ISSN 1870-4069

Hand Gesture Recognition Applied to the Interaction with 3D Models in Virtual Reality

Ángel Leonardo Valdivieso-Caraguay, Óscar Mauricio Rivera-Cajía, Bryan Norberto Flores-Sarango, Lorena Isabel Barona-López, Marco E. Benalcázar

Escuela Politécnica Nacional,
Departamento de Informática y Ciencias de la Computación,
Artificial Intelligence and Computer Vision Research Lab,
Ecuador

angel.valdivieso@epn.edu.ec

Abstract. In this work, we present a real-time hand gesture based Human Computer Interaction (HCI) system for control a Virtual Reality (VR) application by using of Oculus Rift and Myo armband. For this purpose, an Hand Gesture Recognition (HGR) model and a VR application were implemented. The K-Nearest Neighbors (KNN) and Dynamic Time Warping (DTW) algorithms were applied to develop the HGR model. The inputs to this model are signals of 11 hand gestures measured by Myo Armband and G-Force Pro using their built-in surface electromyography (EMG) dry sensors and inertial measurement unit (IMU). The outcome of the HGR model is the designation that characterizes the gesture performed by the user. The VR application was developed by the game engine Unity using Oculus Rift as input device into virtual environment. It allows navigate over an interface and manipulate three-dimensional (3D) objects taking advantage of their properties for a sophisticated experience. The HGR model is used in the VR application where each identified gesture performs an action. The system present a natural communication through hand gestures in a virtual environment. In average, we achieved real-time gesture classification with an accuracy of 82% on eleven distinct gestures. The SUS test results rank our system as excellent in terms of usability.

Keywords: Human computer interaction, hand gesture recognition, virtual reality, K-nearest neighbors, dynamic time warping, electromyography, inertial measurement unit.

1 Introduction

The study of interfaces between humans and computers is known as Human Computer Interaction (HCI). Traditional HCI methods such as keyboards, mouse or touch screens are often unfriendly when interacting with computers. The study of the field of gesture recognition in combination with HCI transcended this barrier because the use of gestures is a more natural way to provide an interface between a user and a computer [30, 1]. These improvements in HCI technology are also leading to advances in another field closely related to HCI: virtual reality (VR).

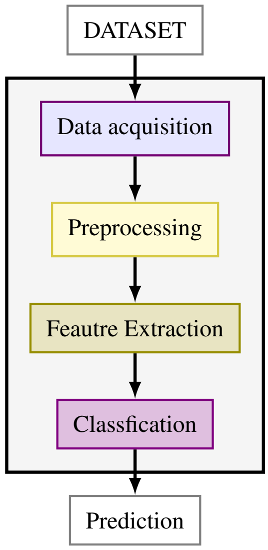


Fig. 1. Hand gesture system framework.

Hand Gesture Recognition (HGR) models are human–computer systems that determine what gesture was performed and when it is performed [12]. In this work, a gesture recognition system based on EMG and IMU data is divided into 4 stages: data acquisition, preprocessing, feature extraction and classification [7] as shown in Figure 1. Those models can acquire data with different instruments, such as gloves, vision sensors, inertial measurement units (IMUs), surface electromyography sensors or a combinations of devices [12].

In this work we combine surface EMG and IMU sensors. Surface EMG is a technique that records the action potentials of the muscle fibers with surface sensors [14]. IMU is a device that records four features: velocity, shape, location and orientation in motion capture on specific body by using a combination of accelerometer, gyroscopy and magnetometer sensor [19, 5]. The acquired information is combined depending on the two categories of gestures according to their type of interaction [27]:

– Static gestures - gestures based on a single posture that is maintained for a certain amount of time.

– Dynamic gestures - gestures based on a motion trajectory.

On this context, we propose a HGR model based on static and dynamic gestures measured by the dry surface EMG and IMU sensors built-in the Myo Armband. Additionally, we propose a virtual reality application controlled by hand gestures where our HGR model can be used to control it.

