Proceeding Paper

Classifier Module of Types of Movements Based on Signal Processing and Deep Learning Techniques †

Manuel Gil-Martín *, Javier López-Iniesta and Rubén San-Segundo

Speech Technology Group, Information Processing and Telecommunications Center,
E.T.S.I. Telecomunicación, Universidad Politécnica de Madrid;
javier.lopeziniesta.diazdelcampo@alumnos.upm.es (J.L.-I.); ruben.sansegundo@upm.es (R.S.-S.)
* Correspondence: manuel.gilmartin@upm.es; Tel.: +34-910672500
† Presented at 8th International Electronic Conference on Sensors and Applications, 1–15 November 2021; Available online: https://ecsa-8.sciforum.net.

Abstract: Human Activity Recognition (HAR) has been widely addressed by deep learning techniques. However, most prior research applied a general unique approach (signal processing and deep learning) to deal with different human activities including postures and gestures. These types of activity typically have highly diverse motion characteristics, which could be captured with wearable sensors placed on the user's body. Repetitive movements like running or cycling have repetitive patterns over time and generate harmonics in the frequency domain, while postures like sitting or lying are characterized for a fixed position with some positional changes and gestures or non-repetitive movements are based on an isolated movement usually performed by a limb. This work proposes a classifier module to perform an initial classification among these different types of movements, which would allow applying afterwards the most appropriate approach in terms of signal processing and deep learning techniques for each type of movement. This classifier is evaluated using PAMAP2 and OPPORTUNITY datasets using subject-wise cross-validation methodology. These datasets used inertial sensors on hands, arms, chest, hip, and ankles, which could collect data in a non-intrusive way. In the case of PAMAP2, the direct approach for classifying the 12 activities using 5-s windows in the frequency domain obtained an accuracy of 85.26 ± 0.25%. However, an initial classifier module could distinguish between repetitive movements and postures using 5-s windows reaching higher performances. Afterward, specific window size, signal format, and deep learning architecture were used for each type of movement module, obtaining a final accuracy of 90.09 ± 0.35% (an absolute improvement of 4.83%).

Keywords: Human Activity Recognition; wearable sensors; classifier module; inertial signals; convolutional neural networks; deep learning; repetitive movements; gestures; postures; PAMAP2; OPPORTUNITY

1. Introduction

Human motion modeling using wearable sensors is a studied field [1,2] with different applications such as Human Activity Recognition (HAR) [3–6], biometrics [7], or health [8]. Concerning physical activity classification, human movements are usually modeled by a system that could recognize these activities afterward. However, each movement presents specific characteristics in terms of motion pattern and duration. A previous work [4] proposed a human motion typology to apply the most appropriate signal processing and deep learning architecture depending on the type of movement. For example, raw signals of repetitive movements such as running, or cycling were processed by a Convolutional Neural Network (CNN) while raw signals of gestures like closing a drawer were processed by a Recurrent Neural Network (RNN). This work highlighted the requirement of developing an initial module to automatically identify the type of

movement before selecting the best signal processing and deep learning strategy to discriminate between movements inside the same group.

The purpose of this work is to develop this initial module as a classifier of types of movements. In addition, this work compares the direct approach of classifying all the activities in each dataset to the approach of using the initial classifier module to distinguish among the types of movements and, afterward, applying specific signal processing and deep learning techniques for movements inside the same group.

2. Materials and Methods

This section provides a description of the datasets, the signal processing, the deep learning approach, and the cross-validation methodology used in this study.

2.1. Datasets

For this work, we used the PAMAP2 dataset [9] and the OPPORTUNITY dataset [10]. The combination of these two datasets contains a wide variety of physical activities, including repetitive, non-repetitive movements (gestures), and postures. Moreover, PAMAP2 includes 27 signals recorded under laboratory conditions while OPPORTUNITY contains 113 signals recorded under wild conditions.

