

Real-Time Hand Gesture Recognition Based on Artificial Feed-Forward Neural Networks and EMG

Marco E. Benalcázar
Departamento de Informática y Ciencias
de la Computación
Escuela Politécnica Nacional
Quito, Ecuador
marco.benalcazar@epn.edu.ec

Carlos E. Anchundia
Departamento de Informática y Ciencias
de la Computación
Escuela Politécnica Nacional
Quito, Ecuador
carlos.anchundia@epn.edu.ec

Jonathan A. Zea
Departamento de Informática y Ciencias
de la Computación
Escuela Politécnica Nacional
Quito, Ecuador
jonathan.a.zea@ieee.org

Patricio Zambrano
Departamento de Informática y Ciencias
de la Computación
Escuela Politécnica Nacional
Quito, Ecuador
patricio.zambrano@epn.edu.ec

Andrés G. Jaramillo
Departamento de Informática y Ciencias
de la Computación
Escuela Politécnica Nacional
Quito, Ecuador
andres.jaramillo@epn.edu.ec

Marco Segura
Departamento de Informática y Ciencias
de la Computación
Escuela Politécnica Nacional
Quito, Ecuador
marco.segura@epn.edu.ec

Abstract—In this paper, we propose a real-time hand gesture recognition model. This model is based on both a shallow feed-forward neural network with 3 layers and an electromyography (EMG) of the forearm. The structure of the proposed model is composed of 5 modules: data acquisition using the commercial device Myo armband and a sliding window approach, pre-processing, automatic feature extraction, classification, and post-processing. The proposed model has an accuracy of 90.1% at recognizing 5 categories of gestures (fist, wave-in, wave-out, open, and pinch), and an average time response of 11 ms in a personal computer. The main contributions of this work include (1) a hand gesture recognition model that responds quickly and with relative good accuracy, (2) an automatic method for feature extraction from time series of varying length, and (3) the code and the dataset used for this work, which are made publicly available.

Keywords—hand gesture recognition, real-time, feed-forward neural networks, electromyography, feature extraction, time series

I. INTRODUCTION

The problem of real-time hand gesture recognition consists of identifying the class of a given movement of the hand as soon as the movement is produced. To be more precise, a hand gesture recognition model must respond in less than 300 ms after the movement ends to work in real time [1]. Solutions for this problem have multiple applications including the control of hand prostheses [2, 3] and robotics [4, 5]; human-computer interaction, including mouse control [6], gaming, sign language translation [7], and virtual reality [8]; and medical applications, including image manipulation [9].

The most common approach for hand gesture recognition is to use a sliding window to extract data for classification [10]. In this way, the result returned by a recognition model is a sequence of labels, where each label corresponds to each window observation. To facilitate the analysis, a hand gesture classification model can be split into the following 5 modules: data acquisition, pre-processing, feature extraction and feature selection, classification, and post-processing. The data acquisition module refers to the type of sensor that we use to acquire data from the hand. For developing hand gesture

recognition models, different types of sensors have been used such as gloves [11]; optical sensors, including webcams [11], infrared cameras [12], and lasers [13]; inertial measurement units (IMUs) [14]; and surface electromyography (EMG) [15-17]. Some models also combine signals from different types of sensors [18].

In this work, we focus on surface EMG sensors since they have several advantages compared with other types of sensors. Surface EMG sensors, or simply EMG sensors from now on, measure the electrical signals that the skeletal muscles produce when they contract. Compared with optical sensors, EMG sensors do not suffer from occlusion, changes of illumination, and changes of the distance between the hand and the sensor. Additionally, implementing portable models is much easier using EMG sensors than using optical sensors, which usually need to be placed in a fixed location. Another advantage of EMG sensors is that they allow the construction of armbands, which are usually more comfortable to wear than gloves. Compared with IMUs, EMG sensors return signals with less noise. In addition to this advantage, recognizing gestures of the hand using angular velocity and acceleration might result uncomfortable for some users since we have to place several IMUs on the fingers and the palm. Finally, EMG sensors are a good option for implementing gesture recognition models for upper-limb amputees, for whom the sensors described above are very difficult and, in some cases, impossible to use [19].