The application provides a natural interaction of actions through gestures [9, 23]. In addition, it allows a detailed examination of the 3D models. Virtual reality gives full control of the environment to the user through the glasses movement and, with our proposal, hand gestures.

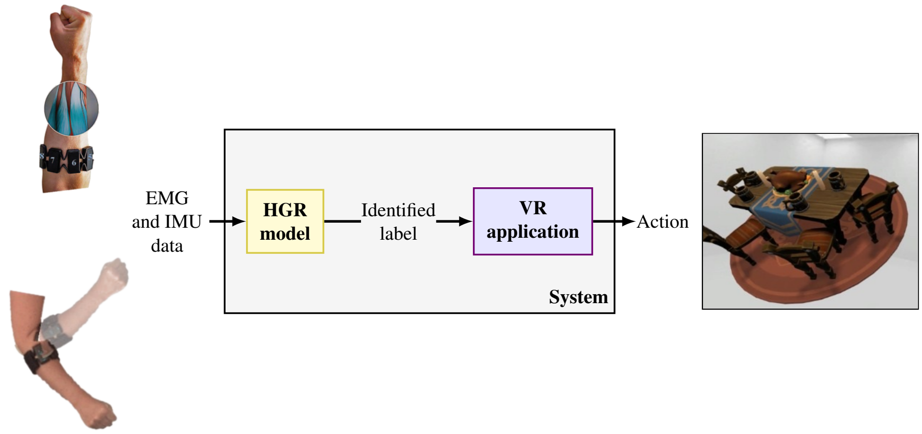


Fig. 2. System.

Figure 2 shows the general purpose of this work, where the HGR model and the application complement each other as a system that combines the potential of the HGR field applied to a VR application. The rest of the paper is organized as follows. Section II describes the works related to gesture recognition and their applications. The dataset, the structure of the HGR model, the description of the application, and integration of both components are explained in Section III. Experimental results and their analysis are presented in Section IV. Finally, Section V concludes this work.

2 Related Works

The development of HGR models is used to perform different tasks including but not limited to motor control, prosthetic device control, and hand motion classification [33]. Gesture recognition applied to VR has a wide array of fields, such as vehicle driving simulation [29, 32], games [28, 20], navigation of maps [16], sign language gestures [26], among others.

The most frequent machine learning algorithms applied to gesture recognition based on EMG and IMU data are artificial neural network-based algorithms [11, 22, 8, 34], classifier-based algorithms [18, 25, 17], and linear discriminant analysis [13, 31]. According to the literature review, for this work we choose the kNN classification algorithm because it is one of the simplest and most optimal models in terms of classification.

The use of Myo Armband and a virtual environment was already evaluated in [10] through SUS, which gives us an idea of the usability qualification, however the overall result of [10] is lower than the obtained in this article. In addition, according to [21], it is easier to use the Myo Armband in the control of tasks related with daily life. Likewise, Rawat[21] affirms that users can interact with various applications simply by making our hands work, which will be demonstrated in this article.