PAMAP2 dataset includes recordings of 12 different physical activities: nine repetitive movements (walking, running, cycling, Nordic walking, ascending stairs, descending stairs, vacuum cleaning, ironing, and rope jumping) and three postures (lying, sitting, standing). These activities were performed by nine subjects, who wore three Inertial Measurement Units (IMUs) with tri-axial accelerometer, gyroscope, and magnetometer. These sensors collected data sampling at 100 Hz and were placed on the dominant hand, chest, and ankle.

OPPORTUNITY dataset contains recordings of 21 different physical activities: one repetitive movement (walking), 17 gestures (open door 1, open door 2, close door 1, close door 2, open fridge, close fridge, open dishwasher, close dishwasher, open drawer 1, close drawer 1, open drawer 2, close drawer 2, open drawer 3, close drawer 3, clean table, drink from a cup, and toggle switch) and three postures (lying, sitting, and standing). These activities were performed by four subjects, who wore five RS485-networked XSense IMUs located in a jacket, two InertiaCube3 on each foot, and 12 Bluetooth acceleration sensors on the limbs. Each IMU includes a tri-axial accelerometer, a tri-axial gyroscope, and a tri-axial magnetic sensor sampling at 32 Hz.

2.2. Signal Processing

We implemented a signal processing module to generate windows of physical activity. We divided the recordings into overlapped windows using a Hamming function and a shift of 0.25 s between consecutive windows. We evaluated the classification performance at the window level, so if an activity transition occurred within a window, the system tried to recognize the activity with the greater presence in the window. We considered different window sizes (3 s, 5 s, 10 s, 15 s, 20 s, and 25 s) but maintaining the same shift (0.25 s). For each window size, we analyzed both time and frequency domain signals as inputs to our deep neural networks. In the first case (raw data), the time samples included in each window directly fed the deep neural network. In the second case, the inputs were the module of the Fast Fourier Transform (FFT) coefficients of each window. These coefficients represented the spectrum from 0 Hz to 50 Hz or 16 Hz (half of the sampling frequency). We decided to limit the spectrogram to 25 Hz for PAMAP2 because the energy in human activities mostly concentrates in low frequencies. Regarding OPPORTUNITY, we considered the frequency range from 0 to 16 Hz. Figure 1 represents the signal processing performed to the acceleration signals in both time and frequency domains for 5-s windows.
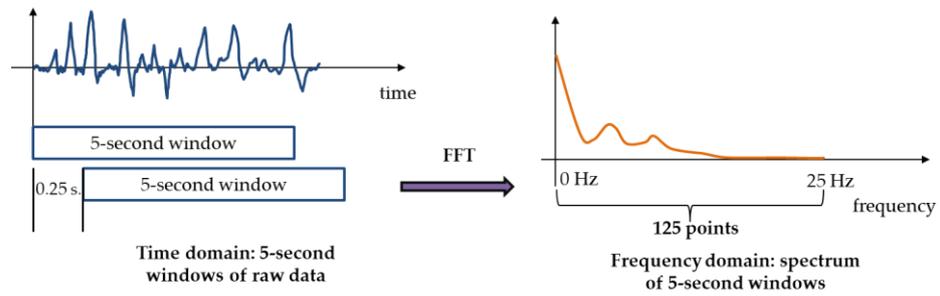
Figure 1. Signal processing performed on the inertial signals.

