of the picket fence effect on the local spectrum, whereas it does not significantly solve the problem caused by spectrum leakage. The Zoom fast Fourier transform (ZFFT) algorithm achieves spectrum correction by reducing the sampling rate of the signal. It blends complex down-conversion, low-pass filtering, and sample-rate change by way of decimation, thereby improving the frequency resolution [26]. Al-Qudsi et al. [27] presented an implementation method of the ZFFT approach to estimate the spectral peak in the FMCW radar, utilizing a field programmable gate array (FPGA). This algorithm can decrease the complexity of calculations and alleviate the influence of picket fence effect. However, it is severely affected by spectrum leakage and WGN.

In view of the unsatisfactory accuracy and resolution of beat signal frequency affected by WGN, spectrum leakage, and picket fence effect, which cannot be solved by the traditional algorithms above, we propose a new spectrum correction algorithm called decomposition filtering-based dual-window correction (DFBDWC). The main contributions are as follows:

(1) This algorithm reduces the influence of WGN, affecting the correction accuracy. In the decomposition and filter part, the beat signal is divided into several components, and each component has its characteristics in the frequency domain. Among them, the first few components possess the widest frequency coverage, and there are no obvious peaks in their power spectrum. The sum can be used as the input of the filter, and the WGN in the beat signal will be mostly removed with the weight parameter.

(2) This algorithm minimizes the impact of spectrum leakage effectively. The Hann window has a narrow main lobe, low side lobe, and fast attenuation speed from the main lobe to the first side lobe. Using two Hann windows in the correction part can concentrate more energy of the signal, thereby making the spectral peak of the desired frequency more obvious.

(3) This algorithm diminishes the picket fence effect that may decrease the frequency resolution of the beat signal. We utilize phase values and the delay value of two signals in the frequency domain after DFT processing. The phase values correspond to the spectral peaks that are at the same position in these signals. Therefore, the calculation error caused by broad adjacent spectral lines near the peak in only one used signal is avoided, and an accurate frequency value of the beat signal is obtained.

(4) This algorithm is different from the traditional spectrum correction algorithm, which can reduce the influence caused by WGN, spectrum leakage, and the picket fence effect at the same time, so that the frequency value obtained by this algorithm is more accurate and the distance ranged by this system is more precise.

This article is organized as follows. In Section 2, the principle of the FMCW laser ranging system is firstly briefly introduced, and we explain the DFBDWC algorithm in detail. In Section 3, we built an experimental platform based on the principle of the FMCW laser-ranging system. The results are obtained via simulation and experiment on this platform. Afterwards, the discussion that evaluates the spectrum correction performance of this algorithm by comparing it with these traditional algorithms is conducted. Finally, Section 4 concludes the article.

2. Methods

Using the method shown in Figure 1, we can obtain the high-precision distance value of the target.

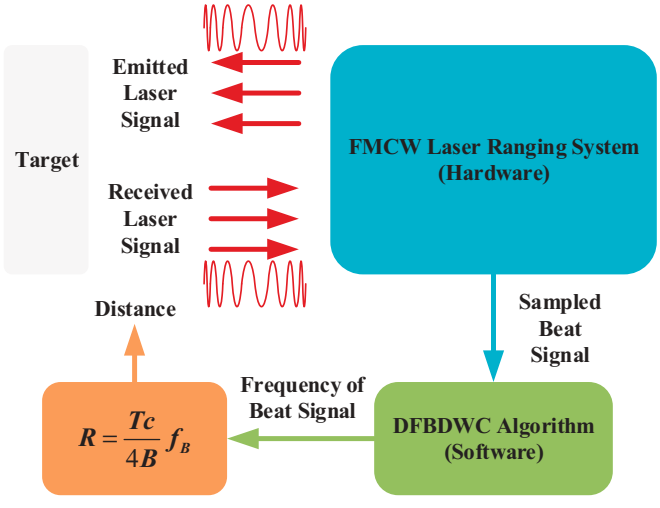


Figure 1. Schematic diagram of distance obtained by the FMCW laser ranging system and DFBDWC algorithm.

The FMCW laser-ranging system emits a modulated laser signal that is reflected by the target and received by the system. After the processing of the received laser signal, the system will output the sampled beat signal. In the software part, we can calculate the precise frequency value of the sampled beat signal with the DFBDWC algorithm and obtain the distance value of this target by taking the frequency value into the equation. In this section, we will introduce the principle of the FMCW laser ranging system and the DFBDWC algorithm, respectively, in detail.

2.1. FMCW Laser Ranging System

The FMCW laser ranging system can be mainly divided into seven parts. The schematic diagram of it is as shown in Figure 2. The signal processing part controls the signal emitting part to generate the FMCW emitted signal, and it drives the laser diode to emit a linear beam, which is the emitted laser signal. The avalanche photo diode (APD) receives the laser signal that is focused by the lens and outputs the echo signal, which is a FMCW signal with a certain delay of emitted signal. The echo signal and the local oscillator signal synchronized by the signal emitting part are mixed in the signal mixing part, and with a series of processing, the beat signal is obtained. In the signal processing

part, the beat signal is digitized and transformed into data, which are stored and sent to the PC. Finally, the beat signal is analyzed and processed by the algorithm in the PC, and the distance is computed.

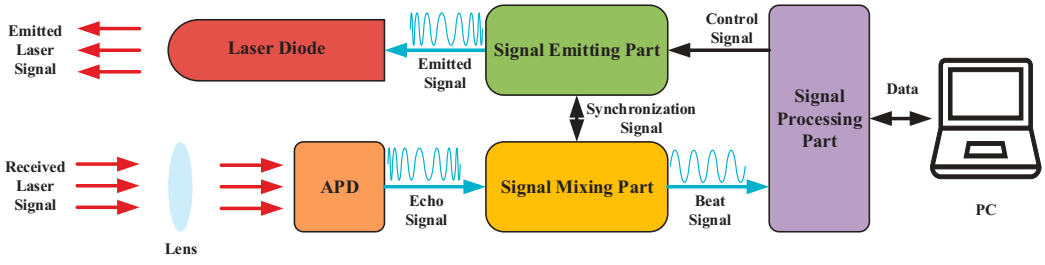


Figure 2. Scheme of FMCW laser ranging system.

In this FMCW laser ranging system, the frequency of the emitted signal is modulated by a triangle wave, which is as shown in Figure 3. Because of the static ranging target, the effect of the Doppler shift does not have to be considered. Then, the emitted signal $s_T(t)$ is expressed by

$$s_T(t) = A_0 \cos\left(2\pi f_0 t + \pi k t^2 + \varphi_0\right), \tag{1}$$

where A_0 is the amplitude of emitted signal, f_0 is the initial frequency, φ_0 is the initial phase, $k = 2B/T$ is the modulation slope, B is the modulation bandwidth, and T is the modulation period. With the delay $\tau = 2R/c$, the echo signal $s_R(t)$ can be obtained by

$$s_R(t) = \eta A_0 \cos\left(2\pi f_0 (t-\tau) + \pi k (t-\tau)^2 + \varphi_0\right), \tag{2}$$

where η is the amplitude decay rate of echo signal, R is the distance, and c is the speed of light. With the mixing of the emitted and echo signal, the beat signal $s_B(t)$ can be calculated by

$$s_B(t) = \frac{\eta {A_0}^2}{2} \cos\left(2\pi f_0 \tau + 2\pi k \tau t - \pi k \tau^2\right). \tag{3}$$

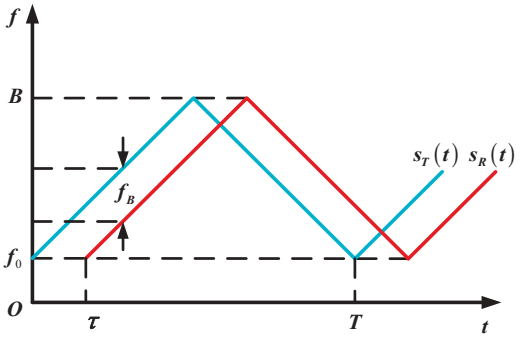


Figure 3. Type of FMCW modulation. The blue and red line represent the emitted signal $s_T(t)$ and echo signal $s_R(t)$, respectively.

Obviously, the frequency of the beat signal $f_B = k\tau$. Thus, the relationship between R and f_B is

$$R = \frac{Tc}{4B} f_B. \tag{4}$$

As shown in the above equation, the factors affecting the accuracy of FMCW laser ranging system are T, B, and f_B. Because T and B are the inherent parameters of this system, and they have already reached the limit of system performance, they do not have a decisive impact on the ranging accuracy. Improving the frequency resolution of the beat signal and obtaining the accurate f_B become the most significant work of this system.

2.2. DFBDWC Algorithm

This section depicts a new spectrum correction algorithm that is different from the six traditional algorithms introduced in Section 1. The key step of it, whose process chart is as shown in Figure 4, is as follows: Improving the signal to noise ratio (SNR) of $s_B(t)$ in decomposition and filter, extracting two sub-signals with a dual-window on each of them, and calculating accurate f_B according to the phase values in correction and DFT spectrum analysis. All the parameters are in digital form as (n).

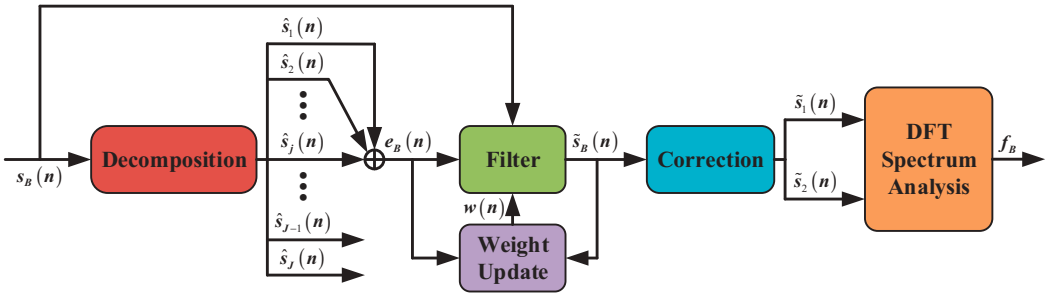


Figure 4. Process of DFBDWC algorithm. The input and output of this algorithm are the beat signal and frequency value of the beat signal, respectively.

WGN is a kind of noise whose probability density function satisfies the statistical characteristics of normal distribution and whose power spectral density function is constant. The most noteworthy feature of it is that the signal contains all frequency components from negative infinity to positive infinity, so it can be apparently distinguished from useful signals with a spectral peak in the spectrum. Accordingly, a similar sequence of WGN can be decomposed from the beat signal.

First, decomposition is a new method based on empirical mode decomposition (EMD) [28–30] that is a direct extraction of the energy associated with various intrinsic time scales and the most important parameters of the system. After processing with EMD, a signal can be expressed as a sum of amplitude- and frequency-modulated functions called modes. Each mode is intrinsic and has unique characteristics in the frequency domain, which means several of them enable the estimation of WGN. However, there are some phenomena in EMD, such as oscillations with very disparate scales in one mode or oscillations with similar scales in different modes. These phenomena will cause a problem called "mode mixing", and some decomposed modes, strictly speaking, will not be a single component signal, so it is not conducive to estimate the noise component accurately.

In order to alleviate "mode mixing", we take advantage of the dyadic filter bank behavior of EMD and the addition of WGN that populates the whole time–frequency space. Then, the K sub-signal of $s_B(n)$ can be expressed by

$$(s_B(n))_k = s_B(n) + (-1)^k \cdot \beta_1 \cdot E_1(G_k(n)), \tag{5}$$

where $G_k(n)$ $(k = 1, 2, \ldots, K-1, K)$ is the kth added WGN signal of zero mean unit variance, K is the number of WGN signals, $E_j(\cdot)$ $(j = 1, 2, \ldots, J-1, J)$ is the jth mode obtained by EMD, and J is the number of modes or components, β_j $(j = 1, 2, \ldots, J-1, J)$ is the jth parameter used to adjust SNR between added WGN signals and $s_B(n)$. The main purpose

of adding the WGN signal with known features operated by EMD is to generate new extreme points. Additionally, with the operation of plus and minus, $s_B(n)$ will be forced to focus on some specific values in the scale energy space.

After that, let $M(\cdot)$ be the operator that produces the mean envelope of each signal in parentheses, which is same as the procedure in EMD and will make use of these new extreme points in parentheses. Additionally, let $A(\cdot)$ be the operator that produces an average signal of all signals in parentheses. With the operation of EMD, we can obtain the first component $\hat{s}_1(n)$ of beat signal $s_B(n)$, which is

$$\begin{aligned}\hat{s}_1(n) &= A(E_1((s_B(n))_k)) = A((s_B(n))_k - M((s_B(n))_k)) \\ &= s_B(n) - A(M((s_B(n))_k)).\end{aligned} \tag{6}$$

Averaging is meant to better estimate the mean envelope value, which reduces "mode mixing" and produces more distinct components. Because of the concept in EMD called residue, the first residue of the beat signal can be expressed by

$$R_1(n) = s_B(n) - \hat{s}_1(n) = A(M((s_B(n))_k)). \tag{7}$$

With $\hat{s}_1(n)$ and $R_1(n)$, we can estimate the second residue $R_2(n)$ and the second component $\hat{s}_2(n)$ of beat signal, respectively, by

$$R_2(n) = A\Big(M\Big(R_1(n) + (-1)^k \cdot \beta_2 \cdot E_2(G_k(n))\Big)\Big), \tag{8}$$

$$\hat{s}_2(n) = R_1(n) - R_2(n) = R_1(n) - A\Big(M\Big(R_1(n) + (-1)^k \cdot \beta_2 \cdot E_2(G_k(n))\Big)\Big). \tag{9}$$

Similarly, for the jth($j = 3, 4, \ldots, J-1, J$) residue $R_j(n)$ and the jth($j = 3, 4, \ldots, J-1, J$) component $\hat{s}_j(n)$ of the beat signal can be calculated by

$$R_j(n) = A\Big(M\Big(R_{j-1}(n) + (-1)^k \cdot \beta_j \cdot E_j(G_k(n))\Big)\Big), \tag{10}$$

$$\begin{aligned}&\hat{s}_j(n) = R_{j-1}(n) - R_j(n) \\ &= R_{j-1}(n) - A\Big(M\Big(R_{j-1}(n) + (-1)^k \cdot \beta_j \cdot E_j(G_k(n))\Big)\Big).\end{aligned} \tag{11}$$

In this way, we not only utilize the advantages of EMD to make the frequency distribution of components more obvious but also reduce the effect of "mode mixing" so as to estimate the noise components more accurately.

Next, these J components can be distinguished according to the characteristics of each component in the frequency domain. Among them, the first to jth components possess the widest frequency coverage, and there are no obvious peaks in their power spectrum. At the same time, they are scattered in the time domain. Therefore, the sum of the first to jth components is regarded as the evaluation signal $e_B(n)$ of $s_B(n)$, which can be considered to be the estimation of WGN, and we can put it into the filter.

The filter in this algorithm has three inputs and one output [31], which is expressed by

$$\widetilde{s}_B(n) = s_B(n) - w(n)e_B(n), \tag{12}$$

where $w(n)$ is a coefficient of weight, $\widetilde{s}_B(n)$ is reconstruction signal of $s_B(n)$. In this filter, $e_B(n)$ is weighted by a parameter at the same instant, so we consider that it is the possible interference signal. If it is subtracted from $s_B(n)$, the useful information can be saved as much as possible in $\widetilde{s}_B(n)$. The weight $w(n)$ is not a fixed parameter, which needs to be updated with the input $e_B(n)$ and the output $\widetilde{s}_B(n)$ at the same instant; then, its value of the next instant will be obtained. In order to ensure the best result of this filter, we consult the calculation method of $w(n)$ in [31].