On the other hand, the use of EMG signals for hand gesture recognition brings some challenges. Most of the classification algorithms proposed in the scientific literature are designed for independent and identically distributed feature vectors (i.e., stationary process) [20, 21]. However, EMG signals behave as a non-stationary Gaussian process modulated by the envelope corresponding to a given movement (Fig. 1). The Gaussian process has zero mean and time-dependent variance [22]. The envelopes of EMG signals have inter- and intra-user variations of their shapes [23]. Therefore, developing general recognition models using EMG is a more difficult problem than developing user-specific models. A general model needs to be trained only once and works for any person, whereas a user-specific model

needs training for each person and for each time it is used. In addition to these problems, EMGs lie in high dimensional spaces, where the number of dimensions depends on the time of measurement and the sampling rate.

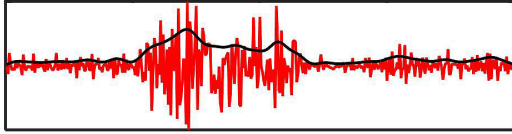


Fig. 1. EMG signal (red line) and its envelope (black line).

The preprocessing module consists of a set of techniques, which can be used to clean and condition the signals returned by the data acquisition module. The most common techniques used for pre-processing EMG include rectification and filtering. The main goal of these techniques is to reduce the noise and detect the envelope of EMG signals. The envelopes of EMGs can be used to determine which type of movement or gesture was performed [24].

The feature extraction module is responsible for extracting independent and non-redundant feature vectors, which need to be discriminative enough for the classification module. For hand gesture recognition, the most common domains for feature extraction include time, frequency, and the combination of these two (e.g., wavelets and spectrograms) [24]. For EMG classification, the features are usually defined manually, which requires deep knowledge of the biological processes underlying the generation of EMG signals during muscle contraction. Additionally, given an initial set of features, deciding which subset of these features is the best option for the classification module (i.e., feature selection) is not a trivial task since this problem is combinatorial [20].

The classification module consists of a function that maps feature vectors to labels. The most frequent classification algorithms used for hand gesture classification include support vector machines [25], feed-forward and recurrent neural networks [26], convolutional neural networks [16], decision trees [27, 28], k -nearest neighbors (k NN) [7], linear discriminant analysis [29], hidden Markov models, and ensembles of classifiers such as random forests [27, 28]. In this work, we focus on the use of feed-forward neural networks because this type of network is a universal approximator [21, 30]. This means that a feed-forward neural network with as many as three layers: input, hidden and output; with arbitrary squashing functions; and with enough number of neurons in the hidden layer is capable of approximating any Borel measurable function with a desirable degree of accuracy [30].

The post-processing module is in charge of refining the decision returned by the classification module. The most common operation used for post-processing is filtering the sequence of labels returned by the classifier. This operation eliminates spurious labels and produces a smooth response from the recognition model.

Developing a real-time gesture recognition model is a very challenging problem since we usually must find an adequate balance between model complexity and computational cost. Models with high complexity, measured through the Vapnik-

Chervonenkis (VC) dimension [23, 24], have usually high computational cost and can give high classification accuracy provided that we have enough training data. However, models with high computational cost are usually not good for developing real-time recognition models due to the long time of processing and large memory requirements. Additionally, training models with high complexity demands large training datasets to avoid overfitting. In the case of user-specific models, acquiring large training datasets is very difficult from the practical point of view since the models need training for each person and for each time they are used. On the other hand, using models with low complexity usually means low computational cost. However, models with low complexity applied to non-linear problems lead to low classification accuracy. Moreover, because EMG signals lie in a high dimensional space, we need an appropriate method for feature extraction in order to reduce the number of dimensions without losing key information for classification. Therefore, the problem of real-time hand gesture recognition is still open for new research which, among other things, must find feature extraction and classification methods with a good balance between model complexity and computational cost. Finally, in this work we propose a real-time hand gesture recognition model composed of 5 modules, which will be described in the next section.