2.3. Deep Learning

A deep learning structure with a feature learning subnet and a classification subnet was used to distinguish between the different types of movements. This architecture was composed of two convolutional layers with intermediate max-pooling layers for feature learning and three fully connected layers for classification. Dropout layers were included after convolutional and fully connected layers to avoid overfitting. The architecture used for the PAMAP2 classifier module is represented in Figure 2.
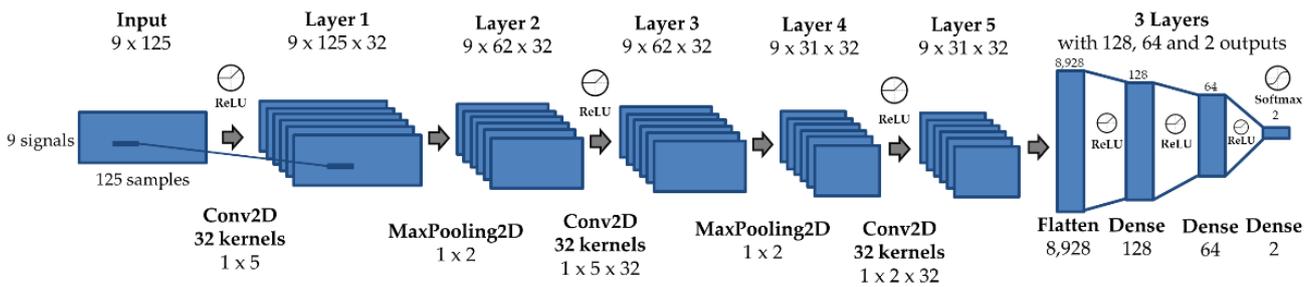

Figure 2. Deep learning architecture used in this work for the PAMAP2 classifier module.

In this architecture, ReLU was used as the activation function in intermediate layers to reduce the impact of gradient vanishing effect and SoftMax is the activation function in the last layer to perform the classification task. The optimizer was fixed to the root-mean-square propagation method [11] with a learning rate of 0.001. In this work, the following hyperparameters of the architecture were optimized for each dataset using a validation subset: number of epochs, batch size, number of convolutional layers, number and size of convolutional kernels, pooling kernel size, numbers of neurons in the fully connected layers and dropout fraction.

Regarding the classification of activities within each type of movement, this work applied and optimized the architectures of the previous work [4] that proposed the typology of types of movements: convolutional and fully-connected layers architecture for repetitive movements and postures and convolutional and recurrent layers architecture for gestures.

2.4. Cross-Validation

In this study, we used the Subject-Wise Cross-Validation (CV) strategy, where the recordings of each subject were included in separated subsets and the testing data were not seen during the training process. For both datasets, we performed 4 folds, using two folds to train the deep learning architecture, one-fold to adjust the main parameters of it (validation subset), and the remaining fold to test the system. This strategy followed a round-robin approach to evaluate the whole dataset and the results in this work are average values obtained throughout the CV procedure. This way, recordings from the same subject did not appear in training and testing subsets at the same experiment. In the case

of OPPORTUNITY, since there are only four subjects, the CV procedure becomes a Leave-One-Subject-Out CV strategy, where the validation and testing subsets only contain data from one subject in each iteration.