However, these parameters above are described in matrices or vectors according to [31], which will lead to the great cost of increased computational complexity and some

stability problems. In order to reduce the complexity of computation and the cost of calculation, we have changed the order of $w(n)$ into 1 without affecting the performance of the filter. The expression of the weight update can be expressed by

$$w(n) = w(n-1) + r(n)\widetilde{s}_B(n)e_B(n), \tag{13}$$

where $r(n)$ is a relevant coefficient. According to [31], it can be obtained by

$$r(n) = \frac{r(n-1)}{\lambda}[1 - g(n)e_B(n)], \tag{14}$$

where λ is the forgetting factor. It is introduced to give a greater forgetting effect to $\widetilde{s}_B(n)$ of the latest moment, and give less forgetting effect to $\widetilde{s}_B(n)$ of the farther moment, so as to ensure that $\widetilde{s}_B(n)$ in the past period is "forgotten" well, so that the filter can work in a more stable state. $g(n)$ is a coefficient of gain, which controls the effect of the output $\widetilde{s}_B(n)$. Referring to the steps in [31], $g(n)$ is calculated by

$$g(n) = \frac{r(n-1)e_B(n)}{\lambda + r(n-1){e_B}^2(n)}. \tag{15}$$

Afterwards, we utilize correction to process $\widetilde{s}_B(n)$ without the interference of WGN. Correction is the processing of $\widetilde{s}_B(n)$ in the time domain [32,33], which is used to reduce the picket fence effect and spectrum leakage by spectral peaks in only two sub-signals of $\widetilde{s}_B(n)$ with the time delay and dual-window of each sub-signal, respectively. With the processing of this step in the DFBDWC algorithm, the calculated frequency value will be more accurate and precise. Firstly, we extract two series of sub-signals, ${\widetilde{s}_B}^{(1)}(n)$ and ${\widetilde{s}_B}^{(2)}(n)$, from $\widetilde{s}_B(n)$. There are L points of delay between them, which means that ${\widetilde{s}_B}^{(2)}(n)$ is L points behind ${\widetilde{s}_B}^{(1)}(n)$. By putting the first L points of ${\widetilde{s}_B}^{(1)}(n)$ and the whole points of ${\widetilde{s}_B}^{(2)}(n)$ together, we can acquire $\widetilde{s}_B(n)$.

After that, the first correction signal $\widetilde{s}_1(n)$ can be obtained by

$$\widetilde{s}_1(n) = S\Big({\widetilde{s}_B}^{(1)}(n) \cdot N(W(n) * W(n))\Big), \tag{16}$$

where $W(n)$ is a window signal, which is usually the Hann window because of its excellent performance in side lobe suppression, $*$ is the operator of convolution, and $N(\cdot)$ is the operator of producing normalization signal. $S(\cdot)$ is the operator that produces a sum signal of the signal's front and back halves in parentheses. Figuratively speaking, there is a signal whose length is $2N$ in parentheses of $S(\cdot)$; this signal's front half means the first to Nth points and back half means the $N+1$th to $2N$th points. The sum signal produced by $S(\cdot)$ is in the length of N, which is formed by adding the values of the first and $N+1$th, the second and $N+2$th, ... , the Nth points and $2N$th points.

Similarly, the second correction signal $\widetilde{s}_2(n)$ can be obtained by

$$\widetilde{s}_2(n) = S\Big({\widetilde{s}_B}^{(2)}(n) \cdot N(W(n) * W(n))\Big). \tag{17}$$

With a dual-window in the time domain, the influence of spectrum leakage can be decreased more than the one-window and none-window, that is, the energy is more concentrated in the main lobe of these signals, which is more conducive to the subsequent operation in the frequency domain. Moreover, there cannot be more than two windows applied to these sub-signals of $\widetilde{s}_B(n)$, because the mathematical model of correction processing is only in two dimensions.

Finally, in order to make signals turn from time domain into the frequency domain, we process $\widetilde{s}_1(n)$ and $\widetilde{s}_2(n)$ with DFT to obtain their spectrum signals $\widetilde{S}_1(q)$ and $\widetilde{S}_2(q)$. Before DFT, $\widetilde{s}_1(n)$ and $\widetilde{s}_2(n)$ can be also expressed in exponential form as

$$\widetilde{s}_1(n) = Ae^{j({\omega_B}^* n + \theta)}, \tag{18}$$

$$\widetilde{s}_2(n) = Ae^{j[\omega_B{}^*(n-L)+\theta]}, \tag{19}$$

where A is the amplitude of $\widetilde{s}_B(n)$, and θ is the initial phase of $\widetilde{s}_B(n)$. $\omega_B{}^*$ is the angular frequency of $\widetilde{s}_B(n)$, and it can be calculated by

$$\omega_B{}^* = \frac{2\pi f_B}{f_S}, \tag{20}$$

where f_S is the sample rate.

After the processing of $\widetilde{s}_B{}^{(1)}(n)$ and $\widetilde{s}_B{}^{(2)}(n)$ with DFT, we can obtain their spectrum signals $\widetilde{S}_1(q)$ and $\widetilde{S}_2(q)$, respectively, by

$$\widetilde{S}_1(q) = Ae^{j\theta}F_g{}^2(q\Delta\omega - \omega_0), \tag{21}$$

$$\widetilde{S}_2(q) = Ae^{j(\theta-\omega_B{}^*L)}F_g{}^2(q\Delta\omega - \omega_0), \tag{22}$$

where F_g is the amplitude spectrum of $W(n)$, q is the serial number of spectral lines, $\Delta\omega$ is the angular frequency between each spectral lines, and ω_0 is the initial angular frequency.

In the amplitude spectrum of $\widetilde{S}_1(q)$, the corresponding serial number of its spectral peak is q^*. According to this, we can find the phase values $\varphi_1(q^*)$ and $\varphi_2(q^*)$ in the phase spectrum of $\widetilde{S}_1(q)$ and $\widetilde{S}_2(q)$, respectively. These phase values are expressed by

$$\varphi_1(q^*) = \theta, \tag{23}$$

$$\varphi_2(q^*) = \theta - \omega^* L. \tag{24}$$

With Equations (23) and (24), we can only obtain an estimation of $\omega_B{}^*$ as

$$\varphi_1(q^*) - \varphi_2(q^*) = \hat{\omega}_B{}^* L. \tag{25}$$

This is because $\varphi_1(q^*) - \varphi_2(q^*)$ is still different from the ideal value; a compensated value of the phase difference has to be introduced. The corresponding angular frequency at the spectral peak q^* is $2\pi q^*/I$, where I is the length of $W(n)$. After the delay of L, this angular frequency will lead to an additional phase shift of $2\pi q^* L/I$, which will increase with this delay. Since the position of spectral peak can be observed, $2\pi q^* L/I$ is considered to be the compensated value of the phase difference. Then, we will calculate $\omega_B{}^*$ by

$$\varphi_1(q^*) - \varphi_2(q^*) + \frac{2\pi q^* L}{I} = \omega_B{}^* L. \tag{26}$$

Eventually, according to Equations (20) and (26), the frequency value f_B of $s_B(n)$ can be calculated by

$$f_B = \frac{f_S}{2\pi}\left[\frac{\varphi_1(q^*) - \varphi_2(q^*)}{L} + \frac{2\pi q^*}{I}\right]. \tag{27}$$

The relationship between $\widetilde{s}_B{}^{(1)}(n)$ and $\widetilde{s}_B{}^{(2)}(n)$ with L delay will overcome the error caused by two wide spectral lines. With $2\pi q^*/I$, we can compute f_B more precisely, and the influence of the picket fence effect will be reduced well. Therefore, the frequency resolution of $s_B(n)$ can be improved, and the purpose of spectrum correction may be achieved.

3. Results and Discussion

In this section, the performance of DFBDWC algorithm is evaluated by both a simulation and an experiment. In the simulation part, an original signal is constructed with Equation (3). Furthermore, to reach the real situation as much as possible, a WGN signal with appropriate SNR value is added to it, which is regarded as a beat signal. In the experiment part, we built an experimental platform according to the scheme shown in Figure 2, and a beat signal obtained with it is analyzed and processed by this algorithm. Table 1 shows the parameters used in the simulation and the experiment.

Table 1. The parameters used in the simulation and experiment.

Parameter	Interpretation	Value
η	The amplitude decay rate of echo signal	0.8
A_0	The amplitude of emitted signal	1
f_0	The initial frequency	1 MHz
c	The speed of light	299,792,458 m/s
B	The modulation bandwidth	99 MHz
T	The modulation period	200 μs
f_S	The sample rate	20 MHz
f_U	The upper limit frequency	0.8
f_L	The lower limit frequency	1

Among them, η and A_0 are only used in the simulation. f_0, B, and T are the key parameters of the FMCW laser ranging system, which influence the theoretical accuracy of distance resolution and are determined by the direct digital synthesizer (DDS) in the experimental platform. Moreover, f_S influences the number of sampled points in beat signal and is determined by analog-to-digital converter (ADC) in the experimental platform. f_U and f_L are considered to be a band-pass filter applied to the beat signal, and they are determined by the performance of the low-pass filter (LPF) in the experimental platform and T, respectively. Because τ cannot be greater than $T/2$, the value of f_L is $2/T$; otherwise, it will violate the principle of the FMCW laser ranging system. According to Equation (4), the distance values that this system can obtain are from 1.5141 to 22.7115 m, which correspond to f_L and f_U, respectively. Moreover, because the minimum sample time is $T/2$, the maximum range resolution of the system is 1.5141 m, which is given by DFT. In order to ensure the best working states of this platform and verify the performance of this algorithm better, the test distance values are shown in Table 2.

Table 2. The test distance value used in the simulation and experiment.

Lower Limit (m)	Test Distance (m)									Upper Limit (m)
1.5141	2 2.5	3 3.5	4 4.5	5 5.5	6 6.5	7 7.5	8 8.5	9 9.5	10	22.7115

3.1. Simulation

As shown in Figure 5, it is a comparison diagram of $\widetilde{s}_B(n)$ with $s_B(n)$ at two specific distances in both the time and frequency domain.

It can be seen that there is a great interference in $s_B(n)$. Compared with $s_B(n)$, the WGN interference in $\widetilde{s}_B(n)$ is reduced well, and the wave pattern of $\widetilde{s}_B(n)$ becomes smoother. Additionally, $\widetilde{s}_B(n)$ also retains the information of shape, amplitude, and frequency. In their power spectrum, the amplitude of the spectral peak in $\widetilde{s}_B(n)$ is greater than the amplitude of the spectral peak in $s_B(n)$. Moreover, the power spectrum of WGN has also been decreased. The above results indicate that the DFBDWC algorithm can suppress the interference of WGN with a large SNR value in the beat signal and retain the useful information in the signal.

In order to verify the comparison of these signals further, we apply SNR and noise power P_{noise} here. They can be expressed, respectively, by

$$SNR = 10\lg\left(\frac{\sum_{n=1}^{N} x^2(n)}{\sum_{n=1}^{N}[x(n) - s_O(n)]^2}\right), \tag{28}$$

$$P_{noise} = \frac{1}{N}\sum_{n=1}^{N}[x(n) - s_O(n)]^2, \tag{29}$$

where $x(n)$ represents $s_B(n)$ or $\widetilde{s}_B(n)$. SNR judges the macroscopic performance of this algorithm. The larger the SNR is, and the smaller P_{noise} is, the better the performance of this algorithm is, that is, the energy will be more concentrated in the position of the spectral peak and the probability of selecting a "fake" peak as the target position due to enormous WGN will be reduced better.

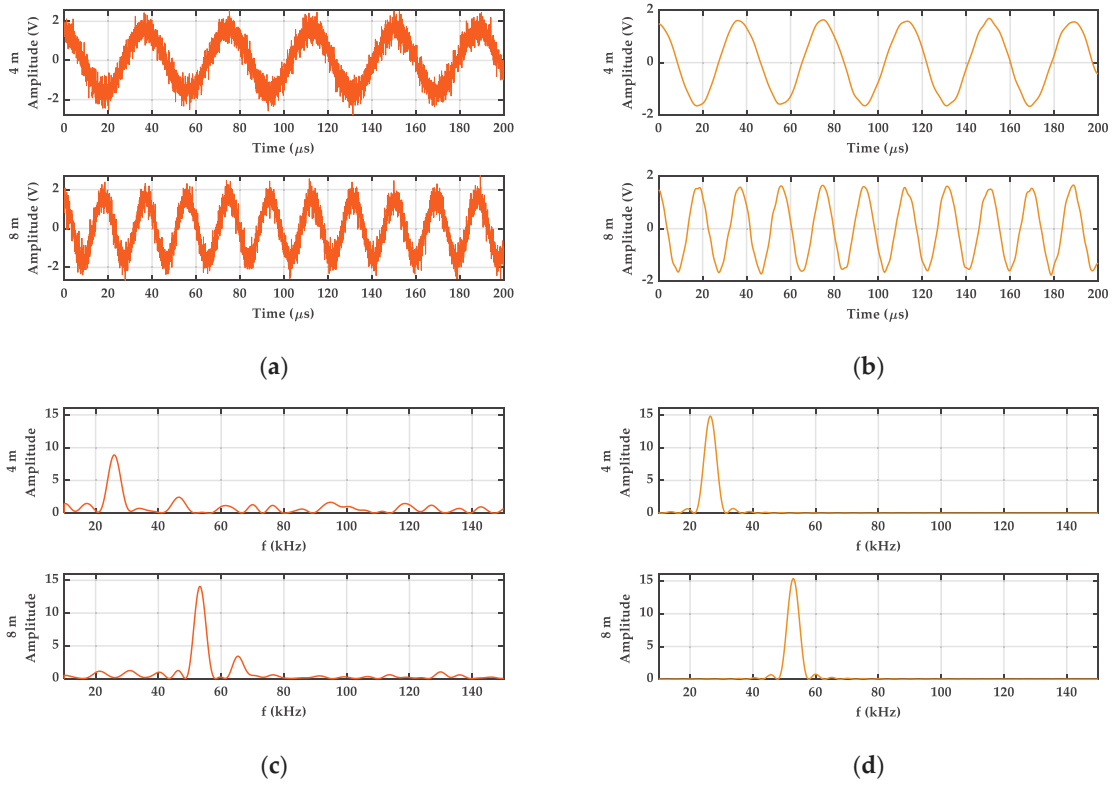


Figure 5. Simulation results of signals comparison at two different distances. (**a**) Beat signals $s_B(n)$ obtained by original signals and WGN signals with SNR of 10 dB. (**b**) Reconstruction signals $\widetilde{s}_B(n)$. (**c**) The power spectrum of $s_B(n)$. (**d**) The power spectrum of $\widetilde{s}_B(n)$.

We calculate the value of these parameters at different distances and noise powers, and the results are as shown in Figure 6. Among the first two figures, SNR increases from 10 dB to more than 25 dB, and P_{noise} reduces from 0.13 W to the order of 10^{-3} W. In the last two figures, SNR increases from 1 dB to more than 11 dB, and P_{noise} reduces from more than 5 W to about 0.1 W. The simulation results above indicate that this algorithm has a great suppression effect on WGN interference, and it saves useful information in $\widetilde{s}_B(n)$ as much as possible. This will improve the accuracy of f_B.

