Automatic EMG-based Hand Gesture Recognition System using Time-Domain Descriptors and Fully-Connected Neural Networks

Ana Antonia Neacsu*, George Cioroiu*,†, Anamaria Radoi*, and Corneliu Burileanu*
*Center for Advanced Research on New Materials, Products and Innovative Processes, University Politehnica of Bucharest
† Research Institute for Artificial Intelligence "M. Draganescu", Romanian Academy Bucharest, Romania

***Abstract*—Hand gesture recognition has numerous applications in medical (e.g., prosthetics), engineering (e.g., robot manipulation) and, even, military research areas (e.g., UAV control applications). This paper proposes a fast and accurate method to identify hand gesture categories based on electromyographic (EMG) signals registered by a commercial sensor (e.g., Myo Armband developed by Ontario-based Thalmic Labs), which is placed on the user's forearm. The proposed method is based on the extraction of time-domain features and a neural network architecture to perform the classification of the EMG signals. In order to evaluate the performance of the proposed algorithm, we use a publicly available dataset with 7 hand gesture categories. The proposed hand gesture recognition system achieves a 99.78 % overall performance accuracy, which is comparable to that reported by applying other state-of-the-art methods, but is able to work in real-time conditions**

***Keywords*—hand gesture recognition, neural networks, EMG signals**

I. Introduction

A significant amount of research has been focused on human-computer interaction based on gestures, vision and voice. Hand gesture recognition provides an intelligent, natural and convenient way of human-computer interaction (HCI). Its main applications are sign language recognition (SLR) and gesture-based control [1]. Sign language recognition has the goal of interpreting signs automatically using a computer, in order to help deaf people communicate easier with the society. Although it is highly structured and based on an alphabet and symbols, it also serves as a good basic for the development of general gesture-based human-computer interaction.

Surface Electromyography (sEMG) is the electrical manifestation of the neuromuscular activity associated with the contracting muscle. This technology may be used by physically impaired persons to control rehabilitation and assistive devices. EMG is also used in many research domains, e.g. biomechanics, gesture-based control applications, neuromuscular physiology, sign language recognition, military applications, games and virtual reality [2]. Among the challenges that accompany a recognition system based on sEMG signals, signal acquisition plays an important role. This process is, in general, affected by noise. For this reason, apart from using adequate signal processing techniques to reduce the effect of noise, powerful recognition algorithms are required.

In the field of gesture classification, several approaches have been proposed, e.g., multistream Hidden Markov Models (HMM) have been combined with decision trees to determine 18 types of hand gestures based on data collected from accelerometer sensors and EMG data [1].

In [3], an integrated approach for the identification of daily hand movements with a view to control prosthetic members is proposed. The EMG signals were acquired using two electrodes attached on two specific muscles of the hand. Features are extracted in the frequency domain, while inserting a dimensionality reduction stage (based either on Principal Component Analysis (PCA) or RELIEF feature selection algorithm [4]) before the application of the classifier. Results have shown that the information carried by the Empirical Mode Decomposition (EMD) extracted features may further increase the classification accuracy [5].

Recently, neural networks have proved great potential in solving gesture classification tasks [6], [7]. In [7], a convolutional network (ConvNet) is augmented with transfer learning techniques to leverage inter-user data from the first data set, alleviating the burden imposed on a single individual to generate a vast quantity of training data for sEMG-based gesture recognition. Specifically, the transfer learning technique proposed in [7] divides the recognition problem into learning a general mapping between the muscle activity and the hand gestures (i.e., source task) and learning a specific mapping (i.e., target task). The proposed transfer learning scheme achieved good results, but the features used are computationally complex, involving time-frequency transformations, e.g. Continuous Wavelet Transform (CWT).

Another method for gesture recognition is mentioned in [8], where an EMG-based pattern recognition algorithm is proposed for classification of joint wrist angular position during flexion and extension movements from EMG signals. The algorithm considers features in both time and frequency domains. The pattern recognition stage uses a recurrent neural network (RNN) as classifier. Results show that shallow Neural Networks have better performance than architectures with numerous layers containing autoencoders.