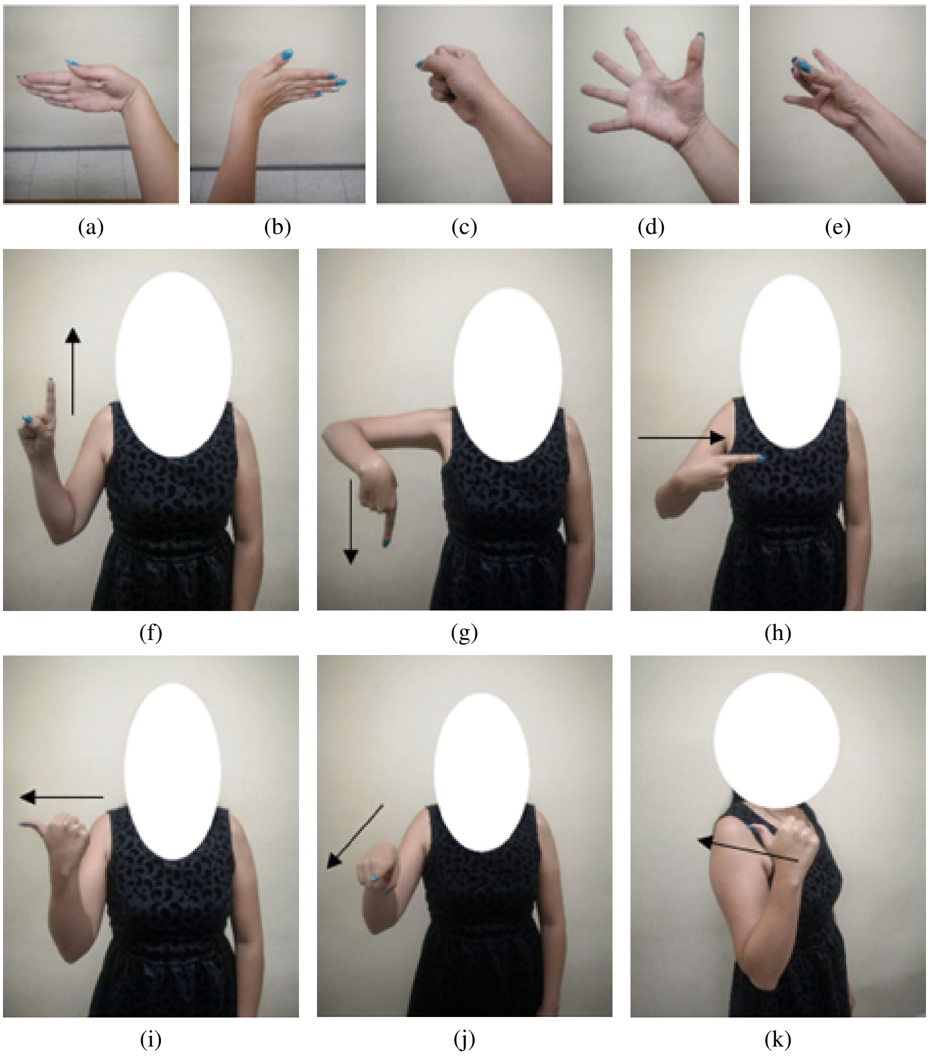


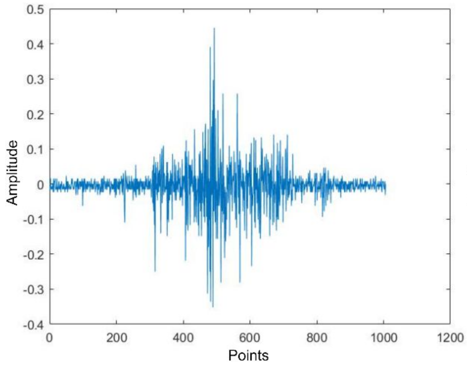
Fig. 3. Gestures used to the Dataset (a) waveIn, (b) waveOut, (c) fist, (d) open, (e) pinch, (f) up, (g) down, (h) left, (i) right, (j) forward and (k) backward.

3 Methodology

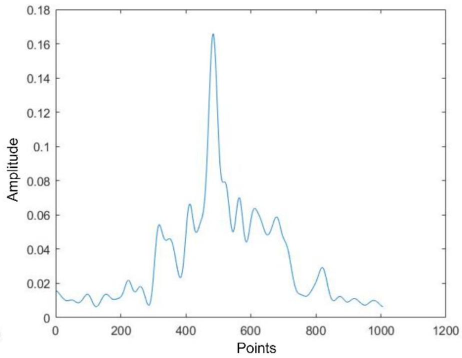
In this section, we describe the dataset and structure of HGR model and the VR application.

3.1 Dataset and HGR Model

This subsection describes the dataset and the HGR model proposed in this work. The HGR model is composed by the following steps: data acquisition, preprocessing, feature extraction and classification.