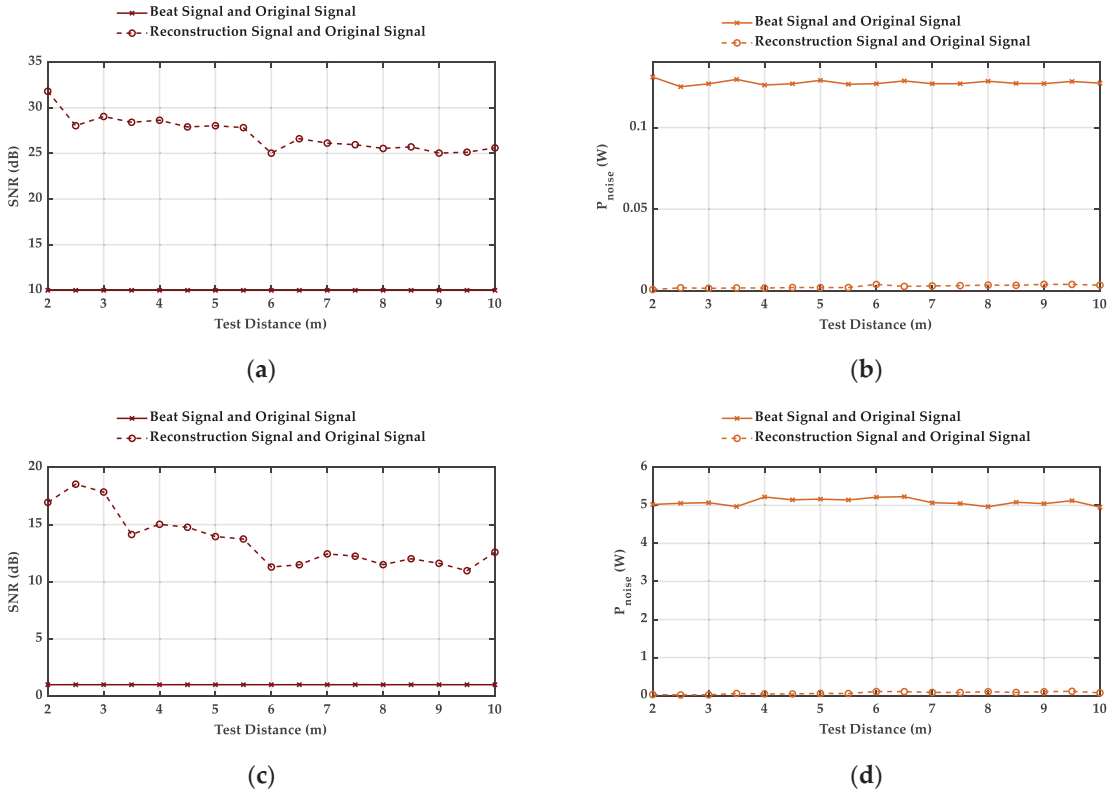


Figure 6. Simulation results of two parameters for evaluating signal comparison at different distances and noise powers. (**a**,**c**) are SNR, (**b**,**d**) are P_{noise}. Additionally, the original SNR of WGN in (**a**,**b**) is 10 dB, and the original SNR of WGN in (**c**,**d**) is 1 dB.

Furthermore, we compare the performance of the DFBDWC algorithm with six other algorithms in terms of computing f_B, that is, the computed distances, and the original SNR of WGN in $s_B(n)$ is 10 dB. To show the results of the comparison better, absolute error (AE) and root mean square error ($RMSE$) are applied here. They can be expressed, respectively, by

$$AE = |R_C(p) - R_T(p)|, \tag{30}$$

$$RMSE = \sqrt{\frac{1}{P}\sum_{p=1}^{P}[R_C(p) - R_T(p)]^2}, \tag{31}$$

where $R_C(p)$ is the computed distance of the pth test distance, $R_T(p)$ is the pth test distance, and P is the number of test distance. The smaller the $RMSE$ is, the better the performance of the algorithm is.

The computed distance and calculation results of AE and $RMSE$ can be seen from Table 3 and Figure 7, respectively. At the first test distance, AE of the PD algorithm has the maximum value, and at the last test distance, AE of DFT algorithm has the maximum value. Among these traditional spectrum algorithm, DFT algorithm has the biggest jump of AE, while the ECC algorithm has the smallest jump of AE. The AE of the Ratio, ECC, CZT, and ZFFT algorithm are relatively stable. At each test distance, the AE of the DFBDWC algorithm basically has the minimum value. $RMSE$ macroscopically evaluates the deviation between $R_C(p)$ and $R_T(p)$. It can be seen that the DFT algorithm has the

maximum value, and the DFBDWC algorithm has the minimum value, which indicates that the DFBDWC algorithm performs the best compared with other traditional algorithms.

Table 3. Simulation results of computed distances comparison between seven algorithms. There are WGN signals with SNR of 10 dB in $s_B(n)$.

Test Distance (Real Distance) (m)	Computed Distance (m)						
	DFT	PD	ECC	Ratio	CZT	ZFFT	DFBDWC
2	1.8483	2.2712	2.0092	2.0162	2.0109	2.1230	2.0021
2.5	2.3103	2.5013	2.4898	2.5102	2.5136	2.5121	2.5011
3	3.2345	2.9997	3.0018	2.9977	3.0164	2.9354	3.0004
3.5	3.6965	3.5102	3.5163	3.5118	3.5191	3.6794	3.5038
4	4.1586	3.9989	3.9992	4.0074	3.9923	4.0685	4.0003
4.5	4.6207	4.4945	4.4967	4.4939	4.5246	4.4919	4.4981
5	5.0827	4.9836	5.0128	5.0064	4.9977	4.8810	4.9988
5.5	5.5448	5.4979	5.4977	5.5069	5.5005	5.6250	5.5021
6	6.0069	5.9938	5.9968	5.9938	6.0032	6.0141	6.0028
6.5	6.4689	6.5056	6.5119	6.5053	6.5059	6.4374	6.4969
7	6.9310	6.9930	6.9942	7.0027	7.0086	6.8265	7.0017
7.5	7.3931	7.4954	7.4987	7.4968	7.5114	7.5705	7.5032
8	7.8552	7.9956	8.0202	8.0079	8.0141	7.9939	8.0022
8.5	8.3172	8.5020	8.5015	8.5079	8.5168	8.3830	8.4985
9	8.7793	8.9948	9.0008	8.9943	9.0196	9.1270	9.0030
9.5	9.7034	9.5065	9.5206	9.5043	9.4927	9.5161	9.4996
10	10.1655	9.9997	10.0038	10.0056	9.9954	9.9394	10.0027

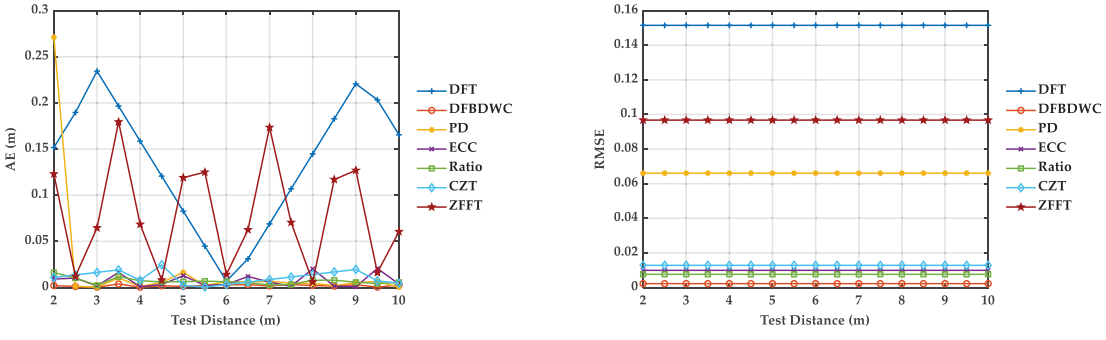


Figure 7. Simulation results of seven algorithms comparison at different distances. There are WGN signals with SNR of 10 dB in $s_B(n)$. (**a**) is AE. (**b**) is $RMSE$.

Additionally, in order to illustrate that this algorithm can reduce the influence of the picket fence effect and spectrum leakage further, we conduct a simulation with $s_B(n)$ without WGN, that is $s_O(n)$, at different distances, whose results of AE are as shown in Table 4. The DFBDWC algorithm still basically has the most minimum value of AE among all the spectrum correction algorithms. According to [34,35], the picket fence effect and spectrum leakage significantly decrease the precision of DFT when applying asynchronous sampling in practical applications. Additionally, there are disadvantages of each traditional spectrum correction algorithm that are described in Section 1. Therefore, the distance calculation accuracy of the DFBDWC algorithm is much better than any other six algorithms when processing with $s_O(n)$, which can prove our new spectrum correction

algorithm not only decreases the influence of spectrum leakage, but also reduces the picket fence effect.

Table 4. Simulation results of AE comparison between seven algorithms. There are no WGN signals in $s_B(n)$.

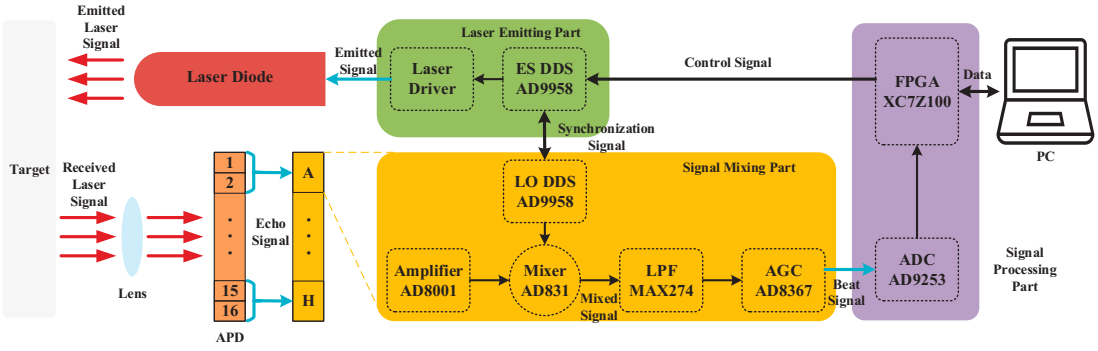
Test Distance (Real Distance) (m)	*AE* (cm)						
	DFT	**PD**	**ECC**	**Ratio**	**CZT**	**ZFFT**	**DFBDWC**
2	15.1729	27.1155	0.7593	1.6049	1.0918	12.2982	0.0064
2.5	18.9662	0.3612	0.5100	1.2172	1.3648	1.2092	0.0066
3	23.4474	0.1285	0.0044	0.3879	1.6378	6.4557	0.0118
3.5	19.6541	0.4905	0.9653	0.9247	1.9107	17.9420	0.0069
4	15.8609	0.0843	0.4251	0.7488	0.7735	6.8530	0.0112
4.5	12.0677	0.1837	0.0044	0.3630	2.4567	0.8119	0.0022
5	8.2744	0.0604	1.1222	0.6495	0.2276	11.9009	0.0109
5.5	4.4812	0.1957	0.3470	0.5298	0.0453	12.4969	0.0273
6	0.6880	0.2423	0.0084	0.3479	0.3183	1.4078	0.0087
6.5	3.1053	0.2089	1.2866	0.5007	0.5913	6.2571	0.0013
7	6.8985	0.2696	0.2793	0.4013	0.8642	17.3461	0.0036
7.5	10.6917	0.3016	0.0156	0.3363	1.1372	7.0517	0.0027
8	14.4850	0.2962	1.4637	0.4075	1.4102	0.6132	0.0093
8.5	18.2782	0.3348	0.2210	0.3157	1.2741	11.7023	0.0044
9	22.0714	0.3613	0.0266	0.3265	1.9561	12.6955	0.0084
9.5	20.3421	0.3672	1.6550	0.3435	0.7282	1.6065	0.0054
10	16.5489	0.3971	0.1714	0.2539	0.4552	6.0584	0.0024

The simulation results demonstrate that the DFT algorithm cannot accurately obtain the distance value, since it cannot overcome the problems described in Section 1. Although the other five algorithms have improved the accuracy of $R_C(p)$ compared with the DFT algorithm and they have achieved a certain spectrum correction effect, they are still inferior to the performance of the DFBDWC algorithm. As a consequence, our algorithm will improve the accuracy of distance calculation.

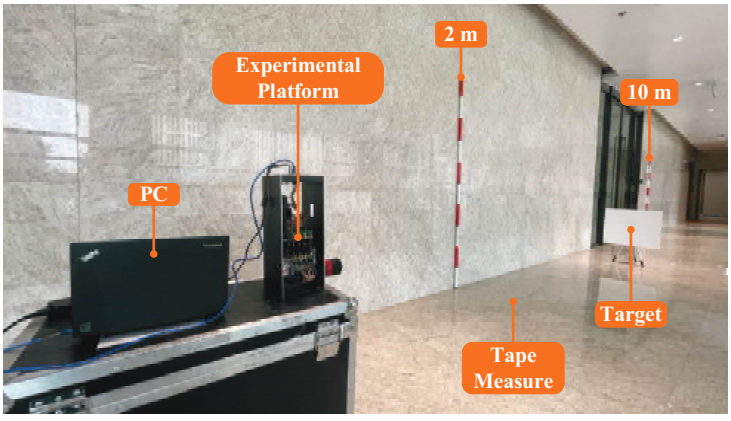
3.2. Experiment

The experimental platform and scene in the experiment part are as shown in Figure 8.

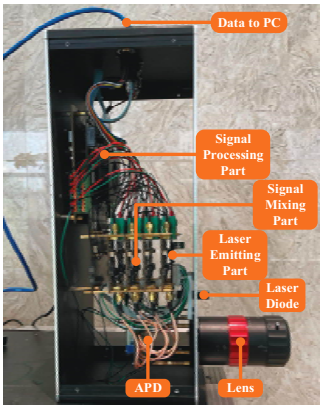
The laser diode whose power is 500 mW and wavelength is 950 nm is driven by an emitted signal generated by an emitted signal (ES) DDS named AD9958. The laser beam is reflected on the target surface at $R_T(p)$ and focused by the lens; then, the echo signal is outputted by an APD with 16 linearly arrayed receiving units. In each signal mixing part, two series of echo signals can be processed. However, the echo signal is too weak and needs to be amplified to a certain amplitude by an amplifier named AD8001, after which a local oscillator (LO) DDS synchronized by ES DDS is mixed with it in a mixer named AD831 to form a mixed signal that contains two frequency values because of the working principle in the mixer. The large frequency value, which is an interference, needs to be filtered out by an LPF named MAX274 whose bandwidth is 150 kHz, and the small frequency value passed through an automatic gain control (AGC) named AD8367 is amplified. Then, the beat signal of appropriate amplitude can be obtained. In the signal processing part, the beat signal is sampled by an ADC named AD9253. Finally, the data are transferred to a FPGA named XC7Z100 for temporary storage and transmitted to the PC for the distance calculation using the DFBDWC algorithm. The experimental platform is placed in a corridor with a length of 15 m, and the distance between the target and APD is considered to be $R_T(p)$, whose value is shown in Table 2. In order to place the target in a precise position, a tape measure with centimeter accuracy is used specially, and its starting position is the surface of APD. At the same time, two benchmarks are placed at 2 and 10 m, respectively, to indicate the placement range of the target.