Following this introduction, in section 2, we describe the sensor used in this work and each module of the proposed model. In section 3, we present, analyze, and compare the results obtained in this work. Finally, in section 4, we draw some conclusions and outline the future work.

II. MATERIALS AND PROPOSED MODEL

A. Materials: Myo Armband and Datasets

The *Myo armband* (Fig. 2) is a commercial low cost EMG sensor built by Thalmic Labs. This sensor can be worn on the forearm. This sensor returns a digital EMG signal composed of 8 channels measured at 200 Hz. Each channel corresponds to the data acquired by each of the 8 pods that are distributed uniformly along the armband. The EMG measured by this sensor is transmitted via Bluetooth to the computer.



Fig. 2. Myo armband and the gestures recognized in this paper.

In this work, we focus on the recognition of 5 gestures: fist (*fi*), wave-in (*wi*), wave-out (*wo*), open (*op*), and pinch (*pi*). The set of all gestures that our model does not recognize, including the rest position, is referred as no-gesture (*no*). The

reason why chose these 5 gestures is because this allows us to compare the performance of the proposed model with the performance of the proprietary system of the Myo armband. Additionally, we can also perform comparisons with the performance of other similar works proposed in the scientific literature. In Fig. 2, we show the 5 gestures that we recognize in this work. It is worth indicating that this figure shows each gesture at its final position.

For this work, we used data from 10 people. The *training dataset* $\mathcal{D}_{train} = \{(\mathbf{E}_1, Y_1), \dots, (\mathbf{E}_N, Y_N)\}$ of each user contains a total of $N = 30$ examples, where the matrix $\mathbf{E}_i \in [-1, 1]^{8 \times 400}$ corresponds to the i th EMG signal measured during 2 s. The categorical variable $Y_i \in \{no, fi, wi, wo, op, pi\}$ denotes the label for the signal \mathbf{E}_i , with $i = 1, 2, \dots, N$. These 30 examples include 5 instances for each of the 6 classes defined previously. During the 2 s of measurement, users were asked to perform an instance of the 5 classes of $\{fi, wi, wo, op, pi\}$. All the instances of these 5 categories went through the following continuous sequence of 3 positions: rest, final position, and rest. For acquiring data for the class no-gesture, users were asked to have the hand in the rest position.

The *testing dataset* $\mathcal{D}_{test} = \{(\mathbf{G}_1, Y_1), \dots, (\mathbf{G}_M, Y_M)\}$ of each user is formed by $M = 150$ examples, recorded during 5 s each. \mathcal{D}_{test} includes 30 instances of each class of $\{fi, wi, wo, op, pi\}$. For testing, we do not include the class no-gesture because the goal is to recognize the 5 classes defined previously from any other movement (no-gesture). Each matrix \mathbf{G}_i from the testing set is an element of the space $[-1, 1]^{1000 \times 8}$, with $i = 1, 2, \dots, M$. We recorded the EMGs for testing longer than for training to position the EMG segment corresponding to the gesture at any region within the 5 s of measurement of the testing signals.

B. Proposed Model

The proposed model is composed of two stages: training and testing. In the training stage, we develop and train the proposed model. In the testing stage, we estimate the actual recognition accuracy of the proposed model.

a) Data Acquisition: For data acquisition, we define a window $W = (w_1, \dots, w_m)$, where w_i is a non-negative integer, with $i = 1, 2, \dots, m$. For this work, we used a window W of $m = 500$ points. The stride between two consecutive window observations is 10 points. We treated each observation as time series represented by the matrix $\mathbf{E} = [\mathbf{E}_1; \dots; \mathbf{E}_8] \in [-1, 1]^{m \times 8}$, where the column vector $\mathbf{E}_i = [E_{i1}, \dots, E_{im}]^T$ corresponds to the i th channel of the observation \mathbf{E} , with $i = 1, 2, \dots, 8$.