This work was supported by the Ministry of Innovation and Research, UEFISCDI, project SPIA-VA, agreement 2SOL/2017, grant PN-III-P2-2.1-SOL-2016-02-0002 and project UAV Platforms, agreement 1SOL/2016, grant PN-III-P2-2.1-SOL-2016-01-0008.

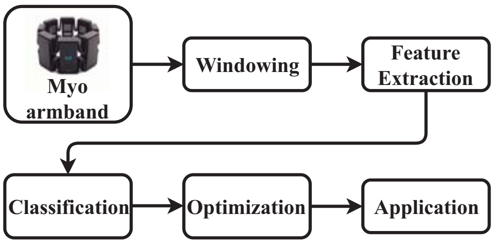

Fig. 1. General overview of the system

This paper proposes a real time automatic gesture recognition system, for seven basic gestures, based on sEMG signals. The signals are acquired using Myo armband, equipped with 8 circularly arranged EMG sensors, placed on the forearm. The classification is performed using a fully connected Neural Nerwork, whose training involves a free-source data set[1], also acquired with Myo armband, detailed in [7].

The rest of the paper is organized as follows: in Section II the proposed method is presented, including the features and the architecture used for the classification. Section III details the experimental setup and the corresponding results. The last section, IV, is dedicated to concluding remarks.

II. Proposed method

A general overview of the proposed method is shown in Figure 1. The signals captured by the armband are transmitted to a computer via Bluetooth. The sampling frequency of the device is $F_s = 200$ Hz, whilst for the analysis, we considered a rectangle sliding window of 250 ms (50 samples) length, with an overlap between consecutive windows of 200 ms (40 samples). The window size is chosen to allow lower statistical variance in the feature sets and a continuous classification of the EMG signals.

A. Feature Extraction

In general, the feature extraction module has an important influence over the recognition system. Numerous signal processing techniques and mathematical models are used to efficiently extract relevant information from the EMG data. So far, research and extensive efforts have been made in the signal processing area, which led to the development of faster and better algorithms. Since the purpose of this work is to design a fast and efficient algorithm for gesture recognition, the features used for classification must be simple and cost-effective to compute. Time-frequency analysis methods involving the Short-Time Fourier Transform, the wavelet transform or the wavelet packet transform would require a larger computation time, with no improvement over the classification performance [9]. For these reasons, only time-domain descriptors were considered. The features used in this paper are: *Mean absolute value (MAV)* (1), *Zero Crossing Rate (ZC)* (2), *Slope Sign Changes (SSC)* (3), *Waveform Length (WL)* (4), *Skewness* (5), *Root Mean Square (RMS)* (6), *Hjorth Activity* (7), *Integrated EMG* (8). Considering x to be the analysed signal of length L, the above mentioned descriptors are defined as follows:

[1]https://github.com/Giguelingueling/MyoArmbandDataset

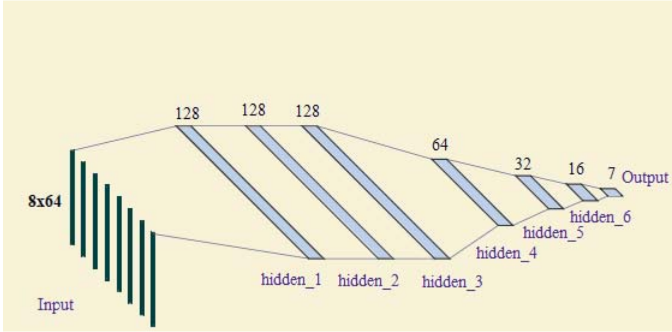

Fig. 2. Proposed Neural Network-based architecture

$$MAV(x) = \frac{1}{L}\sum_{k=0}^{L-1}|x_k| \quad (1)$$

$$ZC(x) = |\{k : (|x_k - x_{k-1}| \geq \alpha) \wedge (sgn(x_i) \neq sgn(x_{i-1}))\}| \quad (2)$$