(a)



(b)

Fig. 4. (a) raw EMG signal (b) preprocessed EMG signal.

Dataset. EMG and IMU signals from 85 users are included in the dataset, which can be found in [2]. The signals were gathered while performing a set of 12 gestures, which includes the 11 gestures specified in Figure 3 (wave in, wave out, fist, open, pinch, up, down, left, right, forward, and backward), along with an additional gesture denoted as the "relax" gesture. Each user performed a total of 180 samples by executing 15 repetitions of 5-second intervals for each gesture.

Data acquisition. The sliding window technique was applied to acquire the input data for the classifier. We used a window of $N = 480$ points for the classification of EMG signals and another window of $M = 480$ points for IMU signals. The complete signal was analyzed, in both cases, with the same number of iterations using the stride of $L = 200$.

Preprocessing. The purpose of this stage is to facilitate the subsequent phases of feature extraction and classification. It was applied only to EMG signals because they have an irregular appearance, as shown in Figure 4a. Preprocessing consists of rectifying and filtering the window N. For rectification, the absolute value of both windows was calculated. Then, for filtering, a fourth-order Butterworth filter with and a cutoff frequency of 5 Hz to reduce the noise was applied. Figure 4b shows the result of the preprocessing as a smoother EMG signal.

Feature extraction. We work with EMG windows to classify static gestures since the execution of these gestures is dominated by muscle activity, which is measured using EMG signals. Similarly, we use IMU windows to classify dynamic gestures which are dominated by arm movement and measured by IMU signals.

In this stage, the two EMG and IMU sliding windows are analyzed in order to select one window for the classification phase. For this purpose, the "energy" feature extraction function is applied over both type sliding windows. It measures the energy distribution of both signals [6]. The output of this function is a feature vector used as input to a switch, as shown in Figure 5. The switch is based on a logistic linear classification model.

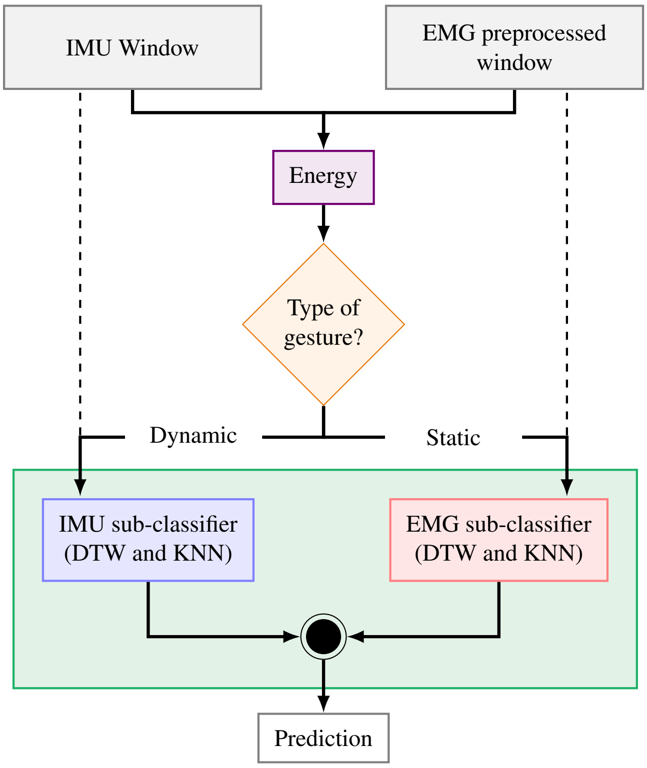


Fig. 5. Feature extraction and classification stages.

This model determines whether a pair of EMG and IMU windows of the same signal corresponds to a static or dynamic gesture. The model was trained using the same "energy" function for the training dataset. Its evaluation achieved an average classification accuracy rate of 93.02%. Through this switch, a feature window is chosen according to the type of gesture determined and, consequently, the subclassifier corresponding to the feature window (EMG or IMU).