b) Pre-processing: Here, we first rectify the signal \mathbf{E} obtaining thus the new signal $\mathbf{R} = abs(\mathbf{E}) \in [0, 1]^{m \times 8}$, where the absolute value is applied to each element of \mathbf{E} . Second, we extract the envelopes \mathbf{V} of each channel of the signal \mathbf{R} . For this task, we apply to each channel of the signal \mathbf{R} a low-pass Butterworth filter Ψ of 5th order, with a cutoff frequency of 10 Hz, obtaining $\mathbf{V} = \Psi(\mathbf{R}) \in [0, 1]^{m \times 8}$. Third, we segment in

\mathbf{V} the region corresponding to a muscle contraction [31]. For this task, we compute the sum $\mathbf{S} = sum(\mathbf{V}, 2) \in [0, 8]^{m \times 1}$ along the rows of the matrix \mathbf{V} . Then, we compute the spectrogram $\mathbf{P}_C = \mathcal{S}(\mathbf{S}) \in \mathbb{C}^{26 \times p}$ of \mathbf{S} by dividing the frequency interval $[0, 100]$ Hz into 26 points. For the time division, we use a Hamming window of 25 points, with a stride of 15 points between two consecutive window observations. The value of p can be computed as $p = floor((m - 10)/15)$, where *floor* rounds toward $-\infty$ and m is the length of the signals \mathbf{E} , \mathbf{R} and \mathbf{V} . Next, we compute the modulus of each complex element of \mathbf{P}_C obtaining thus the new matrix $\mathbf{P} \in \mathbb{R}^{26 \times p}$. Next, we sum along the columns of \mathbf{P} obtaining the vector $\mathbf{U} = sum(\mathbf{P})$. Finally, we find the 2 indices i_s and i_e of \mathbf{U} between which its values are equal to or greater than a given threshold τ_u . For this work, we used the threshold $\tau_u = 10$. If the difference $(i_e - i_s)$ is less than a points, then the segmentation returns the signal $\mathbf{Z} = \mathbf{V}$. Otherwise, the segmentation returns the rows of \mathbf{V} between the indices i_s and i_e : $\mathbf{Z} = \mathbf{V}(i_s:i_e, :)$. For this work, we used $a = 100$ points. Note that this segmentation causes that the \mathbf{Z} signals have different lengths. Finally, if $i_s = 1$ and $i_e = m$, then we execute the next two steps: feature extraction and classification. Otherwise, we return the label no-gesture and proceed to the post-processing module directly. This previous classification is done because, if a window observation \mathbf{E} does not contain the complete EMG of a gesture, we want to reduce the chances of the next two modules returning an incorrect label for \mathbf{E} .

c) Feature Extraction: In this module, we automatically extract a feature vector $\mathbf{X} = (X_1, \dots, X_r) \in \mathbb{R}^r$ for the signal \mathbf{Z} , with $r = 6$ for this work. For this task, we first pre-process each example of \mathcal{D}_{train} using the steps described in the previous module. Then, for each $Y \in \{no, fi, wi, wo, op, pi\}$, we obtain the set ζ_Y of pre-processed EMGs from \mathcal{D}_{train} that belong to the class Y . Next, we find the signal $\mathbf{H}_i^* \in \zeta_Y$ that is closest to all the elements of ζ_Y , with $i = 1, 2, \dots, r$. For this task, we use the Dynamic Time Warping (DTW) distance for multi-channel time series [32]. Thus, we obtain the tuple $(\mathbf{H}_1^*, \dots, \mathbf{H}_r^*)$ of pre-processed EMGs that will be used for feature extraction. Note that the computations described so far in this module need to be performed only once for the training dataset \mathcal{D}_{train} of each user. Next, we compute the feature vector for the signal \mathbf{Z} through $\mathbf{X} = (dtw(\mathbf{H}_1^*, \mathbf{Z}), \dots, dtw(\mathbf{H}_r^*, \mathbf{Z}))$, where *dtw* denotes the DTW distance for multi-channel time series.

