Contemporary Engineering Sciences, Vol. 10, 2017, no. 28, 1377 - 1390
HIKARI Ltd, www.m-hikari.com
https://doi.org/10.12988/ces.2017.711162

CNN Architecture for Robotic Arm Control in a 3D Virtual Environment by Means of by Means of EMG Signals

Natalie Segura Velandia, Robinson Jiménez Moreno and Ruben Dario Hernández

Davinci Research Group, mechatronics department
Faculty of Engineering, Universidad Militar Nueva Granada, Carrera 11# 101-80
Bogotá, Colombia

Abstract

This paper presents the development of a 3D virtual environment to validate the effectiveness of a Convolutional Neural Network (CNN) in a virtual application, controlling the movements of a manipulator or robotic arm through commands recognized by the network. The architecture of the CNN network was designed to recognize five (5) gestures by means of electromyography signals (EMGs) captured by surface electrodes located on the forearm and processed by the Wavelet Packet Transform (WPT). In addition to this, the environment consists of a manipulator of 3 degrees of freedom with a final effector type clamp and three objects to move from one place to another. Finally, the network reaches a degree of accuracy of 97.17% and the tests that were performed reached an average accuracy of 98.95%.

Keywords: deep convolutional neuronal network, electromyography signal, wavelet packet transform, gesture recognition, robotic arm

1 Introduction

Undoubtedly, telecommunications have started a new technological era, making what was believed fiction to be true in the fields of science and engineering, establishing the first communications in 1840 by means of the telegraph handled under the Morse code [1]. Between 1877 and 1883 installing the first telephone lines in Boston, New Haven and then throughout the USA. Additionally, Tomas Alva Edison discovers the lamp and Nikola Tesla builds an alternative power system to replace generators and direct current motors, giving rise to what is known as modern electronics [2]. In 1935, the first studies on radar with a pulsed system at 12 MHz are started, achieving a range of 40 miles, sound broadcasting and transmission of images of acceptable quality.

In the 60s, the era of computers began and a date with great importance appeared in 1969, with the birth of the Internet, which would be commercially public in the 1990s. In 1998, there are developed the optical network systems capable of transmitting 3.2 Terabits per second [2]. According to the above, it is clear to understand the importance of electronic communication, wireless networks and other areas, which have been able to allow areas such as robotics, to create a technological integration.

From the foregoing, the beginnings are given to what is known as: Tele-operation, responsible for making an extension of sensory capabilities and human skills to a remote location. Finally, there is the Tele-robotics, considered as an evolution of Tele-Operation, for presenting a greater degree of autonomy in terms of decisions and actions within a remote system [3].

Consequently, Tele-robotics is driven by telecommunications through wireless connections, in order to control a machine without using cables, create virtual environments or operating interfaces, integrating machine vision systems to visualize the work environment, eliminating peripherals within this area.

In relation to the above, tele-robotic systems focused on medicine have been developed. In 1994, an automatic endoscopic system for voice-controlled positioning (AESOP) is presented, where its objective is to collaborate with the surgeon since they need more than one hand to maneuver the endoscope, this consists of a robotic arm controlled by voice capable of saving three anatomical positions avoiding any type of collision [4]. Likewise, in 2000, one of the most complete surgical robots was developed, consisting of a visualization trolley, dual lighting equipment, dual three-chip cameras, surgeon's console, three instrument arms and a camera arm. The console consists of two controls capable of controlling the robotic arms of 7 degrees of freedom and a machine vision system in 3d, with an infrared sensor that detects the movement of the head by activating the controls for the arms, which is known as Da Vinci Surgical System [5].

In space exploration, in 1997, there are found Sojouner and Rocky 7, built to perform exploration on Mars and known as Mars rovers. They are the first tele-operated vehicles capable of navigating in unknown terrains outside the planet earth with an information delay of 20 min taking into account the distances to which the control system is located [6].