(**a**)



(**b**)

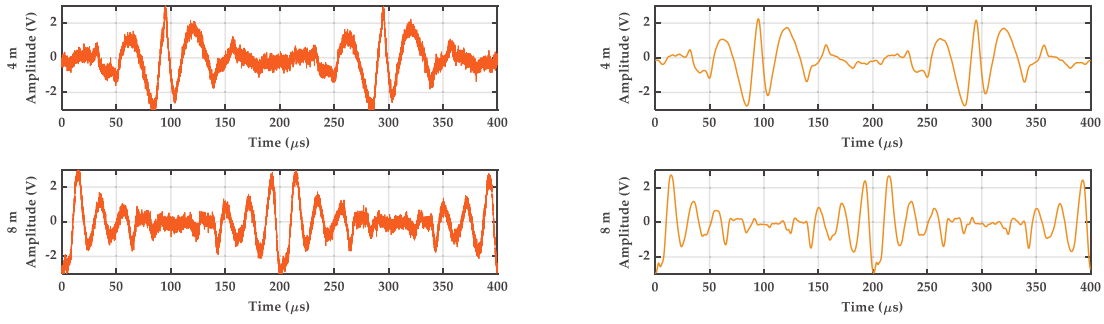


(**c**)

Figure 8. The experimental platform and scene in the experiment part. (**a**) Scheme of this experimental platform. (**b**) Experimental platform in the experimental scene. (**c**) Details of this experimental platform.

Above all, we utilize only one of the units in APD to receive the laser signal from a plane target. With the analysis and processing of the DFBDWC algorithm, we obtain a comparison diagram of $\widetilde{s}_B(n)$ and $s_B(n)$ at two specific distances in the time domain, which is as shown in Figure 9. It can be seen that they are different from $\widetilde{s}_B(n)$ and $s_B(n)$ in the simulation part, but they still contain the frequency information of $s_B(n)$. Compared with $s_B(n)$, the WGN interference in $\widetilde{s}_B(n)$ is reduced well, and the wave pattern of $\widetilde{s}_B(n)$ becomes smoother. This indicates that the algorithm can suppress the interference of noise in the beat signal and retain the useful information in the signal.

Similarly, we compare the performance of the DFBDWC algorithm with other six algorithms in terms of computed distances in Table 5 and error analysis in Figure 10. We can note from Figure 10 that as for the other six algorithms, the AE of DFT and CZT algorithm has the maximum value, respectively, at the first and the last test distance. Overall, the DFT algorithm has the biggest jump of AE, while the ECC algorithm has the smallest jump of AE. At each test distance, the AE and $RMSE$ of the DFBDWC algorithm basically have the minimum value. This indicates that the performance of this algorithm is the best in all these algorithms. Additionally, the maximum and minimum AE of each algorithm are listed in Table 6. The maximum AE is decreased from 0.7937 to 0.0407 m by using the DFBDWC algorithm. As expected, this algorithm overcomes the problem to a certain extent caused by spectrum leakage and the picket fence effect and improves the accuracy of distance calculation.

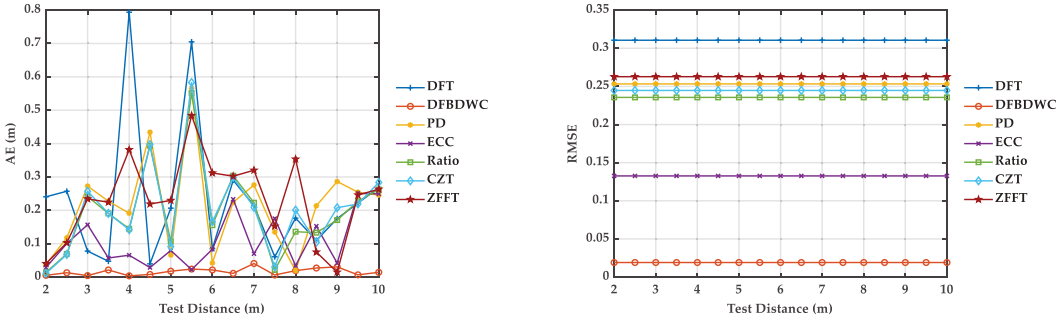


(**a**) (**b**)

Figure 9. Experimental results of signals comparison at two different distances in time domain. (**a**) is beat signals $s_B(n)$. (**b**) is reconstruction signals $\widetilde{s}_B(n)$.

Table 5. Experimental results of computed distances comparison between seven algorithms. There are WGN signals with SNR of 10 dB in $s_B(n)$.

Test Distance (Real Distance) (m)	Computed Distance (m)						
	DFT	PD	ECC	Ratio	CZT	ZFFT	DFBDWC
2	2.2406	2.0378	2.0288	2.0142	2.0112	2.0405	1.9947
2.5	2.2428	2.3821	2.3995	2.4305	2.4324	2.3972	2.5130
3	3.0778	3.2725	3.1569	3.2413	3.2570	3.2344	3.0044
3.5	3.4525	3.2725	3.4428	3.3091	3.3091	3.2757	3.4788
4	4.7937	4.1913	3.9345	4.1448	4.1420	4.3819	4.0037
4.5	4.5397	4.9340	4.5286	4.8935	4.8972	4.7189	4.4921
5	4.7935	4.9340	5.0797	4.8935	4.9086	4.7702	5.0177
5.5	4.7951	4.9340	5.4775	4.9486	4.9170	5.0159	5.5240
6	6.0848	6.0423	6.0834	6.1554	6.1661	6.3121	5.9786
6.5	6.7901	6.7241	6.2667	6.8037	6.7990	6.8023	6.5111
7	6.7856	6.7241	7.0700	6.7772	6.7912	6.6798	6.9593
7.5	7.5608	7.6350	7.6754	7.5213	7.5307	7.6526	7.5062
8	7.8234	7.9807	7.9655	7.8641	7.7992	7.6465	8.0192
8.5	8.3899	8.7136	8.3474	8.3670	8.3959	8.5749	8.5271
9	9.1752	8.7136	9.0420	9.1708	9.2079	9.0146	8.9694
9.5	9.7220	9.7542	9.7507	9.7275	9.7192	9.7459	9.4933
10	9.7324	9.7542	9.7507	9.7374	9.7163	9.7367	10.0143



(**a**) (**b**)

Figure 10. Experimental results of seven algorithms comparison at different distances. (**a**) is AE. (**b**) is $RMSE$.

Table 6. Maximum and minimum AE of each algorithm.

	DFT	PD	ECC	Ratio	CZT	ZFFT	DFBDWC
Maximum AE (m)	0.7937	0.5660	0.2507	0.5514	0.5830	0.4841	0.0407
Minimum AE (m)	0.0397	0.0193	0.0225	0.0142	0.0112	0.0146	0.0037

On the basis of experimental results above, we can conclude that because of decomposition and filter, WGN has been estimated accurately from the beat signal and reduced to a certain extent, which will improve the SNR of the beat signal. Additionally, due to the dual-window applied in correction, the energy is more concentrated and the influence of spectrum leakage has been decreased. Moreover, utilizing two main spectral lines at the peaks of these sub-signals with delay to calculate the frequency of the beat signal has alleviated the problem of poor accuracy of results caused by the picket fence effect, and it also avoid the interference that arises from using multiple spectral lines in some traditional spectrum correction algorithms. When the beat signal is not affected by WGN severely, correction will become the key step that makes the distance calculation more accurate in the DFBDWC algorithm.

In addition to the accuracy comparison of computed distance, we also calculate computation time consumed for these seven algorithms by processing the same group of sampled beat signals so as to judge the efficiency of each algorithm. Using an PC with CPU of Intel i7-7700 and RAM of 16 GB, we obtain the results that are as shown in Table 7. It can be seen that in different sample times, the computation time consumption of the ZFFT algorithm is the least, the DFBDWC algorithm is the most, and the other algorithms are almost the same. Additionally, with the increase in sample time, the computation time consumption of the DFBDWC algorithm is doubled, and there are not many rises in the other algorithms. This is because decomposition and filter in this algorithm have to process by iterative operation. The larger the sampled points of the beat signal, the more computation time consumption will be needed. In practical application, we only focus on the accuracy of the distance calculation, while we do not require any real-time computation. Therefore, we sacrifice the efficiency for the precision of our algorithm.

Table 7. The comparison of average computation time consuming between seven algorithms.

Sample Time (μs)	Computation Time Consuming (s)						
	DFT	PD	ECC	Ratio	CZT	ZFFT	DFBDWC
100	0.0423	0.0449	0.0458	0.0452	0.0441	0.0243	2.3891
200	0.0444	0.0501	0.0468	0.0464	0.0460	0.0259	5.6616

It can be found from the experimental results that the performance of each algorithm is consistent with simulation results. This indicates that simulation has achieved the real situation well, and the parameters for evaluating the performance of these algorithms is also reasonable. However, the values of each parameter obtained in experiment are larger than those in simulation, which is caused by errors from the experimental platform, the factors of the environment and the target placement, such as the sensitivity of APD, the response speed of the laser diode, the bandwidth of the emitted signal, the interference of ambient light, and the accuracy of distance and angle when we place the target. This can still verify that the DFBDWC algorithm reduces the influence of WGN, spectrum leakage, and the picket fence effect. Moreover, it performs the best among the existing spectrum correction algorithms, and the maximum AE of it is not more than 0.05 m.

4. Conclusions

In this article, we proposed a new spectrum correction algorithm named DFBDWC, and built an experimental platform based on the principle of the FMCW laser ranging system. The experimental platform outputs the data of the beat signal, which is analyzed

and processed by the DFBDWC algorithm in the PC, and the target distance detected by the system is obtained. Comparing this algorithm to the traditional DFT algorithm and other spectrum correction algorithms in both the simulation and the experiment, including the PD, ECC, Ratio, CZT, and ECC algorithm, we achieve the performance evaluation of this algorithm. The results indicate that DFBDWC algorithm can reduce the influence of WGN, spectrum leakage, and the picket fence effect. Additionally, it can also improve the accuracy and frequency resolution of the beat signal. The maximum absolute error of the target distance calculated by this algorithm is reduced from 0.7937 m of the DFT algorithm to 0.0407 m, which is the best among all the spectrum correction algorithms. The most remarkable performance of our algorithm is because decomposition can estimate WGN accurately in the beat signal and the filter reduces it to a certain extent. The double Hann window applied in correction concentrates more energy in the spectrum, which minimizes the impact of spectrum leakage well. Utilizing two main spectral lines at the peaks of these sub-signals with a delay to calculate the frequency of the beat signal has alleviated the problem of poor accuracy of results caused by the picket fence effect, and it also avoids the interference that arises from using multiple spectral lines in some traditional spectrum correction algorithms. Therefore, the DFBDWC algorithm can improve the performance of the FMCW laser ranging system. In future work, it is necessary to upgrade this platform of the system, such as by choosing more sensitive APD, selecting a laser diode with a faster response speed, and increasing the bandwidth of the emitted signal, which makes it adapt to this algorithm better. In addition, we still need to optimize the structure and computational complexity of our algorithm so that the efficiency of distance calculation in engineering can be greatly raised while keeping the accuracy. Furthermore, we will carry out experiments by utilizing 16 units in APD to figure out the surface fitting uncertainty for different object shapes of this system so that it could make our algorithm and platform more valuable for 3D scanning.

Author Contributions: Y.H. and P.S. proposed the decomposition filtering based dual-window correction algorithm for FMCW laser ranging system. Y.H., X.W., and Z.P. built the experimental platform and analyzed the simulation and experimental results. All authors have read and agreed to the published version of the manuscript.

Funding: The research was funded by Beijing Advanced Innovation Center for Intelligent Robots and Systems, grant number 2019IRS13.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Amann, M.C.; Bosch, T.; Lescure, M.; Myllyla, R.; Rioux, M. Laser ranging: A critical review of usual techniques for distance measurement. *Opt. Eng.* **2001**, *40*, 10–19.
2. Borkowski, J.; Mroczka, J. LIDFT method with classic data windows and zero padding in multifrequency signal analysis. *Measurement* **2010**, *43*, 1595–1602. [CrossRef]
3. Rife, D.C.; Vincent, G.A. Use of the discrete Fourier transform in the measurement of frequencies and levels of tones. *Bell Syst. Tech. J.* **1970**, *49*, 197–228. [CrossRef]
4. Grandke, T. Interpolation algorithms for discrete Fourier transforms of weighted signals. *IEEE Trans. Instrum. Meas.* **1983**, *32*, 350–355. [CrossRef]
5. Ming, X.; Kang, D. Corrections for frequency, amplitude and phase in a fast Fourier transform of a harmonic signal. *Mech. Syst. Signal Proc.* **1996**, *10*, 211–221. [CrossRef]
6. Agrez, D. Weighted multipoint interpolated DFT to improve amplitude estimation of multifrequency signal. *IEEE Trans. Instrum. Meas.* **2002**, *51*, 287–292. [CrossRef]
7. Belega, D.; Dallet, D. Frequency estimation via weighted multipoint interpolated DFT. *IET Sci. Meas. Technol.* **2008**, *2*, 1–8. [CrossRef]