$$SSC(x) = |k : (x_k - x_{k-1}) \cdot (x_k - x_{k+1}) \geq \alpha| \quad (3)$$

$$WL(x) = \sum_{k=1}^{L-1}|x_k - x_{k-1}| \quad (4)$$

$$Skewness(x) = \frac{1}{L}\sum_{k=0}^{L-1}\left(\frac{x_k - \overline{x}}{\sigma}\right)^3 \quad (5)$$

$$RMS(x) = \sqrt{\frac{1}{L}\sum_{k=0}^{L-1}x_k^2} \quad (6)$$

$$Activity(x) = \frac{1}{L}\sum_{k=0}^{L-1}(x_k - \overline{x})^2 \quad (7)$$

$$IEMG(x) = \sum_{k=0}^{L-1}|\,x_k\,|, \quad (8)$$

where sgn is the sign function, L is the length of the analysis window, x_k is the current vector, σ is the mean value of the window and $\overline{x}$ is the mean value of the vector.

B. Classification

Recently, the use of machine learning algorithms has become more prominent, as they are being employed in various tasks, ranging from simple regressions up to complex multinomial classification. In the field of EMG-based gesture recognition Neural Networks can be successfully used for classification, if the data set contains sufficient examples. Deep Network-based architectures can learn very complex patterns, but they are prone to overfitting. However, such architectures

Fig. 3. The seven hand gestures considered during the experiment [7]

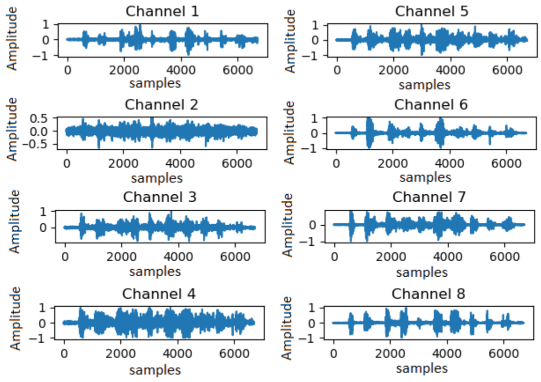

Fig. 4. Raw EMG signal captured with Myo

may be time-consuming, hence, not adequate for real time applications. This article proposes a fully connected architecture with a forward pass of less than 4.5 ms, including the feature extraction stage. The model parameters were determined using the cross entropy loss. Considering l targets, the cross entropy loss for a single example is given by the sum:

$$E = -\sum_{i=1}^{l} (t_i \log_2(y_i) + (1 - t_i) \log_2(1 - y_i)) \qquad (9)$$

where $t_1, t_2, ..., t_l$ are the targets and $y_1, y_2, ..., y_l$ are the outputs of the neural network arhitecture.

The model parameters are determined via backpropagation, using the ADAM optimizer instead of the classical Stochastic Gradient Descent (SGD). The reason for using the ADAM optimizer is its robustness to changes in hyperparameters [10].

The entire architecture diagram is displayed in Fig. 2. The proposed network is composed of 6 hidden layers, having 128, 128,128, 64, 32 and 16 neurons each. After each layer, a batch normalization step is performed to avoid overfitting. The activation used for all the layers is ReLu, except for the output layer which uses Softmax activation.

III. Experiments

A. Experimental Setup

We used a free-source data set, considering seven gestures that relate to the basic movements of the hand, namely four gestures for hand mobility (i.e., left / wrist flexion, right / wrist extension, down / ulnar deviation, up / radial deviation), two gestures for hand grip (hold / hand close, release / hand open) and one gesture for neutral position [7]. Examples of such gestures are displayed in Fig. 3. The EMG activity was recorded using the Myo armband created by Thalmic Labs. The Myo armband has 8 dry sEMG (surface EMG) sensors placed circularly (i.e., each EMG is composed of 8 channels). An example of a normalized 8-channels raw signal is depicted in Fig 4. The main advantage of the Myo armband is the fact that it is non-invasive and it can be used without any preparations. However, these benefits come with severe limitations since dry electrodes are not as accurate in capturing the EMG activity compared to gel-based ones. Even under these conditions, the proposed method is able to detect the movements of the hand in a very accurate manner.