However, tele-robotic systems are accompanied by environments capable of emulating reality. Virtual environments are, in the first instance, low-cost tools capable of allowing the simulation of a system by providing a form of learning or training of the user, since the physical presence of the user near the system is not necessary and the operation of the device can be done remotely through the Internet

or a compatible communication system as VRML [7] which allows recreating a user-friendly virtual environment, interacting simultaneously with the real system.
In the present work it is developed an algorithm capable of recognizing a group of gestures generated by the hand, which reflect their movement in the muscles of the forearm, from where the EMG signals will be captured by means of the development card MySignals HW v2 [8], supported by Arduino and responsible of sending the data to a computer equipment, in which a graphic interface was previously developed, in charge of processing the acquired data and guiding a robotic control application in a virtual environment. The algorithms and the application are implemented through MATLAB® software.
From the above, the contribution of this article is to implement a CNN architecture for the recognition of electromyography signals that will be used for the control of a robotic manipulator in tasks of moving objects within a virtual environment. The article is organized as follows: Section 2 presents, first, the method of acquisition and preprocessing of the signal, second, the configuration of the architecture of the convolutional neuronal network (CNN), and the training and validation of the network, and finally, the virtual environment developed. In section 3 the results obtained integrated with the virtual environment are presented. Finally, section 4 presents the conclusions of the study.

2 Materials and Methods

The development of the application is demarcated by four characteristic steps, the first one corresponds to the acquisition of control-oriented electromyography signals according to a group of characteristic gestures associated with a movement of the robotic manipulator. Subsequently, the preprocessing of these signals is performed by filtering techniques and application of wavelet transform, which allow to debug a database for the training of a CNN in the third stage, which will learn to discriminate the gestures for its use. The fourth stage corresponds to the virtual interface and the robotic manipulator, showing the requirements of each one. Next, each one of these stages is presented.

2.1 Signal acquisition

The MySignals HW development card, designed for the acquisition of bio-medical signals by means of specialized sensors, allows the taking of samples of: Blood Pressure, Pulse, Oxygen in Blood, Spirometer Air Capacity, Glucometer, Snore, Body Position, Body Temperature, Electromyography, Electrocardiogram, Electroencephalography, Airflow or Breathing and Galvanic skin response. These signals and data acquired by the sensors are transmitted through the following means: Bluetooth 2.0, WiFi, GPRS, Xbee, RFID and RS-232, and depending on the need of the programmer, the means by which the communication will be made is chosen.
In this development, the electromyography signals are acquired by means of three surface electrodes, where two of these will be located in the extensor Carpi Radialis

Longus and Brevis muscle on the right forearm, its location is illustrated in Figure 1a, and the last electrode is the reference which will be located in Palmar Carpal Ligament as shown in Figure 1b, because this electrode should not be affected by the movement of the muscle in study, a place that cover this characteristics is chosen.

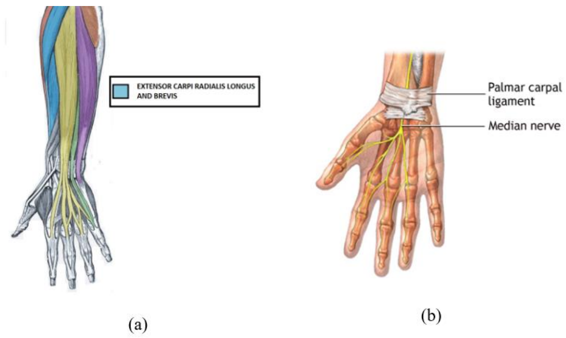


Figure 1. Muscles and Ligaments: (a) Muscle in study, Extensor Carpi Radialis Longus. (b) Ligament in study, Palmar Carpal Ligament.

The EMG signals are generated by muscle contraction, which is why it must be made a correct location of the electrodes to capture the information generated by the muscle, which is in the order of the mV. The cables used are shielded to avoid interference or noise caused by the environment, the electrical network or electronic circuits. The location of the electrodes on the forearm are shown in Figure 2.

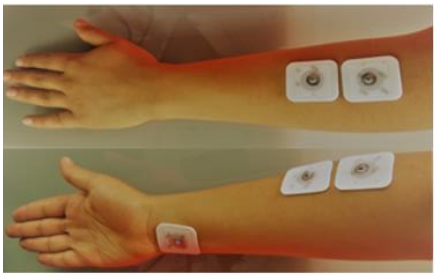


Figure 2. Location of the electrodes on the forearm.

In relation to the selected muscle group, a total of five gestures to be recognized are proposed, however it is important to take into account that the muscle has a reduced area of work, which increases the likelihood of confusion between the gestures and generating the need to implement a method that allows to differentiate the acquired signals such as the implementation of the CNN.

Figure 3 shows the gestures to be recognized, corresponding to Hand Gestures: Relax, Wave_In, Close_Fingers, Gun and Fist. These gestures are chosen because they present differentiable characteristics in muscle contraction and derived from this, from the electromyography signal.