8. Belega, D.; Dallet, D. Multifrequency signal analysis by interpolated DFT method with maximum sidelobe decay windows. *Measurement* **2009**, *42*, 420–426. [CrossRef]
9. Belega, D.; Dallet, D.; Petri, D. Accuracy of sine wave frequency estimation by multipoint interpolated DFT approach. *IEEE Trans. Instrum. Meas.* **2010**, *59*, 2808–2815. [CrossRef]
10. Mcmahon, D.R.A.; Barrett, R.F. An efficient method for the estimation of the frequency of a single tone in noise from the phases of discrete Fourier transforms. *Signal Process.* **1986**, *11*, 169–177. [CrossRef]
11. Zhu, L.; Li, H.; Ding, H.; Xiong, Y. Noise influence on estimation of signal parameter from the phase difference of discrete Fourier transforms. *Mech. Syst. Signal Proc.* **2002**, *16*, 991–1004. [CrossRef]
12. Kang, D.; Ming, X.; Xiaofei, Z. Phase difference correction method for phase and frequency in spectral analysis. *Mech. Syst. Signal Proc.* **2000**, *14*, 835–843. [CrossRef]
13. Luo, J.; Xie, M. Phase difference methods based on asymmetric windows. *Mech. Syst. Signal Proc.* **2015**, *54/55*, 52–67. [CrossRef]
14. Ding, K.; Cao, D.; Li, W. An approach to discrete spectrum correction based on energy centroid. *Key Eng. Mater.* **2006**, *321-323*, 1270–1273.
15. Offelli, C.; Petri, D. A frequency-domain procedure for accurate real-time signal parameter measurement. *IEEE Trans. Instrum. Meas.* **1990**, *39*, 363–368. [CrossRef]
16. Belega, D.; Dallet, D.; Petri, D. Accuracy of the normalized frequency estimation of a discrete-time sine-wave by the energy-based method. *IEEE Trans. Instrum. Meas.* **2012**, *61*, 111–121. [CrossRef]
17. Lin, H.; Ding, K. Energy based signal parameter estimation method and a comparative study of different frequency estimators. *Mech. Syst. Signal Proc.* **2011**, *25*, 452–464.
18. Zhang, Q.; Zhong, S.; Lin, J.; Huang, Y.; Nsengiyumva, W.; Chen, W.; Luo, M.; Zhong, J.; Yu, Y.; Peng, Z.; et al. Anti-noise frequency estimation performance of Hanning-windowed energy centrobaric method for optical coherence velocimeter. *Opt. Lasers Eng.* **2020**, *134*. [CrossRef]
19. Rabiner, L.; Schafer, R.; Rader, C. The chirp z-transform algorithm. *IEEE Trans. Audio Electroacoust.* **1969**, *17*, 86–92. [CrossRef]
20. Leng, J.; Shan, C. Application of chirp-z transformation in high accuracy measurement of radar. *Appl. Mech. Mater.* **2013**, *392*, 730–733. [CrossRef]
21. Li, D.; Liu, H.; Liao, Y.; Gui, X. A novel helicopter-borne rotating SAR imaging model and algorithm based on inverse chirp-z transform using frequency-modulated continuous wave. *IEEE Geosci. Remote Sens. Lett.* **2015**, *12*, 1625–1629.
22. Masso, E.; Bolognini, N. Dynamic multiple-image encryption based on chirp z-transform. *J. Opt.* **2019**, *21*, 035704. [CrossRef]
23. Qin, M.; Li, D.; Tang, X.; Zeng, C.; Li, W.; Xu, L. A fast high-resolution imaging algorithm for helicopter-borne rotating array SAR based on 2-D chirp-z transform. *Remote Sens.* **2019**, *11*, 1669. [CrossRef]
24. Shen, S.; Nie, X.; Tang, L.; Bai, Y.; Zhang, X.; Li, L.; Ben, D. An improved coherent integration method for wideband radar based on two-dimensional frequency correction. *Electronics* **2020**, *9*, 840. [CrossRef]
25. Wei, D.; Nagata, Y.; Aketagawa, M. Partial phase reconstruction for zero optical path difference determination using a chirp z-transform-based algorithm. *Opt. Commun.* **2020**, *463*, 125456. [CrossRef]
26. Lyons, R.G. The Zoom FFT. In *Understanding Digital Signal Processing*, 3rd ed.; Bernard, G., Ed.; Pearson Education: Boston, MA, USA, 2011; pp. 548–550.
27. Al-Qudsi, B.; Joram, N.; Strobel, A.; Ellinger, F. Zoom FFT for precise spectrum calculation in FMCW radar using FPGA. In Proceedings of the 2013 9th Conference on Ph.D. Research in Microelectronics and Electronics, Villach, Austria, 24–27 June 2013.
28. Huang, N.E.; Shen, Z.; Long, S.R.; Wu, M.C.; Shih, H.H.; Zheng, Q.; Yen, N.C.; Tung, C.C.; Liu, H.H. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **1998**, *454*, 903–995. [CrossRef]
29. Torres, M.E.; Colominas, M.A.; Schlotthauer, G.; Flandrin, P. A complete ensemble empirical mode decomposition with adaptive noise. In Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Prague Congress Center, Prague, Czech Republic, 22–27 May 2011.
30. Colominas, M.A.; Schlotthauer, G.; Torres, M.E. Improved complete ensemble EMD: A suitable tool for biomedical signal processing. *Biomed. Signal Process. Control.* **2014**, *14*, 19–29. [CrossRef]
31. Diniz, P.S.R. *Adaptive Filtering: Algorithms and Practical Implementation*, 5th ed.; Springer Nature: Cham, Switzerland, 2020; pp. 157–160.
32. Huang, X.; Wang, Z.; Hou, G. New method of estimation of phase, amplitude, and frequency based on all phase FFT spectrum analysis. In Proceedings of the IEEE 2007 International Symposium on Intelligent Signal Processing and Communication Systems, Xiamen, China, 28 November–1 December 2007.
33. Su, T.; Yang, M.; Jin, T.; Flesch, R.C.C. Power harmonic and interharmonic detection method in renewable power based on Nuttall double-window all-phase FFT algorithm. *IET Renew. Power Gener.* **2018**, *12*, 953–961. [CrossRef]
34. Zhang, D.; Sun, S.; Zhao, H.; Yang, J. Laser Doppler dignal processing based on trispectral interpolation of Nuttall window. *Optik* **2019**, *205*, 163364. [CrossRef]
35. Lin, H.C. Power harmonics and interharmonics measurement using recursive group-harmonic power minimizing algorithm. *IEEE Trans. Ind. Electron.* **2012**, *59*, 1184–1193. [CrossRef]

sensors

MDPI

Study Protocol

Mobile 5P-Medicine Approach for Cardiovascular Patients

Ivan Miguel Pires [1,2,*], Hanna Vitaliyivna Denysyuk [1], María Vanessa Villasana [3], Juliana Sá [4,5], Petre Lameski [6], Ivan Chorbev [6], Eftim Zdravevski [6], Vladimir Trajkovik [6], José Francisco Morgado [7] and Nuno M. Garcia [1]

1 Instituto de Telecomunicações, Universidade da Beira Interior, 6200-001 Covilhã, Portugal; hanna.denysyuk@ubi.pt (H.V.D.); ngarcia@di.ubi.pt (N.M.G.)
2 Escola de Ciências e Tecnologias, University of Trás-os-Montes e Alto Douro, Quinta de Prados, 5001-801 Vila Real, Portugal
3 Centro Hospitalar do Baixo Vouga, 3810-164 Aveiro, Portugal; 72152@chbv.min-saude.pt
4 Faculty of Health Sciences, Universidade da Beira Interior, 6200-506 Covilhã, Portugal; julianasa@fcsaude.ubi.pt
5 Centro Hospitalar e Universitário do Porto, 4099-001 Oporto, Portugal
6 Faculty of Computer Science and Engineering, SS. Cyril and Methodius University, 1000 Skopje, North Macedonia; petre.lameski@finki.ukim.mk (P.L.); ivan.chorbev@finki.ukim.mk (I.C.); eftim.zdravevski@finki.ukim.mk (E.Z.); trvlado@finki.ukim.mk (V.T.)
7 Computer Science Department, Polytechnic Institute of Viseu, 3504-510 Viseu, Portugal; fmorgado@estgv.ipv.pt
* Correspondence: impires@it.ubi.pt; Tel.: +351-966-379-785

Abstract: Medicine is heading towards personalized care based on individual situations and conditions. With smartphones and increasingly miniaturized wearable devices, the sensors available on these devices can perform long-term continuous monitoring of several user health-related parameters, making them a powerful tool for a new medicine approach for these patients. Our proposed system, described in this article, aims to develop innovative solutions based on artificial intelligence techniques to empower patients with cardiovascular disease. These solutions will realize a novel 5P (Predictive, Preventive, Participatory, Personalized, and Precision) medicine approach by providing patients with personalized plans for treatment and increasing their ability for self-monitoring. Such capabilities will be derived by learning algorithms from physiological data and behavioral information, collected using wearables and smart devices worn by patients with health conditions. Further, developing an innovative system of smart algorithms will also focus on providing monitoring techniques, predicting extreme events, generating alarms with varying health parameters, and offering opportunities to maintain active engagement of patients in the healthcare process by promoting the adoption of healthy behaviors and well-being outcomes. The multiple features of this future system will increase the quality of life for cardiovascular diseases patients and provide seamless contact with a healthcare professional.

Keywords: 5P-Medicine; digital health; mobile bio-sensing for medicine; patient empowerment technologies; artificial intelligence; cardiovascular diseases

Citation: Pires, I.M.; Denysyuk, H.V.; Villasana, M.V.; Sá, J.; Lameski, P.; Chorbev, I.; Zdravevski, E.; Trajkovik, V.; Morgado, J.F.; Garcia, N.M. Mobile 5P-Medicine Approach for Cardiovascular Patients. *Sensors* **2021**, *21*, 6986. https://doi.org/10.3390/s21216986

Academic Editor: Manuel José Cabral dos Santos Reis

Received: 16 September 2021
Accepted: 18 October 2021
Published: 21 October 2021

1. Introduction

The health care systems traditionally followed the paternalistic approach [1]. In recent years, there has been a noticeable paradigm shift towards patient- and community-centered strategies, empowering those approaches with modern technologies [2–4]. Today, it is impossible to imagine living without ubiquitous technologies such as smartphones and smart devices. All of us are connected using the Internet, and each of our moves is either being recorded or analyzed. Smartphones, sensors, and smart devices also allow measurement of various parameters, contributing to building a general model of our personal and our health care profile [5–7]. These and other aspects allow employment of the novel technologies for patient empowerment, in a sense that the patients themselves

are directly included in their healthcare management. Healthcare is not an institutional or hospital-based service, but the patients themselves and their communities are at the service center [8]. However, the application of modern paradigms is yet to be studied and employed to their full potential [9]. It is especially true for Cardiovascular patients.

Based on World Health Organization (WHO) data published in 2016, cardiovascular diseases are highly prevalent around the world [10]. Still, the middle- and low-income countries report higher prevalence with significant worldwide rates of morbidity and mortality [11]. Cardiovascular diseases are considered chronic non-infectious diseases related to different risk factors, including arterial hypertension, diabetes, hyperlipidemia, obesity, smoking, an imbalanced diet, and lack of physical activity [12]. The control and treating of these diseases include the normalization of blood pressure, the reduction of stress, healthy nutrition, and the increasing level of physical activity [13].

In general, the technology can be combined for the creation of complex solutions with complex measurements [14]. At this point, the 5P-Medicine concept is essential for the design of solutions to Predictive, Preventive, Participatory, Personalized, and Precision Medicine [15]. Furthermore, it allows the promotion of several actions, prior, during, and after some healthcare problems [16]. Still, the most important is the technology enables medical people to act just in time. Combining the different concepts also allows predicting the future and increasing the effectiveness of treatment to a minimum of 40%.

Nowadays, technology is part of the different daily activities. The various devices include the possibility of connection to the Internet [17]. Most of them also have a set of embedded sensors, including an accelerometer, magnetometer, gyroscope, and microphone, which allows the acquisition of different types of data [18]. In addition, these devices can add connections to other sensors, including electrocardiography, electromyography, pressure sensors, sphygmomanometer, among others, that allow the collection of medical data. It will enable the constant monitoring and the creation of patterns of the different diseases for the acquired data, allowing a remote evaluation of the patients [19]. Technology also allows the communication that promotes telemedicine and telemonitoring, making the patient independent, promoting patient empowerment, and centering the medical treatments in the patient [20].

The main goal of this paper consists in the proposal of a novel system architecture that considers the 5P-Medicine paradigm (Predictive, Preventive, Participatory, Personalized, and Precision Medicine) approach to empowering cardiovascular patients. Currently, the technology is improving, and the different measurements can be performed anywhere with the high commodity for the patients. The secondary endpoints consist of the different prospective achievements needed for the creation of the proposed system:

- Before the planning of the system, analysis of the state-of-the-art about the current applications which use mobile and wearable personal devices for promoting personalized digital health care must be performed.
- Identify the implementation challenges when applying the approach to real implementations.
- Analyze the required device features, because, for the use and implementation of the system, a minimum hardware and software requirements are needed.
- Analyze the required sensors because different sensors are needed for the data acquisition that will help the healthcare professionals in the monitoring of the cardiovascular patients.
- Increase the patients' autonomy with the easy and seamless contact with healthcare providers and professionals.

This paragraph ends the introduction. In continuation, Section 2 offers state-of-the-art patient empowerment, the 5P-Medicine concept, wearable devices, smart mobile devices, cardiovascular diseases and technology, and bio-signals acquisition and processing. Next, the research background, research design, and expected results are presented in Section 3. Finally, the analysis and further implementation of the system are discussed in Section 4, presenting the conclusions of this paper in Section 5.

2. State-of-the-Art

This section started with state-of-the-art representing the current stage of the 5P medicine approach. Section 2.1 aims to identify novel contributions of the 5P medicine to the digital health management practice, in particular for cardiovascular disease. Section 2.2 represented the advanced methods and solutions for empowering patients with different chronic conditions. Section 2.3 captured current trends in the use of mobile and wearable personal devices for monitoring and collecting physiological parameters. Section 2.4 presents the technology to prevent and monitor cardiovascular diseases. Finally, Section 2.5 discussed the most suitable Bio-signals acquisition and processing techniques available in the literature.

2.1. 5P-Medicine Concept

More than ten years ago, Leroy Hood defined the concept of 4-P Medicine, i.e., Predictive, Preventive, Personalized, and Participatory [21], to highlight the future change of medical intervention from care to prevention. This concept evolved, with Pravettoni and Gorrini, into a 5-P model that included Psychocognitive medicine [22]. It recognizes that patients have behaviors, habits, and beliefs that influence their interaction with health. Adding this dimension to the biological entity is essential to empower the person to share decisions over his health [22]. The concept of 5-P Medicine includes Predictive, Preventive, Participatory, Personalized, and Precision Medicine [15]. New technologies have significantly developed with eHealth providing solutions to improve healthcare [23]. More recently, the use of mHealth to improve autonomy in the control of chronic diseases. As public health systems are being modernized worldwide, conventional medicine is undergoing a profound transformation, and new digital 5P-based medical models are emerging. It is becoming crucial to identify new disease monitoring and prevention methods using modern information and communication technologies. However, some challenges remain to enhance accessibility, determine the exact impact on health, know the financial consequences, and improve data security [24].

As public health systems are being modernized worldwide, conventional medicine is undergoing a profound transformation, and new digital 5P-based medical models are emerging. Therefore, developing new disease monitoring and prevention methods is crucial using modern information and communication technologies. The goal is to understand and implement how conventional medical approaches and medicine of the future will co-exist and interact.

2.2. Patient Empowerment

Health care systems have been shifting their delivery of care towards patient- and community-centered approaches. Shared decisions have increasingly replaced the past paternalistic, and hospital-based healthcare service model, self-care, self-management, and home-based care [25]. WHO advocates patient empowerment as an essential tool to promote health. It is defined as a process to educate and give tools to the patients by healthcare professionals to recognize community and cultural differences and the participation of the patients [26]. The effect of patient empowerment on health results has been studied in several chronic conditions such as diabetes and heart diseases [27,28]. Empowered patients tend to have a better quality of life and well-being [29], impacting health outcomes to be consistently proved.

The use of technology to promote patient empowerment is being widely discussed and analyzed. Systems and tools have been developed to encourage and maintain healthy behaviors, education, and disease self-management. Technological development is expected to reduce financial costs and contribute to the sustainability of health care through rethinking interactions between patients and professionals, overcoming geographical barriers, and developing home-based solutions [30,31]. Several projects developed tools and systems to promote patient autonomy using online platforms and mobile devices in many medical fields. Digital tools to encourage autonomy and treatment guidance not only for chronic

diseases, such as respiratory diseases, diabetes, palliative care, and acute infections [32–35]. The transformation of healthcare and medicine through technology has been predicted for several years. However, current changes in healthcare delivery globally, boosted by the COVID-19 pandemic, promote a vital momentum to drive digital transformation in healthcare [36].

2.3. Wearable and Smart Mobile Devices

Researchers and technology companies have explored the use of wearable sensors to monitor physiological parameters and activities for the past decade. These devices can record real-time information through usable gadgets or being incorporated into clothing. They can measure physiological signals, such as heart rate, body temperature, arterial oxygen saturation, breathing rate, and body movement. They also have wireless communication modules integrated with mobile devices [37]. Real-time feedback is useful both for patients and for healthcare professionals. Patients can better understand their disease and see immediate and objective results from their actions, allowing them to improve behaviors and be empowered to make decisions [38]. Health professionals can access individual data to provide personalized advice, predict events, prevent disease, early diagnosis, and chronic control conditions [39]. Of course, these outcomes can only be achieved if data is secure and reliable. However, there are still challenges facing the way sensors and systems are developed.