d) Classification: In this module, we estimate the class $\psi(\mathbf{X}) \in \{no, fi, wi, wo, op, pi\}$ to which the feature vector \mathbf{X} belongs. For this purpose, we use the equation

$$\psi(\mathbf{X}) = \underset{y \in \{no, fi, wi, wo, op, pi\}}{\operatorname{argmax}} \mathbb{P}(Y = y | \mathbf{X}), \quad (1)$$

subject to the constraint that the conditional probability that maximizes (1) is equal to or greater than τ . Otherwise, \mathbf{X} is labeled with no-gesture. In (1), the term $\mathbb{P}(Y = y|\mathbf{X})$ denotes the conditional probability that \mathbf{X} belongs to the class y . For this work, we set the threshold $\tau = 0.5$ to reduce the number of false positives for each class. For estimating the distribution $\mathbb{P}(Y|\mathbf{X})$, we use a shallow feed-forward Neural Network (NN) composed of three layers: input, hidden and output. The input layer is composed of 6 units, where each unit takes each component of the feature vector \mathbf{X} . The hidden layer is composed of 6 neurons with hyperbolic tangent functions. The output layer is also composed of 6 neurons with the softmax transfer function. For training the NN, we use the cross-entropy cost function and the gradient descend method [20, 21], with the training set $\mathcal{D} = \{(\mathbf{X}_1, Y_1), \dots, (\mathbf{X}_N, Y_N)\}$. In this set, each vector \mathbf{X}_i is obtained from the signal $\mathbf{E}_i \in \mathcal{D}_{train}$ by applying the pre-processing and the feature extraction modules defined above, with $i = 1, 2, \dots, N$. Before feeding a feature vector \mathbf{X} to the NN, we standardize its values by $(X_j - \mu)/\sigma$, where μ and σ denote the mean and the standard deviation of \mathbf{X} , respectively, with $j = 1, 2, \dots, r$.

e) Post-processing: Here we eliminate consecutive repetitions of the same label by using a time delay of one window observation. Let us assume we have the sequence of labels $\psi(\mathbf{X})_{i-1}, \psi(\mathbf{X})_i$ for some $i \in \mathbb{Z}^+$. If the current label $\psi(\mathbf{X})_i$ is equal to the previous label $\psi(\mathbf{X})_{i-1}$, then we return $\psi(\mathbf{X})_i = no$. Otherwise, we return the label $\psi(\mathbf{X})_i$ unchanged.

III. RESULTS, ANALYSIS, AND COMPARISONS

In this section, we present, analyze, and compare the results obtained by applying the proposed model to each testing set \mathcal{D}_{test} of the 10 volunteers of this work. The testing dataset of each person is composed of 30 examples of each category of the set $\{fi, wi, wo, op, pi\}$. Therefore, for each person, we have a total of 150 examples. The recognition accuracy that we report in this paper is computed over the 1500 examples that we have in total for testing. The time of processing that we report corresponds to the average time of labeling a window observation using a personal computer with an Intel® Core™ i7-3770S processor and 4GB of RAM memory.

For reproducing the results of the recognition accuracy of the proposed model, we make publicly available in the link: https://drive.google.com/drive/folders/1ZCsaHNc08MYvOS1fMC_wchioix6srpB the Matlab code as well as the datasets used for training and testing the proposed model. Reproducing the time of processing will be a bit difficult because this variable depends on the computational resources used to run the code.

In Table 1, we present the confusion matrix of the proposed model. For comparison purposes, in Table 2 we present the results that we obtained in this paper as well as in our previous works [31] and [33]. In [31] we presented a model using the

k NN classifier based on the DTW distance. In [33] we presented an evolution of the model presented in [31], by incorporating the segmentation of the muscle contraction on the training dataset only. In addition to these results, we also present the results that we obtained by evaluating the recognition accuracy of the proprietary model of the Myo.