3. Results

Repetitive movements are characterized by the presence of harmonics in the frequency domain; in contrast, postures and gestures only have information at low frequencies. However, it is possible to perform a specific gesture like drinking a cup of coffee while a subject is walking or while sitting. For these reasons, an approach consisting of two steps could be applied when dealing with gestures. The first step of the classifier should distinguish movements in a higher level between two groups: repetitive movements and gestures during repetitive movements versus postures and gestures during postures. In this step, the FFT coefficients were able to distinguish between these two groups. During the second step, a specific module could distinguish between the types of movement within each previous module. For these subgroups, the FFT coefficients could be appropriate to distinguish the gestures while performing another repetitive movement or posture using the signals from the different limbs. Afterward, for each type of movement, we followed the configurations of the previous work [4] that proposed the typology of types of movements to boost the recognition performance: raw data and long windows (25 s) for repetitive movements, raw data and short windows (3 s) for gestures and FFT and long windows (10 s) for postures. Particularizing this approach to the datasets used in this work, Figures 3 and 4 show the overview of the final systems of PAMAP2 and OPPORTUNITY respectively, including the window size and input format for each module. These figures include the type of movement recognition performance in green boxes and the activity classification performance in orange boxes. In case of PAMAP2, since there were no gestures, an initial classifier was used to distinguish between repetitive movements and postures using FFT coefficients as inputs to the deep learning architectures. The performance of the repetitive movements module of OPPORTUNITY dataset was the maximum since there was only examples from walking activity.
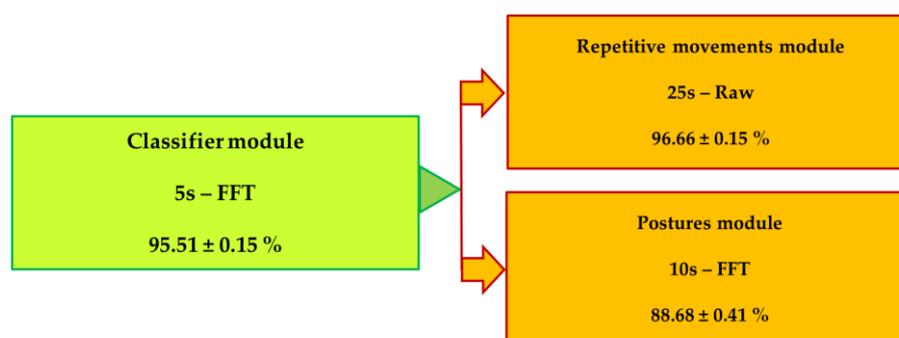
Figure 3. Overview of classification modules for PAMAP2 dataset.
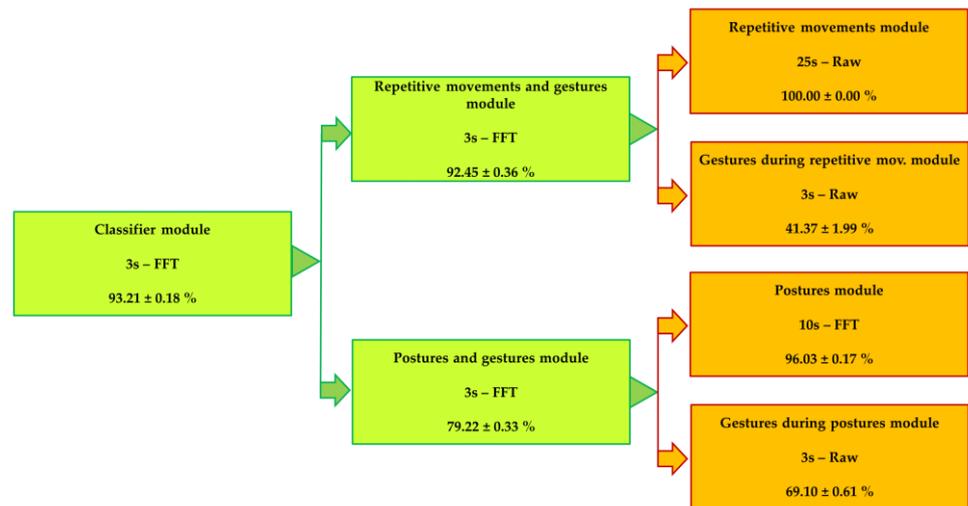
Figure 4. Overview of classification modules for OPPORTUNITY dataset.

Table 1 summarizes the main results of this paper for PAMAP2 and OPPORTUNITY datasets. Final activity classification accuracy and confidence intervals (CI) are included in the tables. Direct systems performances were obtained using short windows (3–5 s) and optimized signal processing and deep learning techniques for each dataset: FFT and convolutional and fully connected layers architecture for PAMAP2 and raw data and convolutional and fully connected layers architecture for OPPORTUNITY. Final activity accuracy of the system with classifier is obtained considering the performance of the modules and the number of examples per type of movement. For example, Equation (1) was applied for computing the final performance of the PAMAP2 dataset, where N_1 is the number of examples of repetitive movements, N_2 is the number of examples of postures and N is the total number of examples. Results suggest that it is possible to significantly increase the recognition performance when using specific modules to distinguish the movement through the movement typology.

$$Accuracy\ (\%) = \frac{Classifier\ module\ Acc.}{100} * \frac{(Rep.\ Acc.* N_1 + Post.Acc.* N_2)}{N} \quad (1)$$

Table 1. Activity classification accuracy for PAMAP2 and OPPORTUNITY datasets.