Additionally, there is also potential to increase sensors and wearables in clinical trials, accelerating knowledge and new treatments. For this, the concern about data safety is essential to improve the acceptance of the systems [40]. Different devices in the market embed or connect to reliable sensors to monitor various health parameters related to cardiovascular diseases [41,42].

2.4. Cardiovascular Diseases and Technology

Cardiovascular diseases are highly prevalent across the globe representing 31% of global deaths [10], with half of deaths occurring in the middle- and low-income countries. These diseases are related to unhealthy behaviors and poor control of chronic conditions such as hypertension, diabetes, obesity, and cardiac failure [12].

The cornerstone of cardiovascular diseases management and prevention is based on interventions to motivate lifestyle modification and adherence to effective cardiovascular medications. Successful strategies to promote smoking cessation, increase physical activity levels, encourage a healthy diet, and improve medication adherence are associated with improvements in morbidity and reductions in mortality [43,44]. However, given the millions of people at risk for or with cardiovascular diseases, there are practical, logistical, geographical, and financial challenges in delivering comprehensive risk factor management to diverse populations. Health systems worldwide are charged with finding ways to reach more people in efficient and scalable ways.

The use of technology to prevent and monitor cardiovascular diseases has been tested with positive results [45,46], leading to new clinical practice recommendations [47]. Moreover, there is evidence of the need to develop these tools to achieve more accurate results and disseminate their adoption [48].

2.5. Bio-Signals Acquisition and Processing

The data acquisition and processing of sensors' data have been studied in the literature [49–51]. They consist of the instrumentation of the different individuals with wearable and smart mobile devices connected to other external sensors to increase the data acquisition capabilities [52–54]. Various studies use cloud servers to store the data acquired in natural environments [55–58]. In general, acquiring the data is also part of a system that includes data processing, cleaning, imputation, fusion, and classification [59]. The data cleaning mainly consists of removing the noisy data for the correct perception of the data acquired with different methods, including low-pass filter and high-pass filter [52]. The

data imputation measures the data that failed in the data acquisition process. Different methods can be implemented, including K-Nearest Neighbors imputation [60]. The feature extraction consists of analyzing the data related to cardiovascular diseases, including the heart rate, heart rate variability, different variables, and measurements associated with the QRS complex, and other measures [61]. However, one sensor/variable is not sufficient for complex measurements, and the fusion of the different data must be performed [62–64]. The final stage consists of the data classification. It may include various AI techniques, including Artificial Neural Networks, Support Vector Machines, Decision Trees, and Ensemble Learning algorithms [65]. Finally, the different measurements and machine learning methods will be more accurate using the Big Data concept in healthcare [66–69].

3. Methods and Expected Results

This section is the main section of this scientific paper, presenting the research background in Section 3.1. Next, Section 3.2 presents the different stages of the design of the proposed system. Section 3.3 presents the different methods that will be expected to be used for the measurement of patients' empowerment. Finally, Section 3.4 presents an overview of the expected results to be obtained with the proposed system.

3.1. Research Background

The proposed approach, presented in Figure 1 intends to give a solution that implements the 5P-Medicine paradigm [15], including Precision, Predictive, Preventive, Participatory, and Personalized Medicine related to cardiovascular diseases. For the final implementation, each of the 5Ps implementations will be researched for the final combination. This study aims to integrate the knowledge of different sciences, including computer, mathematics, and medical sciences, to research, develop, validate, and disseminate the developed technological solution.

For this purpose, there are different solutions available in the market. Still, no one includes the different concepts proposed in this paper. We intend to test the proposed system with cardiovascular patients, where the data acquisition will be performed with various sensors available in the market, including sensors available in mobile devices, sensors that can be connected to a BioPlux device [70], and others. These sensors have easy positioning, and all people may use these sensors to acquire health parameters. Still, some of these sensors are expensive. Finally, it includes the connection to a server for further analysis by medical people. In addition, the system must allow communications between healthcare professionals and patients.

The design and development of the proposed solution will consist of executing the main tasks, such as planning and development of data acquisition methods with the technological devices and collaboration of professionals, execution of data acquisition process for the creation of a database with different kinds of data, data analysis for the design of models for each of the 5P-Medicine, and integration of all developments.

The proposed system will be divided into five main parts consisting of each of the different concepts, including Precision, Predictive, Preventive, Participatory, and Personalized medicine. Then, as expected results, a system that integrates all the concepts will be presented in a solution that implements the 5P-Medicine paradigm for cardiovascular diseases.

Following the first P related to Prediction Medicine, the data acquired from the various sensors and wearable and smart mobile devices will be treated to create a method to predict future events or healthcare problems. It consists in utilizing the cloud computing tools for the expected model development. After being prepared, data set integration, the machine learning algorithms, modeling technique, and test design will be applied. Finally, the performance of predictive training models for cardiovascular diseases will be assessed to ensure the quality and reliability of the results. The metrics are numerical measures obtained from the confusion matrix that quantify the performance of a given classifier.

Figure 1. Workflow of the proposed system.

Following the second P related to Prevention Medicine, the medical and clinical health-related data sets, which are the prime concerns in healthcare decision-making, will be used on preventive models development. In addition, these models will be augmented with additional non-clinical background features to enhance the predictive capabilities of the preventive models [71]. With the method developed, the goal is to change people's behaviors or take adequate actions before healthcare problems occur.

Following the third P related to Participatory Medicine, the patient is the center of the system that will be part of the system. The patient must report important information in the system. The patient will be involved in the system's further development, where the decision is centered on the patient involved in the treatment and accompanying processes.

Following the fourth P related to Personalized Medicine, the system uses the sensors' data and the patients to adapt the intervention, recommendations, medication changes, and adequate medical exams. The monitoring and evaluation of the patient's process must be continuous, with no particular time for the decision-making. It is helpful because it gives independence to the patient. The system performs different measurements when the patients are in their regular environments, adapting their daily living approach. The patients' examination in their environment will be more reliable because they are examined in their daily living activities. Thus, the system will be integrated into the patients' living.

Finally, the fifth P, related to Precision Medicine, is intended to create an algorithm capable of predicting the moment and the healthcare problems with high accuracy and exactness.

3.2. Research Design

Based on the research background presented in Section 3.1, a global system was designed for the analysis of the data collected by the different sensors available in mobile and wearable devices, which will be sent to the cloud for further processing and generation of the notification to the patients and health systems. The whole design of the proposed system is presented in Figure 2.

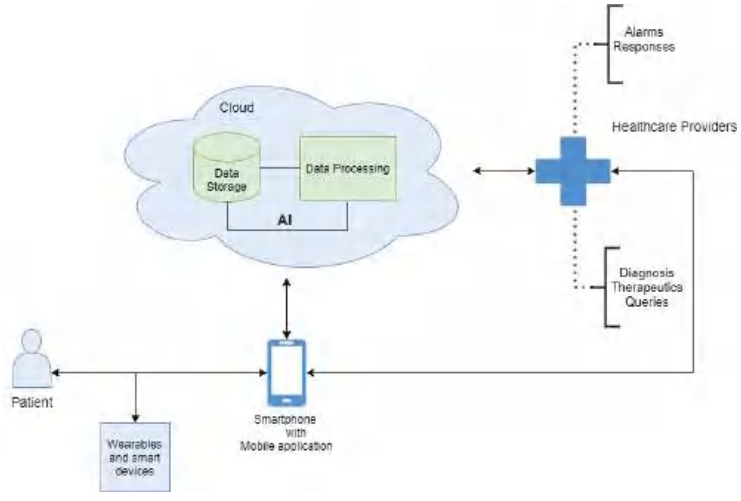


Figure 2. Architecture of the proposed system.

This study will be performed in five different iterations composed of different stages, presented in Figure 3 for each one of the 5Ps. In the following subsections, each of the parts for the research and development of the system will be presented. The proposed system will follow the Regulation (EU) 2017/745 of the European Parliament and the Council established for medical devices.

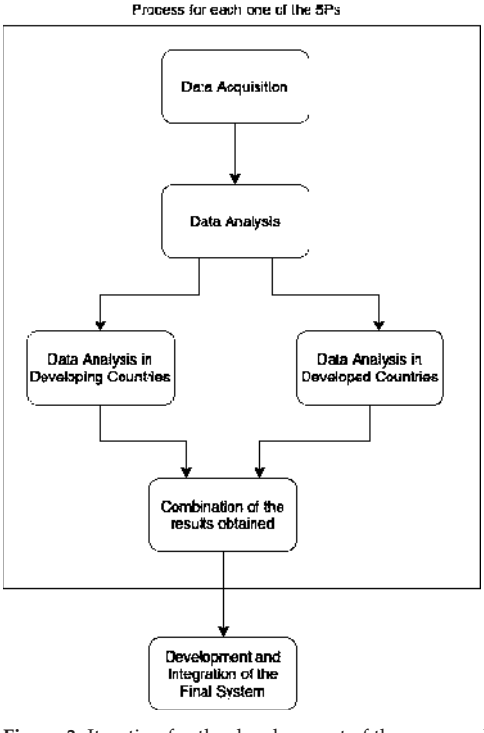


Figure 3. Iteration for the development of the proposed system.

3.2.1. Data Acquisition

Smartphones and connected wearable devices can quickly generate and collect a large-scale amount of diverse, complex, and dynamic data for analysis. In addition, these technologies are more economical and time-saving than traditional research mode since the entire process can be done remotely, anytime, and everywhere.

Gathering and analysis of clinical information will be performed with support from medical experts. This process involves human participants with cardiovascular diseases. Participation in the study is voluntary, and there is no compromise of the clinical diagnosis or management if the patient does not want to participate. Processing personal data will be carried out in compliance with the General Data Protection Regulation (GDPR), the data protection law of the European Union.

It consists of collecting medical data, including Electrocardiography (ECG) data, bioimpedance data, location data, personal data, data related to the various diseases of the participants, and data acquired from different sensors. The used sensors may be inertial, acoustic, and imaging, stored in a secure server.

The conditions to participate in the study are related to cardiovascular diseases, distributed equally by gender, and different age groups. A waiver of informed consent will be requested for patients in the hospital without conditions to consent and, when possible, given to the patient to sign afterward, as this is a non-experimental study. In the event of death or not signing the informed consent, the legal representative or the closest family members will be asked to determine the patients' participation in research studies, as it is a study of non-experimental character. In addition, the participants must have an Internet connection available to store the collected data remotely. The acquired data include vulnerable individuals, including older adults, limited capabilities, and other vulnerable people. The study consists of intervention activities by medical doctors for the acquisition of imaging data (or other types of data with medical equipment) to be sent to the data processing stage to control the treatments' evolution. As it involves different countries, the data from EU and non-EU countries will be imported/exported for further processing by the various technological and medical teams.

Before acquiring the different datasets related to each of the 5P-Medicine concepts, the analysis of the existing methods is performed for the correct planning of the data acquisition process. As the system must be patient-centered, the planning stage is crucial for appropriately developing the technique for data acquisition. Finally, the patients with cardiovascular diseases and group control of healthy people will be recruited.

The positioning of the sensors and the positioning of the mobile device are widely essential for the correct data acquisition, and the different constraints, including environmental and sensor conditions, affect the data acquisition process. Therefore, it can be considered as a limitation of this stage. Another limitation is the definition of the timeline for the acquisition using the device because these devices have limited memory, power processing, and battery capabilities. The research is intended to use open-source development. Thus, the data acquisition will be developed for mobile devices with the Android operating system for smartphones and smartwatches. The smartphones will collect the data from the other devices over-the-air.

The study protocol was approved by the Ethics committee of Universidade da Beira Interior, Covilhã, Portugal, with the reference CE-UBI-Pj-2021-041:ID969.

3.2.2. Data Analysis

It consists of analyzing the data acquired based on the information provided by medical people. It includes artificial neural networks, statistical, and other computational methods to analyze the data. The analysis of the data occurs after the data acquisition.

Before starting the study, the participants fill on a questionnaire related to personal data, i.e., age, gender, diseases, habits, location, biometric data, health data, and other personal data that may be used for the analysis comparison of the different participants on the study. However, the data are always anonymized for technical analysis. The

participant is only identified by one ID attributed by the healthcare professional in the study's invitation, which is only known by the healthcare professional. This data are only used to support the study results, and it will never be used to identify the participants for the scientific community.

Before analyzing the different datasets related to each 5P-Medicine concept, the definition of the possible measurements is needed with the support of medical people. The correct measurements have been essential as the beginning for the excellent performance of the data analysis. It is predicted that a possible limitation is that the data collected is not known, and several experiments are needed before the definition of the concrete variables to analyze. The analyses will be performed with open-source programming technologies, and only some statistical analysis will require more robust software.

3.2.3. Data Analysis in Developing Countries

It consists of analyzing the data acquired in developing countries based on the information provided by healthcare professionals, patients' lifestyles, and healthcare and environmental conditions for data acquisition of the people from these countries. The developing countries are particular, and they need to be analyzed separately because the people's characteristics are different from developed countries.

3.2.4. Data Analysis in Developed Countries

It consists of analyzing the data acquired in developing countries based on the information provided by healthcare professionals, patients' lifestyles, and healthcare and environmental conditions for data acquisition of the people from these countries. The developing countries are particular, and they need to be analyzed separately because the people's characteristics are different from developing countries.

3.2.5. Development and Integration of Final System

This stage involves developing and integrating several self-contained components such as wearable personal devices connecting for data acquisition, biometrics signal processing, and machine learning algorithms utilizing the cloud computing tools for data processing and model development. As a result, the system will advance the smart solution in personalized and precision eHealth. Furthermore, its innovative solutions will increase the quality of life for cardiovascular diseases patients and be carried out by healthcare professionals, contributing to 5P-Medicine.

3.3. Methods for Patient Empowerment and System Analysis

Regarding measures improvements for patient empowerment and system effectiveness, the quantitative and qualitative methods will be applied. In the literature, many authors defended that patients' active engagement through digital tools in their health could increase self-monitoring effectiveness and empowerment, improving and maintaining a healthier lifestyle.

As engagement is important for long-term use, it is crucial to design technology in such a way that the chance of engagement is high [72]. One way to improve engagement is by incorporating a combination of persuasive design and behavioral change techniques [73]. The combined term for this is persuasive features. Another factor that influences engagement is usability. Usability relates to the quality of the technology in terms of easiness to learn and use it [74]. Usability and engagement together form the user experience. Usability and persuasive features will have a small direct influence on the adherence but will mainly affect engagement. High engagement is likely to improve adherence.

The log data analysis is a quantitative study. Log data analysis can provide information about the use of the technology for numerous users without interfering with normal behavior [75]. Although log data analysis can show interesting differences, it often remains unclear why these differences in use exist. Qualitative studies can only reach some total users, but it gives more insight into barriers and facilitators [76]. The qualitative method

is the usability tests and interviews. In this study, the log data analysis will be provided insight into general use and difference in use from the proposed system for a specific time. The usability tests and interviews will be held after the log data analysis.

The availability of more information about the facilitators will motivate patients into adherent behavior. It is important to achieve effective long-term usage of our proposed system for cardiovascular patients. Giving the patient more suitable and credible information, instant support, and feedback by a health care provider about their health will ensure an opportunity to increase their motivation for self-management capability.