TABLE I. CONFUSION MATRIX OF THE PROPOSED MODEL

		Targets					% PRECISION % ERROR
		FIST (fi)	WAVE-IN (wi)	WAVE-OUT (wo)	OPEN (op)	PINCH (pi)	
Predictions	NO-GESTURE (no)	16 1.1%	21 1.4%	8 0.5%	14 0.9%	26 1.7%	0.0% 100%
	FIST (fi)	282 18.8%	3 0.2%	0 0.0%	3 0.2%	1 0.1%	97.6% 2.4%
	WAVE-IN (wi)	2 0.1%	276 18.4%	9 0.6%	2 0.1%	16 1.1%	90.5% 9.5%
	WAVE-OUT (wo)	0 0.0%	0 0.0%	278 18.5%	10 0.7%	0 0.0%	96.5% 3.5%
	OPEN (op)	0 0.0%	0 0.0%	5 0.3%	271 18.1%	12 0.8%	94.1% 5.9%
	PINCH (pi)	0 0.0%	0 0.0%	0 0.0%	0 0.0%	245 16.3%	100% 0.0%
	%SENSITIVITY	94.0%	92.0%	92.7%	90.3%	81.7%	90.1%
	%ERROR	6.0%	8.0%	7.3%	9.7%	18.3%	9.9%

TABLE II. RESULTS OF DIFFERENT RECOGNITION MODELS

Model	Accuracy (%)	Time (ms)
Proposed	90.1	11.0
k NN + DTW distance + segmentation [31]	89.5	193.1
k NN + DTW distance [33]	86.0	245.5
Myo	83.1	-----

The results presented in Table 1 show that our model have a recognition accuracy of 90%. The highest and the lowest sensitivities occur for the gestures fist (94.0%) and pinch (81.7%), respectively. The highest and the lowest precisions occur for the gestures pinch (100%) and wave-in (90.5%), respectively. The results of this table also evidence that our model fails the most at discriminating between the classes pinch and wave-in followed by the classes pinch and open. Additionally, the values of the row corresponding to the class no-gesture evidence that some gestures are not recognized at all. This problem might be caused because, for some EMGs, our model predicts a sequence of labels belonging to two or more classes from the set $\{fi, wi, wo, op, pi\}$.

The results of Table 2 evidence that the recognition accuracy of the proposed model is slightly higher than the accuracy of the model presented in [31] and higher than the accuracies of both [33] and the proprietary system of the Myo. Regarding the time of processing, the proposed model is 17 times much faster than the real-time models used for the comparison. This occurs because these two previous models use the k NN classifier, whose time complexity of classification depends on the number of training examples. In contrast to this fact, the time complexity of classification model using a NN is constant with the number of training examples.

IV. CONCLUSIONS

In this paper, we have made three main contributions. (1) We have proposed a real-time hand gesture recognition model that responds quickly (i.e., in 11 ms) as well as with good accuracy (90.1%). (2) We have also proposed a method for

automatic feature extraction of time series of varying length. Finally, (3) we have made both the code and the dataset that we used for this paper publicly available. This last contribution will allow the readers to reproduce the recognition accuracy obtained in this paper and check the fine details of our implementation.

Future work includes the research of new algorithms to find automatically the centers of the clusters of each category for the feature extraction module. Another problem to research about is the use of recurrent neural networks for the classification module in order to improve the recognition accuracy using context information.

ACKNOWLEDGMENT

The authors gratefully acknowledge the financial support provided by Escuela Politécnica Nacional for the development of the research project PIJ-16-13.