Classification. The classification module works exclusively with EMG signals or with IMU signals separately. Consequently, the classifier is constituted by 2 sub-classifiers as shown in Figure 5: the EMG sub-classifier which classifies static gestures and the IMU sub-classifier which classifies dynamic gestures. Both subclassifiers were built using the KNN algorithm.

The approach on which KNN is based in this work is the estimation of the conditional probability based on the relative frequency of the nearest neighbors to the window to be classified. For both subclassifiers, the value of k nearest neighbors was determined with the formula $k = \mathrm{ceil}[\log_2 N]$, where N is the number of samples composing the dataset corresponding to the subclassifier. The threshold was set at 80%, to avoid false positives. The code of the proposed HGR model was written with MATLAB version 2021b and is publicy available in [3].

Table 1. Technical specifications of the application development environment.

Aspect	Specification
Unity Version	2021.2.3f1
GPU	NVIDIA GeForce GTX 1650
CPU	Intel Core i7-10700F
RAM	16 GB
HDD	932 GB
SSD	233 GB
Motherboard	Prime H410M-E

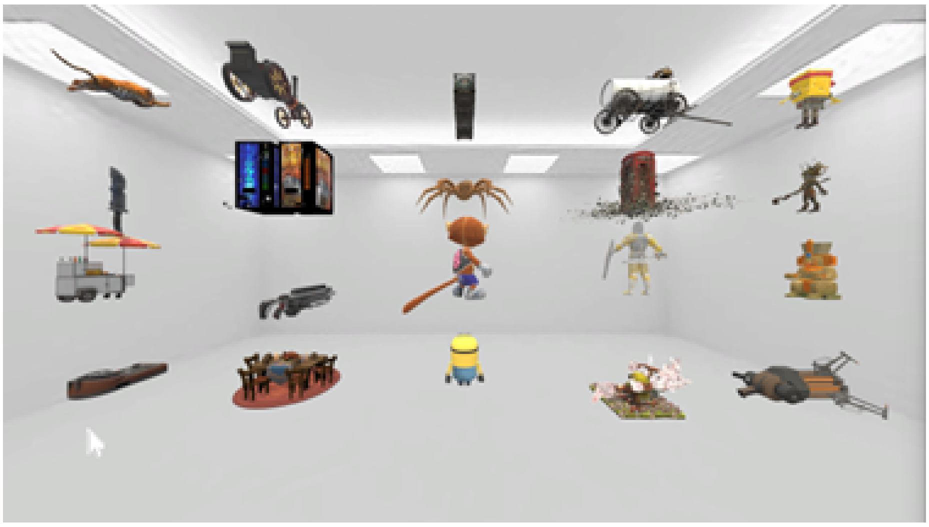


Fig. 6. 3D model viewer “Gallery” scene.

3.2 3D Model Viewer

To demonstrate the feasibility of the model, a software application (3D Model Viewer) controlled by the HGR system is developed. The 11 hand gestures are used to execute interaction actions with models within the application, such as: moving between models, moving the model, accepting an action, among others. The gestures are detected thanks to the use of a recognition model detailed above.

The methodology for the 3D model viewer uses the waterfall model. [24] . The phases of the methodology include: requirements specification, design, development, verification and validation, and operation of the application. The requirements specification phase collected user needs through meetings with the end user. This phase expresses the user’s intention in 13 use cases, as shown in Annex B.

In the design phase, the space names, scenes, class diagram, and complements were described. This phase expresses the architecture of the application through the class diagram, as observed in Annex A. Likewise, in the development phase, the source code was implemented using C# as the programming language.