3.4. Expected Results

After all the developments and tests, it is expected to be available a system for implementing the different concepts of 5P-Medicine for the accompanying, monitoring, and empowerment of cardiovascular patients. Furthermore, the system is expected to be accepted by the medical communities integrated with this system's design. The system will have the opportunity to be integrated into the National Health System to expand its use and empowering patients and healthcare professionals to achieve better health outcomes.

4. Discussion

Implementing a solution that empowers cardiovascular patients based on the 5P-Medicine concept will promote combining technology and medicine. It is an innovative system, and the exploration of the digitalization of the 5P-Medicine is only in the beginning, and further developments are needed. However, several challenges are hard to solve due to the different devices' characteristics and variety for medical purposes. One main reason for these challenges is the complexity of these developments, and the challenges of long-term and continuous monitoring, sending data and receiving results in real-time.

4.1. Multidisciplinary Approach Required

The proposed approach to implementing the treatment and monitoring process between healthcare professionals and computer science experts requires constant communication to plan and develop the final solution. The development of solutions is part of technological people, but knowing the correct collection, data treatment, and purposes is part of the medical people. Therefore, different joint expertise from other people will increase the reliability and validity of the system and future usage. Furthermore, this approach combines various sciences, including medicine, mathematical, and computer science knowledge, making the presented problem fascinating.

4.2. Experimental Challenges

The experimental challenges include several variants of potential problems related to the different activities to perform. The limited power processing and battery capabilities are significant challenges that must be considered while developing the software for experimental procedures using lightweight technologies. The next challenge consists of correctly positioning the sensors for accurate data acquisition, where the correct information must be provided to the patients by the mobile application. Furthermore, the connection between different devices and sensors must be available for proper data acquisition. The various connections must be inspected before the experimental data acquisition to avoid this problem. Another challenge consists in the Internet connection needed to store the acquired data into the cloud server. In this case, the system for obtaining the data must store the received data in offline mode to the mobile device's memory, and it will be sent to the cloud server when an Internet connection is available. The correct positioning of the sensors also reduces the data noise. During the data acquisition, the sensors may have failures that must have reduced effects with the proper modeling of the system.

4.3. Complexity and Specificity of System

Following the challenges with the sensors' data acquisition presented as experimental challenges, other challenges are related to the other components of the proposed system. Thus, several specificities of the system will increase the complexity of the system:

- *Data acquisition and processing:* It must have low latency times to acquire and process data. The frequencies of data acquisition from the different sensors/equipment must be adjusted to obtain better results.
- *Data processing:* The time for data processing can sometimes be high, and the solution's development time is affected. The nonexistence of rules for analyzing the data is another problem related to data processing, where the data should be tested with different processing methodologies.
- *Data fusion:* The acquired data have different natures, and the complexity of the data fusion and processing may be adapted to the different kinds of data. Various sensors retrieve distinct types of data.
- *Amount of data:* The proposed approach must be prepared to transmit and store a large amount of data related to different sensors.
- *Interaction between sensors and patients:* The patients must be taught and familiarized with the different sensors and mobile devices before using the system.
- *Acquisition of health parameters:* As the proposed approach is related to the acquisition of health parameters, the system must consider different rules for data protection and security during the data transmission and analysis.
- *Patients:* The proposed approach must be adapted to the patients. The data acquired from the different patients must be anonymized and labeled for the healthcare professional's knowledge responsible for the patient. Only features related to the different data types must be processed because the different data types may identify the patients. Finally, the time zones of the various countries must be controlled to synchronize the results obtained and contacts with different intervenients.
- *Features extracted from the data acquired:* During the development and testing of the developed methods, the best set of features must be identified to increase the results' accuracy.
- *User interface:* The user interface of the proposed approach must be user-friendly for the different ages of the people as the movements for older adults are more limited than children.

4.4. Modeling Challenges

The proposed approach includes technological and medical people from different countries and patients from the selected countries. The main challenge with this system is the existence of varying healthcare diseases related to cardiovascular problems that healthcare professionals must have previously identified. Furthermore, the data acquisition and processing methods must be planned with the knowledge about the diseases [77]. Furthermore, the algorithms for the automatic analysis must be developed for the different disorders. Moreover, the developed methods must be lightweight as preferable or executed in a remote server with the data sent by an Internet connection.

This approach consists of implementing a system that aggregated the measurements related to each of the 5P-Medicine concepts [15]. The method may have different specificities for each development stage that the literature must discover, and the nature of the sensors used healthcare professionals' knowledge.

The most important piece of the proposed system is the patient, and the developed system must be intelligent to be adapted to the different patients. Therefore, the system's adaption is crucial for the patient, and it is also vital for the healthcare professionals with a suitable communication method between them. The system must be comfortable for the patient to increase its use. As it includes communication, all the information must be securely stored and transmitted, and the different procedures on the system must be performed in an authenticated mode. Finally, the system should notify the patient of

different things in a non-intrusive method. The patients will have the opportunity to have contact with healthcare professionals from other countries. Still, the system must take into account the different time zones to schedule the meeting.

4.5. Integration with Real World

After the development and testing of the system, another challenge is integrating the system with the National Health Systems. Changes must be made in the National Health Systems, or adapters should be implemented to allow a seamless operation of both systems and achieve a wide usage of the proposed approach.

However, as this system is developed with different countries' cooperation, multilingual support must also be implemented in each nation. Furthermore, due to the diversity of cultures and different levels of development of the countries and their respective healthcare systems, the integration approach must take all the varying specificities of each country into consideration, both cultural and legal.

The final system must perform the different actions and analyses in real-time, adjusting the participating countries' various time zones. Furthermore, the marketing of the final product must be completed in the participating countries and languages to adapt the different methods to the people's perception. Finally, the system's integration must be performed with technical people working on the current systems.

4.6. Limitations of Study

This study presents the conceptual foundation of the proposed approach for integrating the 5P-Medicine paradigm within a healthcare system that will be, first and foremost, patient-centric and will implement all of the requirements of the 5P paradigm. This study analyzes the initial concepts and compares them with the current state of affairs while also presenting the current and future challenges this approach will face. The primary study's shortcoming is that the system is not yet implemented, and all of the analysis is based on the proposed concepts. The direct implications to the patients of such systems can be evident only after the system is implemented and a pilot run is executed in multiple countries.

5. Conclusions

Cardiovascular diseases are disorders of the heart and blood vessels and are a significant cause of disability and premature death worldwide. Therefore, individuals at higher risk of developing cardiovascular diseases must be encouraged to maintain active engagement in the healthcare process by promoting the adoption of healthy behaviors and well-being outcomes to prevent premature deaths. Advances in the field of computational intelligence, together with data from connected wearable and smart mobile devices, have made it possible to create recognition systems capable of identifying hidden patterns and valuable information.

In this paper, we presented a new efficiency system for a 5P-Medicine approach to the healthcare systems to manage cardiovascular diseases. This system will facilitate the administration of continual care and offer opportunities for maintaining patients' active engagement in the care process by promoting patients' psychological skills and well-being outcomes. The proposed system is defined as a vehicle to enrich patients and stakeholders through the intersection of medical informatics and public health business. As such, through our system, we will promote a new "state of mind" for medical professionals, marked by a global attitude and intention to improve worldwide health.

Author Contributions: Conceptualization, I.M.P., H.V.D., M.V.V. and J.S.; methodology, I.M.P., H.V.D. and N.M.G.; formal analysis, I.M.P., H.V.D., M.V.V. and J.S.; investigation, I.M.P., M.V.V. and J.S.; writing—original draft preparation, I.M.P., M.V.V., J.S., P.L., I.C., E.Z., V.T., J.F.M. and N.M.G.; writing—review and editing, I.M.P., H.V.D., M.V.V., J.S., P.L., I.C., E.Z., V.T., J.F.M. and N.M.G.; funding acquisition, I.M.P. and N.M.G. All authors have read and agreed to the published version of the manuscript.

Funding: This work is funded by FCT/MEC through national funds and co-funded by FEDER—PT2020 partnership agreement under the project **UIDB/50008/2020** (*Este trabalho é financiado pela FCT/MEC através de fundos nacionais e cofinanciado pelo FEDER, no âmbito do Acordo de Parceria PT2020 no âmbito do projeto UIDB/50008/2020*). This article is based upon work from COST Action IC1303–AAPELE–Architectures, Algorithms and Protocols for Enhanced Living Environments and COST Action CA16226–SHELD-ON–Indoor living space improvement: Smart Habitat for the Elderly, supported by COST (European Cooperation in Science and Technology). More information in www.cost.eu (accessed on 5 September 2021).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Meskó, B.; Drobni, Z.; Bényei, É.; Gergely, B.; Győrffy, Z. Digital health is a cultural transformation of traditional healthcare. *Mhealth* **2017**, *3*, 38. [CrossRef] [PubMed]
2. Bridges, J.F.; Crossnohere, N.L.; Schuster, A.L.; Miller, J.A.; Pastorini, C.; Aslakson, R.A. A patient and community-centered approach selecting endpoints for a randomized trial of a novel advance care planning tool. *Patient Prefer. Adherence* **2018**, *12*, 241. [CrossRef]
3. Goldfield, N.I.; Crittenden, R.; Fox, D.; McDonough, J.; Nichols, L.; Rosenthal, E.L. COVID-19 Crisis Creates Opportunities for Community-Centered Population Health: Community Health Workers: At the Center. *J. Ambul. Care Manag.* **2020**, *43*, 184–190. [CrossRef] [PubMed]
4. McAllister, M.; Dunn, G.; Payne, K.; Davies, L.; Todd, C. Patient empowerment: The need to consider it as a measurable patient-reported outcome for chronic conditions. *BMC Health Serv. Res.* **2012**, *12*, 1–8. [CrossRef]
5. Islam, S.R.; Kwak, D.; Kabir, M.H.; Hossain, M.; Kwak, K.S. The internet of things for health care: A comprehensive survey. *IEEE Access* **2015**, *3*, 678–708. [CrossRef]
6. Vashist, S.K.; Schneider, E.M.; Luong, J.H. Commercial smartphone-based devices and smart applications for personalized healthcare monitoring and management. *Diagnostics* **2014**, *4*, 104–128. [CrossRef] [PubMed]
7. Xu, K.; Chen, Y.; Okhai, T.A.; Snyman, L.W. Micro optical sensors based on avalanching silicon light-emitting devices monolithically integrated on chips. *Opt. Mater. Express* **2019**, *9*, 3985–3997. [CrossRef]
8. Dick, R.S.; Steen, E.B.; Detmer, D.E. *The Computer-Based Patient Record: An Essential Technology for Health Care*; National Academies Press: Washington, DC, USA, 1997.
9. Miah, S.J.; Hasan, J.; Gammack, J.G. On-cloud healthcare clinic: An e-health consultancy approach for remote communities in a developing country. *Telemat. Inform.* **2017**, *34*, 311–322. [CrossRef]
10. WHO. Cardiovascular Diseases (CVDs). 2016. Available online: https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds) (accessed on 5 September 2021).
11. Goradel, N.H.; Hour, F.G.; Negahdari, B.; Malekshahi, Z.V.; Hashemzehi, M.; Masoudifar, A.; Mirzaei, H. Stem cell therapy: A new therapeutic option for cardiovascular diseases. *J. Cell. Biochem.* **2018**, *119*, 95–104. [CrossRef]
12. Francula-Zaninovic, S.; Nola, I.A. Management of Measurable Variable Cardiovascular Disease' Risk Factors. *Curr. Cardiol. Rev.* **2018**, *14*, 153–163. [CrossRef] [PubMed]
13. Tang, G.Y.; Meng, X.; Li, Y.; Zhao, C.N.; Liu, Q.; Li, H.B. Effects of vegetables on cardiovascular diseases and related mechanisms. *Nutrients* **2017**, *9*, 857. [CrossRef] [PubMed]
14. Baraldi, E.; Gregori, G.L.; Perna, A. Network evolution and the embedding of complex technical solutions: The case of the Leaf House network. *Ind. Mark. Manag.* **2011**, *40*, 838–852. [CrossRef]
15. Gardes, J.; Maldivi, C.; Boisset, D.; Aubourg, T.; Vuillerme, N.; Demongeot, J. Maxwell®: An Unsupervised Learning Approach for 5P Medicine. *Stud. Health Technol. Inform.* **2019**, *264*, 1464–1465. [CrossRef]
16. Zheng, Y.L.; Ding, X.R.; Poon, C.C.Y.; Lo, B.P.L.; Zhang, H.; Zhou, X.L.; Yang, G.Z.; Zhao, N.; Zhang, Y.T. Unobtrusive Sensing and Wearable Devices for Health Informatics. *IEEE Trans. Biomed. Eng.* **2014**, *61*, 1538–1554. [CrossRef] [PubMed]
17. Kos, A.; Milutinović, V.; Umek, A. Challenges in wireless communication for connected sensors and wearable devices used in sport biofeedback applications. *Future Gener. Comput. Syst.* **2019**, *92*, 582–592. [CrossRef]
18. Amerini, I.; Becarelli, R.; Caldelli, R.; Melani, A.; Niccolai, M. Smartphone fingerprinting combining features of on-board sensors. *IEEE Trans. Inf. Forensics Secur.* **2017**, *12*, 2457–2466. [CrossRef]
19. Appelboom, G.; Camacho, E.; Abraham, M.E.; Bruce, S.S.; Dumont, E.L.; Zacharia, B.E.; D'Amico, R.; Slomian, J.; Reginster, J.Y.; Bruyère, O.; et al. Smart wearable body sensors for patient self-assessment and monitoring. *Arch. Public Health* **2014**, *72*, 1–9. [CrossRef] [PubMed]
20. Kvedar, J.; Coye, M.J.; Everett, W. Connected health: A review of technologies and strategies to improve patient care with telemedicine and telehealth. *Health Aff.* **2014**, *33*, 194–199. [CrossRef] [PubMed]