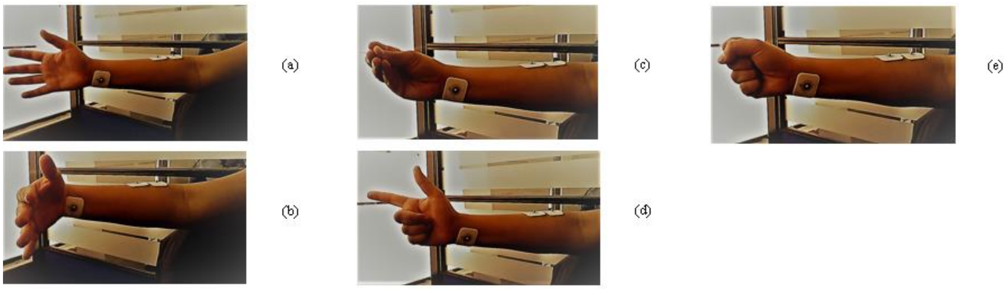


Figure 3. Hand Gestures: a) Relax, b) Wave_In, c) Close_Fingers, d) Gun and e) Fist

2.2 Signal preprocessing

For the recognition of the EMG signals, an initial pre-processing based on a Butterworth bandpass filter is performed in order to eliminate noise at low and high frequencies, generated by electronic noise and the impedance of the skin. Figure 4 shows the behavior of an EMG signal to which first- and second-order bandpass filters were applied, where it is evident that the implementation of these filters helps to extract a cleaner signal, reducing noise and defining action potentials [9]. To implement the digital filters, it was used the Butter function of the development software (MATLAB) [10], developing the design of a Butterworth filter of order 2 and a crop frequency (Wn) of [35Hz 500Hz]. This filter is characterized by having a smooth pass band and keeping the output constant until the cutoff frequency decreases at a rate of 20n dB per decade.

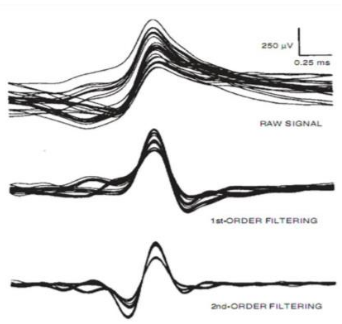


Figure 4. Effects of bandpass filter as order increases [11].

After the filtering, the Wavelet Packed Transform (WPT) is used. WPT is an extension of the Wavelet Transform (WT) responsible for representing a generalization multi-resolution in frequency, generated by the first level of approximation of the signal and increasing the qualities of the discriminant characteristics [12] and by means of the use of families of sub-bands to decompose the signal. In relation to the WT, its main advantage is that, instead of dividing the signal in a single approximation space (cA), the Packet allows to divide in approximation (cA) and in detail (cD) generating a binary recursive tree illustrated

In Figure 5 where each parent node is divided into two orthogonal sub-spaces. This Packet has been implemented in the detection of multiple biomedical signals, for example, in the detection of uterine changes with an efficiency of around 98% [13].

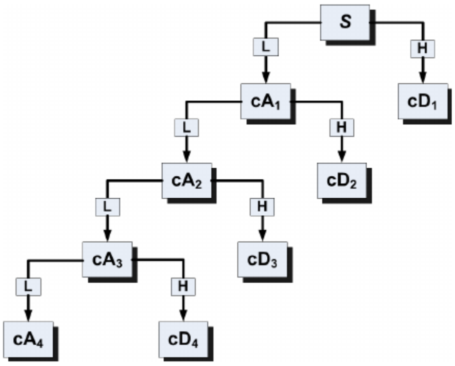
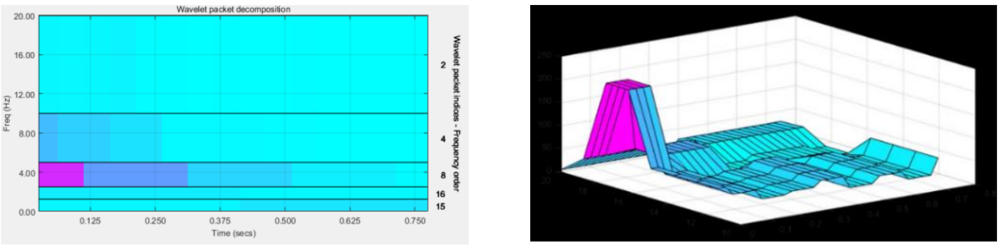


Figure 5. Wavelet decomposition tree [14]

The WPS is in charge of processing the information extracted from the WPT with the objective of recreating the feature map, which is presented and processed in a matrix form, in which it is clear to appreciate the information of the coefficients in each of the frequency bands (See Figure 6a), presenting variations of hue in the violet to blue scales. The violet color presents greater activity in the band and the highest values of the coefficients, and the cyan represents null activity. As shown in Figure 6, its behavior in 2D (Time and Frequency) and 3D dimensions (Time, Frequency and Spectrum) is presented.