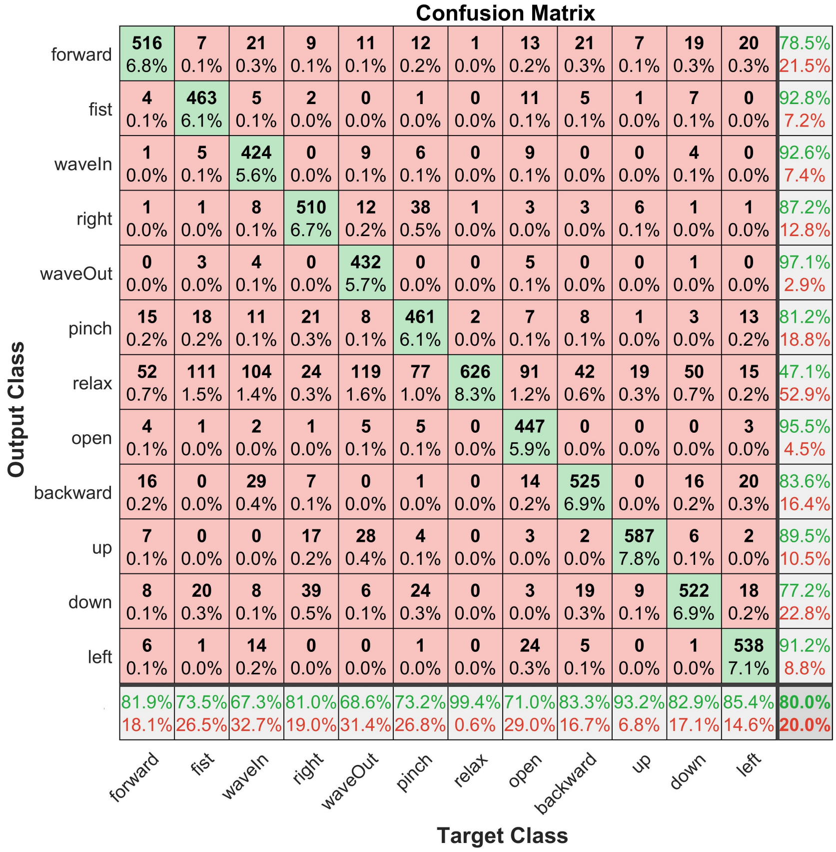


Fig. 7. Confusion matrix for proposed model.

Among other tools for coding the viewer, the IDE VS 2022 version 17.2 and Unity version 2021.2.3f1 were used. Additional technical specifications for the development are described in the Table 1. To comply with the verification and validation phase, integration and acceptance tests were carried out. The integration tests were ascending to perform evaluations of the modules from lower to higher levels. For its part, the acceptance tests are based on each use case specified in Annex B. Finally, in the operation phase, the application was compiled with support for virtual reality in the 64-bit Windows 10 operating system. One of the scenes (Gallery) of the operational application is shown in Figure 6.

3.3 Integration

Once the developed application operational, communication between the application development environment and the recognition model was established. This integration used socket technology where Unity and Matlab took the roles of server and client, respectively.