The training data set consists of 18 able-bodied subjects, whilst the evaluation data set consists of 17 able-bodied subjects performing two rounds of 7 gestures. In both cases, the subjects were asked to perform the seven gestures in two rounds, for a period of 20 seconds.

As a side remark, the position of the armband is not important because, for every user, the armband was slid until the forearm circumference matched the armband. However, the orientation is important in order to keep the 8 channels in the same order for every user.

Moreover, we consider that if the user performs the same gesture for 20 seconds, then any continuous part of the signal is that gesture. Having this in mind, we partition the signals into 50 continuous samples, each equivalent to 250 milliseconds. Each sample receives the same label as the signal they originated from. For each sample, we extract eight features as detailed in Section II A.

For our neural network, we propose an architecture consisting on only Fully Connected Layers. We carried out several experiments with varying depth, from 4 to 7 hidden layers, but, in all cases, the number of neurons in a hidden layer was set to be bigger than the number of neurons in the next hidden layer. The architecture consisting of 6 hidden layers achieved the highest performance, whereas the number of neurons considered are 128, 128, 128, 64, 32, 16.

B. Experimental Results

In order to measure the performance of the proposed classification system, we compute the overall classification accuracy (OA), the per-class accuracies (PC) and Kappa index ($\mathcal{K}$). These measures are computed starting from the confusion matrix C which has the number of predicted labels on the columns and the ground truth on the rows:

$$OA = \sum_{i=1}^{l} \frac{C_{ii}}{N} \qquad (10)$$

$$PC_i = \frac{C_{ii}}{C_{i+}}, \quad \forall i \in \{1, ..., l\} \qquad (11)$$

$$\mathcal{K} = \frac{\frac{1}{N}\sum_{i=1}^{l} C_{ii} - \frac{1}{N^2}\sum_{i=1}^{l} C_{i+}C_{+i}}{1 - \frac{1}{N^2}\sum_{i=1}^{l} C_{i+}C_{+i}} \qquad (12)$$

TABLE I. PER-CLASS ACCURACY RATES (PC), OVERALL ACCURACY (OA) AND KAPPA INDEX ($\mathcal{K}$) OF PROPOSED METHOD FOR HAND GESTURE RECOGNITION.

Class	Proposed method	DL-TL [7]
Neutral	99.96	98.89
Radial Deviation	99.75	99.46
Wrist Flexion	99.86	98.42
Ulnar Deviation	99.72	96.52
Wrist Extension	99.72	99.55
Hand Close	99.86	99.43
Hand Open	99.58	94.61
OA	**99.78**	98.12
$\mathcal{K}$	**0.99**	0.98

TABLE II. AVERAGE RUNNING TIME PER GESTURE (DECOMPOSED ON FEATURE EXTRACTION AND PREDICTION STAGES).

Method	Feature extraction	Prediction	Total
Proposed	**1.9 ms**	**2.5 ms**	**4.4 ms**
DL-TL [7]	50.2 ms	19.4 ms	69.9 ms

where N represents the total number of analysed EMG signals, l is the number of gesture categories, C_{ij} is the number of signals in ground truth class j and classified as class i and the values C_{i+} and C_{+j} are computed as:

$$C_{i+} = \sum_{j=1}^{l} C_{ij} \tag{13}$$

$$C_{+j} = \sum_{i=1}^{l} C_{ij} \tag{14}$$

Moreover, the proposed method is compared to other state-of-the-art methods, such as the transfer learning-based method presented in [7] (abbreviated DL-TL). The results are synthesized in Table I.

Compared to the solution proposed in [7], the overall accuracy achieved by our recognition system (99.78%) is higher. Moreover, for all 7 hand gestures, the reported per-class accuracies are greater than the ones obtained by the DL-TL method [7].