REFERENCES

- [1] H. Mizuno, N. Tsujiuchi, and T. Koizumi, "Forearm motion discrimination technique using real-time emg signals," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*. IEEE, 2011, pp. 4435-4438.
- [2] N. Wang, K. Lao, and X. Zhang, "Design and Myoelectric Control of an Anthropomorphic Prosthetic Hand," *Journal of Bionic Engineering*, vol. 14, pp. 47-59, 2017.
- [3] M. Tavakoli, C. Benussi, and J.L. Lourenco, "Single channel surface EMG control of advanced prosthetic hands: A simple, low cost and efficient approach," *Expert Systems with Applications*, vol. 79, pp. 322-332, 2017.
- [4] J. Mi, Y. Sun, Y. Wang, Z. Deng, L. Li, J. Zhang, and G. Xie, "Gesture recognition based teleoperation framework of robotic fish," in *Proceedings of the 2016 IEEE International Conference on Robotics and Biomimetics (ROBIO)*. IEEE, 2016, pp. 137-142.
- [5] H. Liu, and L. Wang, "Gesture recognition for human-robot collaboration: A review," *International Journal of Industrial Ergonomics*, vol. 60, pp. 1-13, 2017.
- [6] H.S. Grif, and C.C. Farcas, "Mouse Cursor Control System Based on Hand Gesture," *Procedia Technology*, vol. 22, pp. 657-661, 2016.
- [7] L. Jiménez, M. Benalcázar, and N. Sotomayor, "Gesture Recognition and Machine Learning Applied to Sign Language Translation," in *IFMBE Proceedings book series*, vol. 1, Torres, J. Bustamante, and D.A. Sierra, Eds. Springer Singapore, 2016, pp. 233-236.
- [8] S.S. Rautaray and A. Agrawal, "Manipulating Objects through Hand Gesture Recognition in Virtual Environment," in *Advances in Parallel Distributed Computing*, D. Nagamalai, E. Renault, and M. Dhanuskodi, Eds. Springer Berlin Heidelberg, 2011, pp. 270-281.
- [9] R. Wipfli, V. Dubois-Ferrière, S. Budry, P. Hoffmeyer, and C. Lovis, "Gesture-Controlled Image Management for Operating Room: A Randomized Crossover Study to Compare Interaction Using Gestures, Mouse, and Third Person Relaying," *PLoS One*, vol. 11, pp. 1-10, 2016.
- [10] P.K. Pisharady and M. Saerbeck, "Recent methods and databases in vision-based hand gesture recognition: A review," *Computer Vision and Image Understanding*, vol. 141, pp. 152-165, 2015.
- [11] L. Lamberti and F. Camastra, "Handy: A real-time three color glove-based gesture recognizer with learning vector quantization," *Expert Systems with Applications*, vol. 39, pp. 10489-10494, 2012.
- [12] W. Lu, Z. Tong, and J. Chu, "Dynamic Hand Gesture Recognition With Leap Motion Controller," in *IEEE Signal Processing Letters*, vol. 23, no. 9, pp. 1188-1192, 2016.
- [13] P.R. Bhamre, R.G. Chauhan, and S. Kumar, "Gesture recognition using Laser Sensor Enhanced with different parameters," in *2014 IEEE Global Conference on Wireless Computing & Networking (GCWCN)*, Lonavala, 2014, pp. 79-83.
- [14] A. Moschetti, L. Fiorini, D. Esposito, P. Dario, and F. Cavallo, "Recognition of Daily Gestures with Wearable Inertial Rings and Bracelets," *Sensors*, vol. 16, no. 8, pp. 2-18, 2016.
- [15] Y. Huang, W. Guo, J. Liu, J. He, H. Xia, X. Sheng, H. Wang, X. Feng, and P.B. Shull, "Preliminary Testing of a Hand Gesture Recognition Wristband Based on EMG and Inertial Sensor Fusion," in *Intelligent Robotics and Applications*, H. Liu, N. Kubota, X. Zhu, R. Dillmann, and D. Zhou, Eds. Springer International Publishing, 2015, pp. 359-367.
- [16] W. Geng, Y. Du, W. Jin, W. Wei, Y. Hu, and J. Li, "Gesture recognition by instantaneous surface EMG images," *Nature Scientific Reports*, vol. 6, pp. 1-8, 2016.