21. Hood, L.; Friend, S.H. Predictive, personalized, preventive, participatory (P4) cancer medicine. *Nat. Rev. Clin. Oncol.* **2011**, *8*, 184–187. [CrossRef] [PubMed]
22. Pravettoni, G.; Gorini, A. A P5 cancer medicine approach: Why personalized medicine cannot ignore psychology: P5 medicine. *J. Eval. Clin. Pract.* **2011**, *17*, 594–596. [CrossRef] [PubMed]
23. Van den Heuvel, J.F.; Groenhof, T.K.; Veerbeek, J.H.; van Solinge, W.W.; Lely, A.T.; Franx, A.; Bekker, M.N. eHealth as the Next-Generation Perinatal Care: An Overview of the Literature. *J. Med. Internet Res.* **2018**, *20*, e202. [CrossRef]
24. Granja, C.; Janssen, W.; Johansen, M.A. Factors Determining the Success and Failure of eHealth Interventions: Systematic Review of the Literature. *J. Med. Internet Res.* **2018**, *20*, e10235. [CrossRef] [PubMed]
25. Gluyas, H. Patient-centred care: Improving healthcare outcomes. *Nurs. Stand. (Royal Coll. Nurs. (Great Britain): 1987)* **2015**, *30*, 50–57. [CrossRef] [PubMed]
26. WHO Guidelines on Hand Hygiene in Health Care: First Global Patient Safety Challenge Clean Care Is Safer Care *Patient Empowerment and Health Care*; World Health Organization: Geneva, Switzerland, 2009.
27. Kambhampati, S.; Ashvetiya, T.; Stone, N.J.; Blumenthal, R.S.; Martin, S.S. Shared Decision-Making and Patient Empowerment in Preventive Cardiology. *Curr. Cardiol. Rep.* **2016**, *18*, 49. [CrossRef]
28. Lau, M.; Campbell, H.; Tang, T.; Thompson, D.J.; Elliott, T. Impact of Patient Use of an Online Patient Portal on Diabetes Outcomes. *Can. J. Diabetes* **2014**, *38*, 17–21. [CrossRef] [PubMed]
29. Risling, T.; Martinez, J.; Young, J.; Thorp-Froslie, N. Evaluating Patient Empowerment in Association With eHealth Technology: Scoping Review. *J. Med. Internet Res.* **2017**, *19*, e329. [CrossRef] [PubMed]
30. Calvillo, J.; Román, I.; Roa, L.M. How technology is empowering patients? A literature review. *Health Expect.* **2015**, *18*, 643–652. [CrossRef] [PubMed]
31. Lettieri, E.; Fumagalli, L.P.; Radaelli, G.; Bertele', P.; Vogt, J.; Hammerschmidt, R.; Lara, J.L.; Carriazo, A.; Masella, C. Empowering patients through eHealth: A case report of a pan-European project. *BMC Health Serv. Res.* **2015**, *15*, 309. [CrossRef] [PubMed]
32. Huniche, L.; Dinesen, B.; Grann, O.; Toft, E.; Nielsen, C. Empowering patients with COPD using Tele-homecare technology. *Stud. Health Technol. Inform.* **2010**, *155*, 48–54.
33. LeBaron, V.; Hayes, J.; Gordon, K.; Alam, R.; Homdee, N.; Martinez, Y.; Ogunjirin, E.; Thomas, T.; Jones, R.; Blackhall, L.; et al. Leveraging Smart Health Technology to Empower Patients and Family Caregivers in Managing Cancer Pain: Protocol for a Feasibility Study. *JMIR Res. Protoc.* **2019**, *8*, e16178. [CrossRef] [PubMed]
34. Tripoliti, E.E.; Karanasiou, G.S.; Kalatzis, F.G.; Naka, K.K.; Fotiadis, D.I. The Evolution of mHealth Solutions for Heart Failure Management. *Adv. Exp. Med. Biol.* **2018**, *1067*, 353–371. [CrossRef] [PubMed]
35. Ting, D.S.W.; Carin, L.; Dzau, V.; Wong, T.Y. Digital technology and COVID-19. *Nat. Med.* **2020**, *26*, 459–461. [CrossRef] [PubMed]
36. Smith, A.C.; Thomas, E.; Snoswell, C.L.; Haydon, H.; Mehrotra, A.; Clemensen, J.; Caffery, L.J. Telehealth for global emergencies: Implications for coronavirus disease 2019 (COVID-19). *J. Telemed. Telecare* **2020**, *26*, 309–313. [CrossRef] [PubMed]
37. Majumder, S.; Mondal, T.; Deen, M. Wearable Sensors for Remote Health Monitoring. *Sensors* **2017**, *17*, 130. [CrossRef] [PubMed]
38. Dunn, J.; Runge, R.; Snyder, M. Wearables and the medical revolution. *Pers. Med.* **2018**, *15*, 429–448. [CrossRef] [PubMed]
39. Greiwe, J.; Nyenhuis, S.M. Wearable Technology and How This Can Be Implemented into Clinical Practice. *Curr. Allergy Asthma Rep.* **2020**, *20*, 36. [CrossRef]
40. Kruse, C.S.; Frederick, B.; Jacobson, T.; Monticone, D.K. Cybersecurity in healthcare: A systematic review of modern threats and trends. *Technol. Health Care* **2017**, *25*, 1–10. [CrossRef] [PubMed]
41. Fanucci, L.; Saponara, S.; Bacchillone, T.; Donati, M.; Barba, P.; Sanchez-Tato, I.; Carmona, C. Sensing Devices and Sensor Signal Processing for Remote Monitoring of Vital Signs in CHF Patients. *IEEE Trans. Instrum. Meas.* **2013**, *62*, 553–569. [CrossRef]
42. Haghi, M.; Thurow, K.; Stoll, R. Wearable Devices in Medical Internet of Things: Scientific Research and Commercially Available Devices. *Healthc. Inform. Res.* **2017**, *23*, 4. [CrossRef]
43. Chow, C.K.; Jolly, S.; Rao-Melacini, P.; Fox, K.A.; Anand, S.S.; Yusuf, S. Association of Diet, Exercise, and Smoking Modification With Risk of Early Cardiovascular Events After Acute Coronary Syndromes. *Circulation* **2010**, *121*, 750–758.
44. Artinian, N.T.; Fletcher, G.F.; Mozaffarian, D.; Kris-Etherton, P.; Horn, L.V.; Lichtenstein, A.H.; Kumanyika, S.; Kraus, W.E.; Fleg, J.L.; Redeker, N.S.; et al. Interventions to Promote Physical Activity and Dietary Lifestyle Changes for Cardiovascular Risk Factor Reduction in Adults. *Circulation* **2010**, *122*, 406–441. [CrossRef] [PubMed]
45. Chow, C.K.; Ariyarathna, N.; Islam, S.M.S.; Thiagalingam, A.; Redfern, J. mHealth in Cardiovascular Health Care. *Hear. Lung Circ.* **2016**, *25*, 802–807. [CrossRef] [PubMed]
46. Hickey, K.T.; Hauser, N.R.; Valente, L.E.; Riga, T.C.; Frulla, A.P.; Creber, R.M.; Whang, W.; Garan, H.; Jia, H.; Sciacca, R.R.; et al. A single-center randomized, controlled trial investigating the efficacy of a mHealth ECG technology intervention to improve the detection of atrial fibrillation: The iHEART study protocol. *BMC Cardiovasc. Disord.* **2016**, *16*, 152. [CrossRef] [PubMed]
47. Varma, N.; Cygankiewicz, I.; Turakhia, M.P.; Heidbuchel, H.; Hu, Y.F.; Chen, L.Y.; Couderc, J.P.; Cronin, E.M.; Estep, J.D.; Grieten, L.; et al. 2021 ISHNE/HRS/EHRA/APHRS Expert Collaborative Statement on mHealth in Arrhythmia Management: Digital Medical Tools for Heart Rhythm Professionals: From the International Society for Holter and Noninvasive Electrocardiology/Heart Rhythm Society/European Heart Rhythm Association/Asia-Pacific Heart Rhythm Society. *Circ. Arrhythmia Electrophysiol.* **2021**, *14*, e009204. [CrossRef]
48. DeVore, A.D.; Wosik, J.; Hernandez, A.F. The Future of Wearables in Heart Failure Patients. *JACC Heart Fail.* **2019**, *7*, 922–932. [CrossRef]

49. Hershman, S.G.; Bot, B.M.; Shcherbina, A.; Doerr, M.; Moayedi, Y.; Pavlovic, A.; Waggott, D.; Cho, M.K.; Rosenberger, M.E.; Haskell, W.L.; et al. Physical activity, sleep and cardiovascular health data for 50,000 individuals from the MyHeart Counts Study. *Sci. Data* **2019**, *6*, 24. [CrossRef]
50. Oresko, J.J.; Jin, Z.; Cheng, J.; Huang, S.; Sun, Y.; Duschl, H.; Cheng, A.C. A Wearable Smartphone-Based Platform for Real-Time Cardiovascular Disease Detection Via Electrocardiogram Processing. *IEEE Trans. Inf. Technol. Biomed.* **2010**, *14*, 734–740. [CrossRef]
51. Rapin, M.; Braun, F.; Adler, A.; Wacker, J.; Frerichs, I.; Vogt, B.; Chetelat, O. Wearable Sensors for Frequency-Multiplexed EIT and Multilead ECG Data Acquisition. *IEEE Trans. Biomed. Eng.* **2019**, *66*, 810–820. [CrossRef]
52. Banaee, H.; Ahmed, M.; Loutfi, A. Data Mining for Wearable Sensors in Health Monitoring Systems: A Review of Recent Trends and Challenges. *Sensors* **2013**, *13*, 17472–17500. [CrossRef]
53. Miao, F.; Cheng, Y.; He, Y.; He, Q.; Li, Y. A Wearable Context-Aware ECG Monitoring System Integrated with Built-in Kinematic Sensors of the Smartphone. *Sensors* **2015**, *15*, 11465–11484. [CrossRef]
54. Poongodi, T.; Rathee, A.; Indrakumari, R.; Suresh, P. IoT Sensing Capabilities: Sensor Deployment and Node Discovery, Wearable Sensors, Wireless Body Area Network (WBAN), Data Acquisition. In *Principles of Internet of Things (IoT) Ecosystem: Insight Paradigm*; Intelligent Systems Reference Library; Peng, S.L., Pal, S., Huang, L., Eds.; Springer International Publishing: Geneva, Switzerland, 2020; Volume 174, pp. 127–151. [CrossRef]
55. Dubey, H.; Yang, J.; Constant, N.; Amiri, A.M.; Yang, Q.; Makodiya, K. Fog Data: Enhancing Telehealth Big Data Through Fog Computing. In Proceedings of the ASE BigData & SocialInformatics 2015, ASE BD & SI'15, Kaohsiung, Taiwan, 7–9 October 2015; pp. 1–6.
56. Melillo, P.; Orrico, A.; Scala, P.; Crispino, F.; Pecchia, L. Cloud-Based Smart Health Monitoring System for Automatic Cardiovascular and Fall Risk Assessment in Hypertensive Patients. *J. Med. Syst.* **2015**, *39*, 109. [CrossRef] [PubMed]
57. Mohammed, K.I.; Zaidan, A.A.; Zaidan, B.B.; Albahri, O.S.; Alsalem, M.A.; Albahri, A.S.; Hadi, A.; Hashim, M. Real-Time Remote-Health Monitoring Systems: A Review on Patients Prioritisation for Multiple-Chronic Diseases, Taxonomy Analysis, Concerns and Solution Procedure. *J. Med. Syst.* **2019**, *43*, 223. [CrossRef] [PubMed]
58. Ping, P.; Hermjakob, H.; Polson, J.S.; Benos, P.V.; Wang, W. Biomedical Informatics on the Cloud: A Treasure Hunt for Advancing Cardiovascular Medicine. *Circ. Res.* **2018**, *122*, 1290–1301. [CrossRef] [PubMed]
59. Mora, H.; Gil, D.; Terol, R.M.; Azorín, J.; Szymanski, J. An IoT-Based Computational Framework for Healthcare Monitoring in Mobile Environments. *Sensors* **2017**, *17*, 2302. [CrossRef] [PubMed]
60. Azimi, I.; Pahikkala, T.; Rahmani, A.M.; Niela-Vilén, H.; Axelin, A.; Liljeberg, P. Missing data resilient decision-making for healthcare IoT through personalization: A case study on maternal health. *Future Gener. Comput. Syst.* **2019**, *96*, 297–308. [CrossRef]
61. Gia, T.N.; Jiang, M.; Rahmani, A.M.; Westerlund, T.; Liljeberg, P.; Tenhunen, H. Fog Computing in Healthcare Internet of Things: A Case Study on ECG Feature Extraction. In Proceedings of the 2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing, Liverpool, UK, 26–28 October 2015; pp. 356–363. [CrossRef]
62. Kenneth, E.; Rajendra, A.; Kannathal, N.; Lim, C.M. Data Fusion of Multimodal Cardiovascular Signals. In Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, Shanghai, China, 1–4 September 2005; pp. 4689–4692. [CrossRef]
63. Kannathal, N.; Acharya, U.R.; Ng, E.; Krishnan, S.; Min, L.C.; Laxminarayan, S. Cardiac health diagnosis using data fusion of cardiovascular and haemodynamic signals. *Comput. Methods Programs Biomed.* **2006**, *82*, 87–96. [CrossRef]
64. King, R.C.; Villeneuve, E.; White, R.J.; Sherratt, R.S.; Holderbaum, W.; Harwin, W.S. Application of data fusion techniques and technologies for wearable health monitoring. *Med. Eng. Phys.* **2017**, *42*, 1–12. [CrossRef]
65. Van den Heuvel, E.R.; Vasan, R.S. Statistics in cardiovascular medicine: There is still gold in the old. *Heart* **2018**, *104*, 1227. [CrossRef]
66. Ang, L.M.; Seng, K.P. Big Sensor Data Applications in Urban Environments. *Big Data Res.* **2016**, *4*, 1–12. [CrossRef]
67. Dash, S.; Shakyawar, S.K.; Sharma, M.; Kaushik, S. Big data in healthcare: Management, analysis and future prospects. *J. Big Data* **2019**, *6*, 54. [CrossRef]
68. Rajabion, L.; Shaltooki, A.A.; Taghikhah, M.; Ghasemi, A.; Badfar, A. Healthcare big data processing mechanisms: The role of cloud computing. *Int. J. Inf. Manag.* **2019**, *49*, 271–289. [CrossRef]
69. Saheb, T.; Izadi, L. Paradigm of IoT big data analytics in the healthcare industry: A review of scientific literature and mapping of research trends. *Telemat. Inform.* **2019**, *41*, 70–85. [CrossRef]
70. PLUX Wireless Biosignals. Available online: https://plux.info/ (accessed on 10 September 2021).
71. Matias, I.; Garcia, N.; Pirbhulal, S.; Felizardo, V.; Pombo, N.; Zacarias, H.; Sousa, M.; Zdravevski, E. Prediction of Atrial Fibrillation using artificial intelligence on Electrocardiograms: A systematic review. *Comput. Sci. Rev.* **2021**, *39*, 100334. [CrossRef]
72. Kelders, S.M.; van Zyl, L.E.; Ludden, G.D.S. The Concept and Components of Engagement in Different Domains Applied to eHealth: A Systematic Scoping Review. *Front. Psychol.* **2020**, *11*, 926. [CrossRef]
73. Baumel, A.; Kane, J.M. Examining Predictors of Real-World User Engagement with Self-Guided eHealth Interventions: Analysis of Mobile Apps and Websites Using a Novel Dataset. *J Med. Internet Res* **2018**, *20*, e11491. [CrossRef] [PubMed]
74. Van Gemert-Pijnen, L.; Kelders, S.M.; Kip, H.; Sanderman, R. *eHealth Research, Theory and Development: A Multi-Disciplinary Approach*; Routledge: Abingdon, UK, 2018.

75. Baumel, A.; Yom-Tov, E. Predicting user adherence to behavioral eHealth interventions in the real world: Examining which aspects of intervention design matter most. *Transl. Behav. Med.* **2018**, *8*, 793–798. [CrossRef] [PubMed]
76. Gorst, S.L.; Armitage, C.J.; Brownsell, S.; Hawley, M.S. Home Telehealth Uptake and Continued Use Among Heart Failure and Chronic Obstructive Pulmonary Disease Patients: A Systematic Review. *Ann. Behav. Med.* **2014**, *48*, 323–336. [CrossRef]
77. Xu, K. Silicon electro-optic micro-modulator fabricated in standard CMOS technology as components for all silicon monolithic integrated optoelectronic systems. *J. Micromech. Microeng.* **2021**, *31*, 054001. [CrossRef]

MDPI AG
Grosspeteranlage 5
4052 Basel
Switzerland
Tel.: +41 61 683 77 34

Sensors Editorial Office
E-mail: sensors@mdpi.com
www.mdpi.com/journal/sensors

Academic Open
Access Publishing

mdpi.com

ISBN 978-3-7258-1562-3