(a) (b)

Figure 6. Wavelet Packet Spectrum in (a) 2D and (b) 3D.

2.3 Configuration of the architecture of the CNN

Dataset

Based on the processed EMG signals, the discrimination of each of the 5 gestures shown in Figure 3 is performed, through the recognition of patterns by means of a convolutional neuronal network, where each gesture is a category of classification at the output of the network. For this, the EMG signal is acquired during a time interval of 1 second, with an equivalent to 40 samples which will be processed by means of the WPT for the decomposition of the signal and later by the Wavelet Packet Spectrum that generates the map of characteristics with which the construction of the database that enters the Network is made. Following, the training parameters of the network are presented.

Architecture Implemented

It was implemented an architecture composed of 2 packages of Convolution/Convolution/MaxPooling with rectangular filters, combining them between time domain (Figure 7a) frequency domain (Figure 7b), in order to extract the greatest number of characteristics and evaluate the effectiveness of this architecture with combined filters.

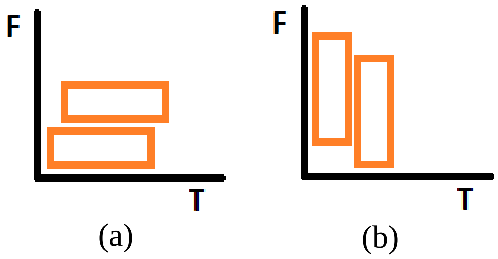


Figure 7. (a) Time Domain Filter. (b) Frequency Domain Filter.

Initially, in the first package it was decided to have a Padding of 2 since it is important to have information on the edges and with this, a better extraction of that is performed. Finally, for the classification and extraction of detailed features, three Fully-Connected Layers were used (Figure 8).

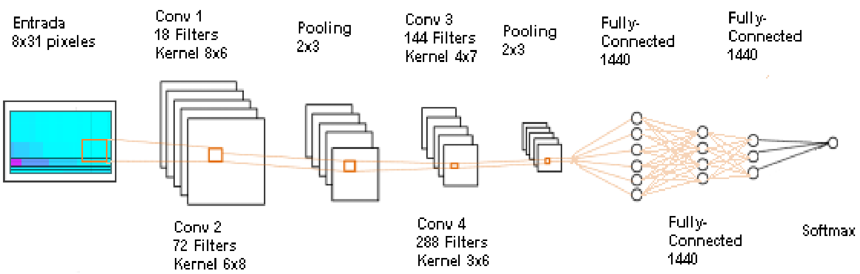


Figure 8. Architecture Implemented

Table I shows in a detailed way the construction of the architecture, where S is the Stride and P is the Padding.

TABLE I. CNN Architecture

Type	Kernel	# Filters
Input	8 x 31	-
Convolution	8x6 S=1 P=2	18
Convolution	6x8 S=1 P=2	72
MaxPooling	2x3 S=2 P=0	-
Convolution	4x7 S=1 P=1	144
Convolution	3x6 S=1 P=1	288
MaxPooling	2x3 S=2 P=0	-
Fully-Connected/ ReLU	1	1440
DropouLayer	0.5	-
Fully-Connected/ ReLU	1	1440
DropouLayer	0.5	-
Fully-Connected	1	-
Softmax	5	-

The architecture was trained with a dataset of 180 samples per category (Close_Fingers, Wave_In, Gun, Fist and Relax) for a total of 900 samples, these samples were processed by Wavelet Packet for the feature extraction mentioned before. Likewise, the behavior of the training can be seen in Figure 9, where each 9 iterations represent an epoch, and 800 epochs were done, and Training Accuracy represents the behavior of the precision that the training is having in each iteration with respect to the image batch used in that iteration, reaching a stabilization of 100% per epoch around the iteration number 3600 (epoch 400). However, it is important that you keep track of the graph in order to ensure that the training is having an appropriate behavior without overfitting.
However, it must be taken into account that, after performing the training, the network must be evaluated in order to obtain the best trained epoch. To do this, the validation is carried out in order to observe how accurate is the network when classifying the movements, using a confusion matrix to tabulate the result, as seen in Figure 10, where there are the correct and wrong predictions of the network with respect to each category.