- [17] X. Liu, M. Zhang, A. Richardson, T. Lucas, and J. Van Der Spiegel, "The Virtual Trackpad: an Electromyography-based, Wireless, Real-time, Low-Power, Embedded Hand Gesture Recognition System using an Event-driven Artificial Neural Network," *IEEE Trans. Circuits Syst. II Express Briefs*, vol. 64, no. 11, pp. 1257-1261, 2016.
- [18] V. Pathak, S. Mongia, and G. Chitranshi, "A framework for hand gesture recognition based on fusion of Flex, Contact and accelerometer sensor," *2015 Third International Conference on Image Information Processing (ICIIP)*, Wagnaghat, 2015, pp. 312-319.
- [19] D. Farina, N. Jiang, H. Rehbaum, A. Holobar, B. Graimann, H. Dietl, and O.C. Aszmann, "The Extraction of Neural Information from the Surface EMG for the Control of Upper-Limb Prostheses: Emerging Avenues and Challenges," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 22, no. 4, pp. 797-809, 2014.
- [20] R. Duda, D. Stork, and P. Hart, "Pattern Classification," Wiley 2001.
- [21] L. Devroye, L. Györfi, and G. Lugosi, "A Probabilistic Theory of Pattern Recognition," Springer, 1996.
- [22] K.C. McGill, "Surface electromyogram signal modelling," *Medical and Biological Engineering and Computing*, vol. 42, pp. 446-454, 2004.
- [23] J. Rodriguez, A. Malanda, and J. Navallas, "EMG modeling," INTECH Open Access Publisher, 2012.
- [24] M. Reaz, M. Hussain, and F. Mohd-Yasin, "Techniques of EMG signal analysis: detection, processing, classification and applications," *Biological Procedures Online*, vol. 8, no. 1, pp. 11-35, 2006.
- [25] S. Saha, A. Konar, and J. Roy, "Single Person Hand Gesture Recognition Using Support Vector Machine," in *Computational Advancement in Communication Circuits and Systems*, K. Maharatna, G.K. Dalapati, P.K. Banerjee, A.K. Mallick, and M. Mukherjee, Eds. Springer, 2015, pp. 161-167.
- [26] E. Tsironi, P. Barros, C. Weber, and S. Wermter, "An analysis of Convolutional Long Short-Term Memory Recurrent Neural Networks for gesture recognition," *Neurocomputing*, 2017.
- [27] M.K. Sohn, S.H. Lee, H. Kim, and H. Park, "Enhanced hand part classification from a single depth image using random decision forests," in *IET Computer Vision*, vol. 10, no. 8, pp. 861-867, 2016.
- [28] A. Joshi, C. Monnier, M. Betke, and S. Sclaroff, "Comparing random forest approaches to segmenting and classifying gestures," *Image and Vision Computing*, vol. 58, pp. 86-95, 2017.
- [29] Y. Wang and L. Zhang, "3D hand gesture recognition based on Polar Rotation Feature and Linear Discriminant Analysis," in *2013 Fourth International Conference on Intelligent Control and Information Processing (ICICIP)*, Beijing, 2013, pp. 215-219.
- [30] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, no. 5, pp. 359-366, 1989.
- [31] M. Benalcázar, C. Motoche, J. Zea, A. Jaramillo, C. Anchundia, P. Zambrano, M. Segura, F. Benalcázar, M. Pérez, "Real-time hand gesture recognition using the Myo armband and muscle activity detection," *2017 IEEE Second Ecuador Technical Chapters Meeting (ETCM)*, Salinas, 2017, pp. 1-6.
- [32] M. Müller, "Information Retrieval for Music and Motion," Springer, 2007.
- [33] M. Benalcázar, A. Jaramillo, J. Zea, A. Páez, and V. Andaluz, "Hand gesture recognition using machine learning and the Myo armband," *2017 25th European Signal Processing Conference (EUSIPCO)*, Kos, 2017, pp. 1040-1044.