Experiment	Test Accuracy (%)	
	PAMAP2	**OPPORTUNITY**
Direct system	85.26 ± 0.25	67.33 ± 0.33
System with classifier	90.09 ± 0.35	68.45 ± 0.66

4. Discussion and Conclusions

The classifier module developed in this work allows recognizing types of movements to apply specific signal processing and deep learning techniques for each type of movement afterward. Results suggest that using short windows (3–5 s) to discriminate between different types of movements is required since higher decision resolution might skip examples of gestures. After that, it is recommended to increase the window size when classifying repetitive movements or postures because the recognition performance will improve. In this sense, the window size becomes a crucial optimization parameter that depends on the type of movement. For example, when an athlete is performing physical exercise during lengthy series, long analysis windows could be directly used to obtain high recognition performance. However, if a person is performing unknown activities, a lower decision resolution is required to detect the type of movement beforehand. In this

regard, a trade-off between decision resolution and performance is critical for leveraging the capabilities of HAR systems.

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

Institutional Review Board Statement:

Informed Consent Statement:

Data Availability Statement:

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

References

1. Wang, J.; Chen, Y.; Hao, S.; Peng, X.; Hu, L. Deep learning for sensor-based activity recognition: A survey. *Pattern Recognit. Lett.* **2019**, *119*, 3–11, doi:10.1016/j.patrec.2018.02.010.
2. Slim, S.O.; Atia, A.; M.A., M.; M.Mostafa, M.-S. Survey on Human Activity Recognition based on Acceleration Data. *Int. J. Adv. Comput. Sci. Appl.* **2019**, *10*, 84–98, doi:10.14569/ijacsa.2019.0100311.
3. Gil-Martín, M.; San-Segundo, R.; Fernández-Martínez, F.; Ferreiros-López, J. Improving physical activity recognition using a new deep learning architecture and post-processing techniques. *Eng. Appl. Artif. Intell.* **2020**, *92*, 103679, doi:10.1016/j.engappai.2020.103679.
4. Gil-Martín, M.; San-Segundo, R.; Fernández-Martínez, F.; de Córdoba, R. Human activity recognition adapted to the type of movement. *Comput. Electr. Eng.* **2020**, *88*, 106822, doi:10.1016/j.compeleceng.2020.106822.
5. Gil-Martín, M.; San-Segundo, R.; Fernández-Martínez, F.; Ferreiros-López, J. Time Analysis in Human Activity Recognition. *Neural Process. Lett.* **2021**, 1–19, doi:10.1007/s11063-021-10611-w.
6. Gil-Martin, M.; San-Segundo, R.; Lutfi, S.L.; Coucheiro-Limeres, A. Estimating gravity component from accelerometers. *IEEE Instrum. Meas. Mag.* **2019**, *22*, 48–53, doi:10.1109/mim.2019.8633352.
7. Gil-Martín, M.; San-Segundo, R.; de Córdoba, R.; Pardo, J.M. Robust Biometrics from Motion Wearable Sensors Using a D-vector Approach. *Neural Process. Lett.* **2020**, *52*, 2109–2125, doi:10.1007/s11063-020-10339-z.
8. Gil-Martín, M.; Montero, J.M.; San-Segundo, R. Parkinson's Disease Detection from Drawing Movements Using Convolutional Neural Networks. *Electronics* **2019**, *8*, 907, doi:10.3390/electronics8080907.
9. Reiss, A.; Stricker, D. Introducing a New Benchmarked Dataset for Activity Monitoring. In Proceedings of the 2012 16th International Symposium on Wearable Computers, Newcastle upon Tyne, UK, 18–22 June 2012; pp. 108–109, doi:10.1109/iswc.2012.13.
10. Chavarriaga, R.; Sagha, H.; Calatroni, A.; Digumarti, S.T.; Tröster, G.; Millan, J.D.R.; Roggen, D. The Opportunity challenge: A benchmark database for on-body sensor-based activity recognition. *Pattern Recognit. Lett.* **2013**, *34*, 2033–2042, doi:10.1016/j.patrec.2012.12.014.
11. Weiss, N.A. *Introductory statistics*; Pearson: 2017.