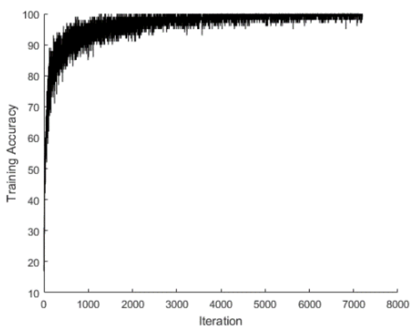


Figure 9. Training Accuracy vs Iteration

For this, a validation database must be made with the movements to be classified, this must contain a smaller number of samples than the training one and with different data. In this case, 85 samples were taken per category for a total of 425.
In general, the category with the highest misclassification (True Negatives) was the Gun, with a 92.9% of accuracy, predicting 4 of its gestures in Fist category, and 2 in Close_Fingers, and also with the highest confusion between the other categories (False Positives), where classified 5 gestures not belonging to that category, however, it has an excellent accuracy. In this way, the overall accuracy achieved by the network is 97.2%.

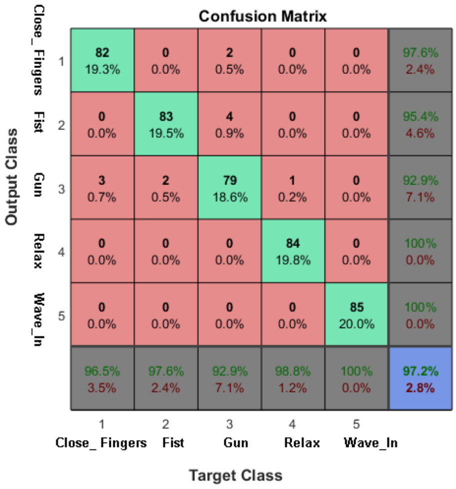


Figure 10. Confusion Matrix of the validation of the CNN.

2.4 Graphical interface and 3D virtual environment development

A graphical interface is developed (Figure 11) that allows the user to observe the response of the network and its degree of membership to the category classified, as illustrated in the green box and, in the upper part, the response of the gesture that was recognized characterized for the control of the manipulator is observed as seen in the red box, where four of the gestures are related to the movement of the joints characterized by being linear or rotational movements (Elbow, Shoulder, Gripper and Shoulder Rotation) and one to the change of direction (Mode). The behavior of the acquired and processed signal is displayed in the two panels as shown in the blue box. Finally, a virtual environment is made where it is observed the response that gestures has with the movement of a manipulator of 3 degrees of freedom and composed of three links (Elbow, Shoulder and Gripper), made with MATLAB and Simulink program and the 3D animation function for V-Realm Builder, as presented in the yellow box.

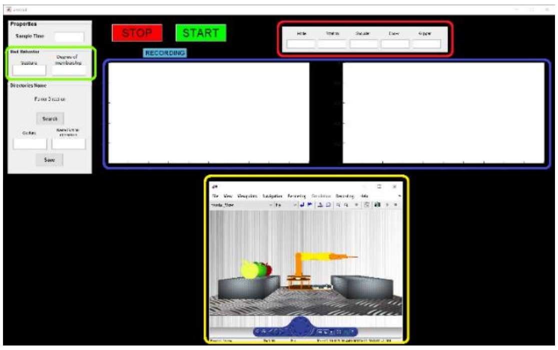


Figure 11. Graphical Interface and 3D virtual environment

For the control of the manipulator developed in a virtual environment by means of gestures generated by the forearm, it is necessary to relate each gesture to a movement that the manipulator must perform.

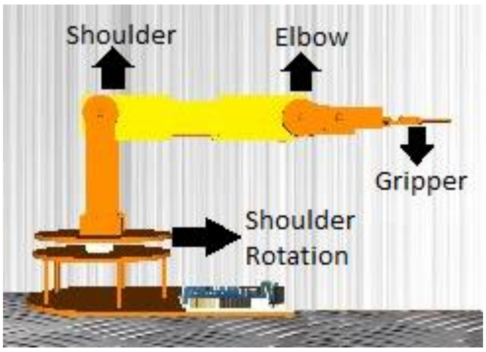


Figure 12. Manipulator joints.