An important aspect for real-time applications is the average running time, which, in the case of the proposed recognition technique, is 4.4 ms. Compared to the DL-TL method, this represents a speedup of 16 times. However, apart from the feature extraction and prediction stages, the most time-consuming part of the DL-TL method is the transfer learning stage which requires almost 5.25 minutes. This represents a drawback for real-time applications in the context of sEMG-based gesture recognition, since the convolutional network (ConvNet) scheme presented in [7] relies upon transfer learning techniques. These techniques leverage inter-user data and increase the overall accuracy by pre-training a model on multiple subjects before training it on a new participant, but come at the cost of spending additional time for learning a specific mapping.

The average running times for both methods, decomposed on feature extraction and prediction stages, are shown in Table II. We mention that all the experiments were carried on an NVIDIA Quadro M4000 GPU.

During our experiments, we observed that increasing the number of layers would result in a model that fails to generalize, thus, being prone to overfitting. The same phenomenon occurs when there are too many identical layers repeated in the model architecture.

C. Optimization and Application

The system for hand gesture recognition described in this paper is working fast because the computations needed for a forward pass are not time-consuming if compared to Convolutional Neural Networks (ConvNets). For real-time application, since the inference part of the system is done almost instantly, an optimization module can be added without creating a perceptible delay for the user.

IV. CONCLUSIONS

This paper presents a novel method for hand gesture recognition based on simple and easy to compute time-domain features. Using a relative easy to train neural network architecture, the proposed system is able to accurately recognize 7 basic hand gestures in a timely manner, being a candidate for online recognition of rapidly varying hand gestures.

REFERENCES

[1] X. Zhang, X. Chen, Y. Li, V. Lantz, K. Wang, and J. Yang, "A framework for hand gesture recognition based on accelerometer and emg sensors," *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, vol. 41, no. 6, pp. 1064–1076, 2011.

[2] J. Kim, S. Mastnik, and E. André, "Emg-based hand gesture recognition for realtime biosignal interfacing," in *Proceedings of the 13th international conference on Intelligent user interfaces*. ACM, 2008, pp. 30–39.

[3] C. Sapsanis, G. Georgoulas, and A. Tzes, "Emg based classification of basic hand movements based on time-frequency features," in *Control & Automation (MED), 2013 21st Mediterranean Conference on*. IEEE, 2013, pp. 716–722.

[4] R. J. Urbanowicz, M. Meeker, W. L. Cava, R. S. Olson, and J. H. Moore, "Relief-based feature selection: Introduction and review," *CoRR*, vol. abs/1711.08421, 2017.

[5] Y. Guo, Q. Wang, S. Huang, and A. Abraham, "Hand gesture recognition system using single-mixture source separation and flexible neural trees," *Journal of Vibration and Control*, vol. 20, no. 9, pp. 1333–1342, 2014.

[6] M. R. Ahsan, M. I. Ibrahimy, and O. O. Khalifa, "Electromygraphy (emg) signal based hand gesture recognition using artificial neural network (ann)," in *(ICOM), 2011 4th International Conference On Mechatronics*. IEEE, 2011, pp. 1–6.

[7] U. Côté-Allard, C. L. Fall, A. Drouin, A. Campeau-Lecours, C. Gosselin, K. Glette, F. Laviolette, and B. Gosselin, "Deep learning for electromyographic hand gesture signal classification using transfer learning," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2019.

[8] A. D. Orjuela-Cañón, A. F. Ruíz-Olaya, and L. Forero, "Deep neural network for emg signal classification of wrist position: Preliminary results," in *Computational Intelligence (LA-CCI), 2017 IEEE Latin American Conference on*. IEEE, 2017, pp. 1–5.

[9] K. Englehart and B. Hudgins, "A robust, real-time control scheme for multifunction myoelectric control," *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 7, pp. 848–854, July 2003.

[10] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *CoRR*, vol. abs/1412.6980, 2014.