Next, the Gesture-Movement association for manipulator control is presented. Figure 12 shows the joints that will be controlled with the gestures: “Close_Fingers” is in charge of controlling shoulder rotation or $\vartheta 1$, “Fist” controls the shoulder or $\theta 2$, “Gun” controls the Elbow or $\vartheta 3$, “Wave_In” controls the closing and opening of the clamp and “Relax” controls the direction in which the arm is going to move, clockwise or counterclockwise.
Therefore, in order to grasp the object, it is necessary to know its location within the virtual environment, avoiding to capture the wrong object. Likewise, to see and simulate the movement within the virtual environment, direct kinematics was implemented (see Table II), to determine the coordinates where the gripper is located, and in this way grasp the object. It must immediately take the coordinates of the gripper, in order to see the trajectory of the object according to the movement made by the manipulator.

TABLE II. Denavit-Hatenberg Parameters

Join	θ	d	a	$\propto$
1	$\vartheta 1$	L	0	90°
2	$\theta 2$	0	A	0
3	$\vartheta 3$	0	L3	0

3 Experimental Results

Initially, the signal of the MySignals HW v2 card is acquired, which is illustrated in the left part of Figure 13. This signal is processed by means of the WDT - WPS for the feature extraction generating the 3D feature map as shown in the right part of Figure 13, which varies depending on the gesture that is made. Additionally, it is possible to observe the behavior of the signals for each gesture and to notice the similarity of some of these. For example, Gun and Close_Fingers present a similar behavior in their input signal, but the 3D map shows in a more detailed way the variations in the behavior of the coefficients generated by the WPS. On the other hand, the tests that were performed in real time presented problems related to the confusion of some gestures. For example, in Figure 14. Errors presented in tests in real time.

a it is observed that the user performs the "Gun" gesture and recognizes "Close_Fingers". Similarly, in Figure 14. Errors presented in tests in real time.

b the user makes the "Close_Fingers" gesture and recognizes "Gun".

On the other hand, for the evaluation of the whole implementation, it was proposed to create a virtual environment, where the manipulator should move three blocks from one surface to another. It consists of controlling the movements of the manipulator to carry out the transfer by means of the gestures performed by the user. Figure 15 illustrates the sequence that was made for the transfer of the three apples.

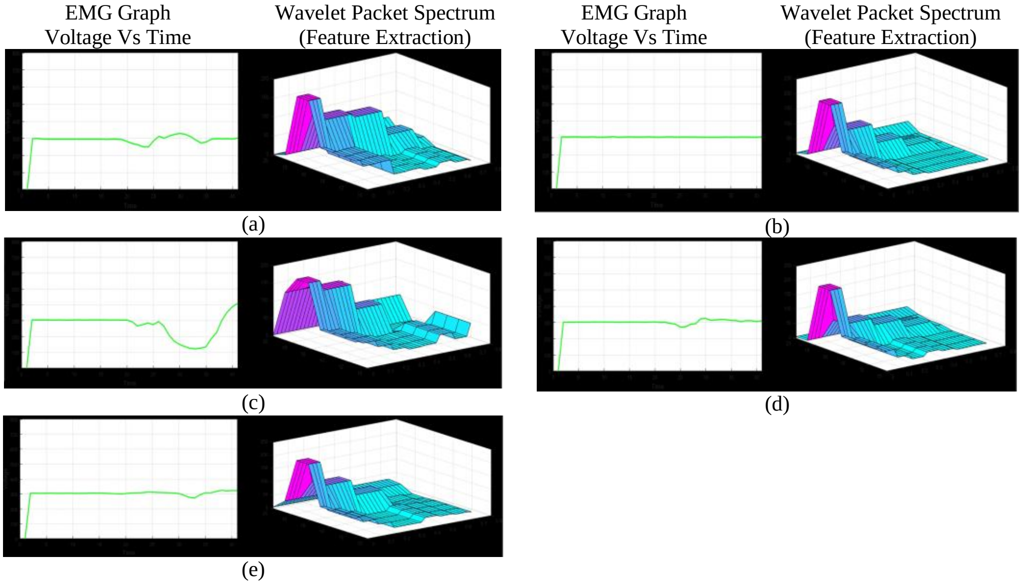
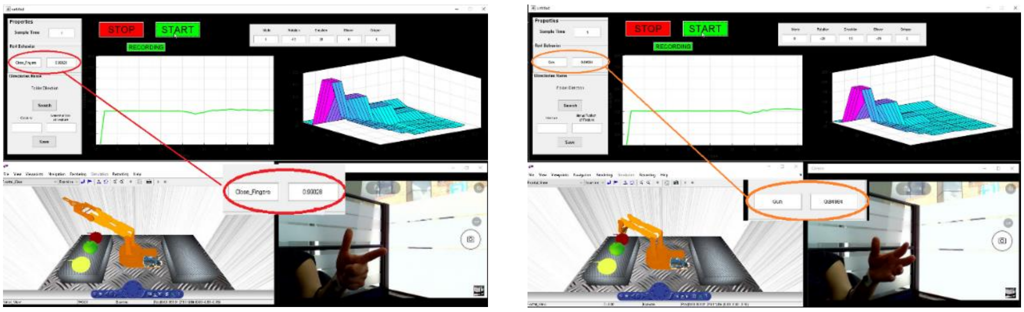


Figure 13. Behavior of the input signal and feature map for each gesture: (a) Fist, (b) Relax, (c) Wave_In, (d) Gun and (e) Close_Fingers.



(a) (b)

Figure 14. Errors presented in tests in real time.

Table III relates the results obtained by moving the three blocks from one side to the other. The accuracy is related to the tests that were carried out to transfer all the apples from one place to another. In this way, it represents the average of the gestures that were successfully recognized among the total gestures made.

TABLE III. ACCURACY ACHIEVED IN 5 USERS MOVING APPLES

User	**1**	**2**	**3**	**4**	**5**
Accuracy (%)	94,76	97,43	96,47	98,95	93,24

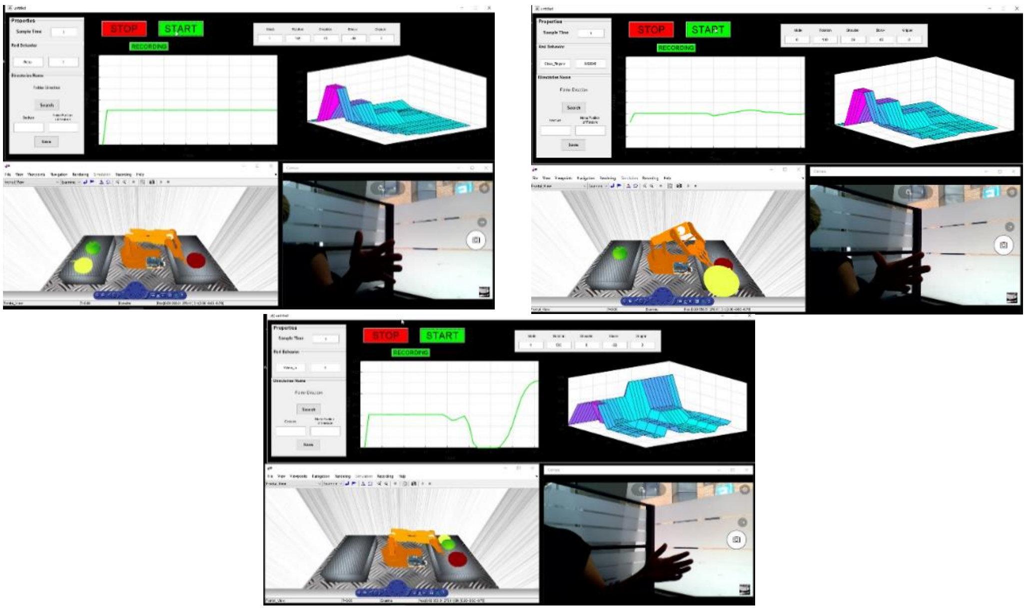


Figure 15. Results obtained in the transfer of the apples.

From the results, it can be seen that the task is carried out successfully with an acceptable degree of accuracy ranging from 93 to 98%. In this way, it reaches the highest accuracy (98.95%) in user 4 and the lowest (93.24%) in user 5. However, one of the factors that affect precision is muscle fatigue, since having to repeatedly perform the same movement causes fatigue in the hand, generating a different movement and thus causes greater confusion with gestures. In the same way, it was found that another factor depends on the force used to make the gesture.

4 Conclusions

A virtual 3D environment was designed for the control of a manipulator, by means of which the CNN operation was verified, designed for the recognition of EMG signals using the WPT, in such a way that this application allows the development of Tele-robot training environments.
The proposed CNN architecture turns out to be successful and novel, obtaining an accuracy of 97.2%. However, it is important to highlight that such success was achieved from each of the stages that conform this work. The preprocessing stage and the feature extraction method were decisive, eliminating the noise and redefining the most important characteristics of the signal, presented in more detail in each of the WPS frequency bands.
In relation to the results obtained, the development of CNN architectures focused on electromyography signals present promising developments, the implementation of this new technology in tele surgery or tele robotic, implemented from anywhere with connection. This work goes a step further in advancing research for future applications focused on the development of medical equipment that integrate this technology.

Acknowledgements. The authors are grateful to the Nueva Granada Military University, which, through its Vice chancellor for research, finances the present project with code IMP-ING-2290 and titled "Prototype of robot assistance for surgery", from which the present work is derived.

References

[1] Telecomm, Museo del Telegrafo Semblanza Historica Del Telegrafo Al Satelite, [Online]. Available: http://www.telecomm.net.mx/telecomm/dmdocuments/conocenos_telegrafo_al_satelite.pdf. [Accessed 15 09 2017].

[2] Historia de las Telecomunicaciones, [Online]. Available: http://www.uv.es/~hertz/hertz/Docencia/teoria/Historia.pdf. [Accessed 15 09 2017].

[3] A. Ollero Baturone, Robotica: Manipuladores y robots moviles, España: Marcomo, Alfaomega, 2001.

[4] S. W. Unger, H. M. Unger and R. T. Bass, AESOP robotic arm, *Surgical Endoscopy*, **8** (1994), no. 9, 1131. https://doi.org/10.1007/bf00705739

[5] R. Valero, Y. H. Ko, S. Chauhan, O. Schatloff, A. Sirvaraman, R. F. Coelho, F. Ortega, K. J. Palmer, R. Sanchez Salas, H. Davila, X. Cathelineau and V. R. Patel, Robotic Surgery: History and Teaching Impact, *Actas Urologicas Españolas*, **35** (2011), 540-545. https://doi.org/10.1016/j.acuroe.2011.12.004

[6] R. Volpe, J. Balaram, T. Ohm and R. Ivlev, "The Rocky 7 Mars Rover Prototype, *Proceedings of IEEE/RSJ International Conferenceon Intelligence Robots and Systems*, Osaka, Japan, (1996). https://doi.org/10.1109/iros.1996.569020

[7] Virtual Reality Modelling Language Society, [Online]. Available: http://www.vrml.org. [Accessed 14 09 2017].

[8] C. hacks, MySignals HW v2 -eHealth and Medical IoT Development Platform for Arduino, Libelium, [Online]. Available: https://www.cooking-hacks.com/mysignals-hw-ehealth-medical-biometric-iot-platform-arduino-tutorial/. [Accessed 08 02 2017].

[9] I. A. Cifuentes Gonzalez, Diseño y construccion de un sistema para la deteccion de señales electromiograficas, Yucatan: Universidad Autonoma de Yucatan, 2010.

[10] G. Trejo Alcantara and N. Castañeda Villa, Efecto del pre-procesamiento del EEG en el Analisis por componente independientes: reducción del artefacto del implante coclear en los potenciales Evocados Auditivos, *Revista Mexicana de Ingenieria Biomedica*, **38** (2017), no. 1, 382-389. https://doi.org/10.17488/rmib.38.1.34

[11] H. Bhullar, G. Loudon, J. Fothergil and N. Jones, Selective noninvasive electrode to study myoelectric signals, Medical and Biological Engineering and Computing, **28** (1990), no. 6, 581-586. https://doi.org/10.1007/bf02442611

[12] M. Misiti, Y. Misiti, G. Oppenheim and J.-M. Poggi, "MATLAB MathWorks," 2017. [Online]. Available: https://www.mathworks.com/help/pdf_doc/wavelet/wavelet_ug.pdf. [Accessed 03 08 207].

[13] M. Chendeb, M. Khalil and J. Duchene, Methodology of wavelet packet selection for event detection, *Signal Processing*, **86** (2006), no. 12, 3826-3841. https://doi.org/10.1016/j.sigpro.2006.03.029

[14] A. Phinyomark, C. Limsakul and P. Phukpattaranont, Wavelet-based Denoising Algorithm for Robust EMG Pattern Recognition, *Fluctuation and Noise Letters*, **10** (2011), no. 02, 157-167. https://doi.org/10.1142/s0219477511000466

Received: November 17, 2017; Published: December 10, 2017