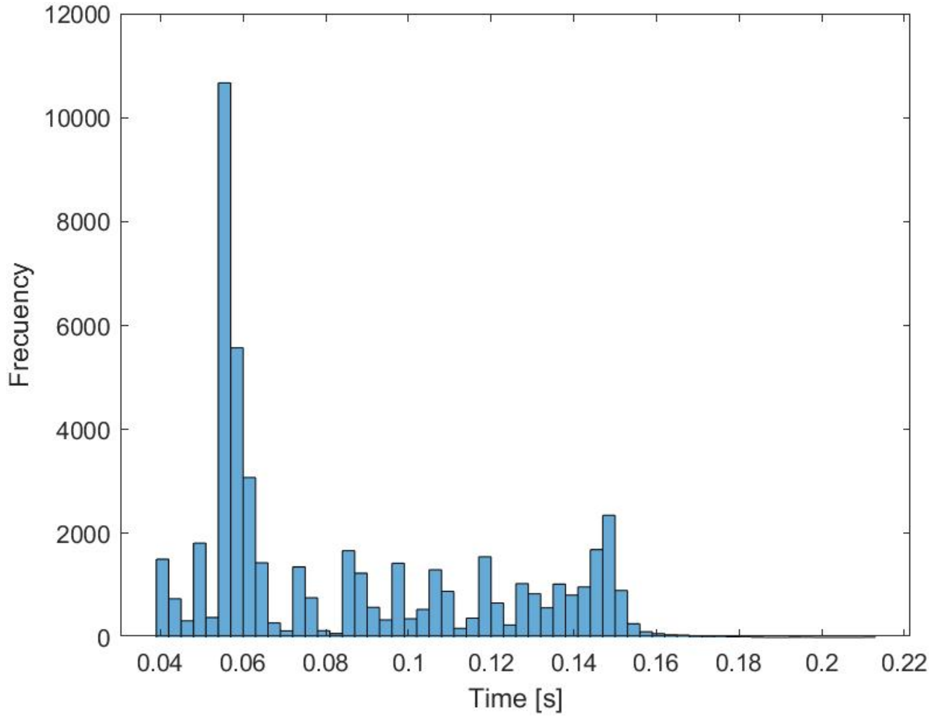


Fig. 8. Histogram of the processing time of each window observation.

The IP address used was "127.0.0.1", the port was 55001 and 10 seconds as time out. The source code of the proposed HGR model and the VR application is publicy available in [3]. A video demonstration of the execution of the complete system can be found in [4].

4 Results

In this section, we present the results of the tests applied to the HGR model and the system.

4.1 Performance of the HGR Model

We evaluate the performance of HGR model with 42 users of dataset described in Section III. Classification results are shown in the Figure 7. The label that presents the highest sensitivity is Up (93.2%) while the label with the lowest sensitivity is WaveIn (67.3%).

On the other hand, the labels with the highest and lowest precision are WaveOut (97.1%) and Down (77.2%), respectively. It is also illustrated that the overall classification accuracy of the model is 80% while the average accuracy is 84.45%.

Table 2. Result of SUS.

Person\Item	1	2	3	4	5	6	7	8	9	10	**Total**
1	5	0	0	0	4	0	5	0	5	0	**97,5**
2	5	1	5	2	5	0	5	0	5	2	**100**
3	4	0	4	2	5	0	4	0	4	1	**95**
4	5	0	4	4	5	0	5	0	5	0	**100**
5	5	0	4	0	5	5	5	1	5	0	**95**
6	3	1	4	1	3	3	4	1	5	4	**72,5**
7	3	3	4	2	1	1	3	4	3	1	**57,5**
8	3	3	3	0	3	1	3	3	5	0	**75**
9	1	0	4	1	2	1	4	1	5	0	**82,5**
10	4	2	4	4	4	2	2	2	3	5	**55**

It is important to note that the results obtained in terms of classification (80% accuracy) taking into account the high number of labels to classify (11 gestures shown in Figure 3) are considered highly satisfactory for our RV application. The response time is defined as the time that each sub-classifier (EMG or IMU) takes to process and classify each window of the signal.

The processing time of each window was measured and stored for later analysis. Figure 8 shows the processing times for each processed window. The highest processing time is 0.22 seconds while most of measures remains below 0.16 seconds. In conclusion, our HGR model returns a real time response below 0.3 seconds, which it can be considered as real time.

4.2 Usability Test Results

The system was subjected to the System Usability Scale (SUS). The evaluation was carried out by 10 participants between 21 and 37 years old. The 10 statements received a rating between 0 and 5. The rating of 0 is for the evaluator to show that they totally disagree while rating 5 is to show total agreement. The results of the 10 users for each statement are shown in Table 2. To calculate the global rating of the application for each user, the following formula was used:

$$t = (x - 5) + 2.5(25 - y), \quad (1)$$

where:

$x =$ Sum of even item responses.
$y =$ Sum of odd item responses.
t = Global.

The average of the global was 83. According to [15], the usability score of the system is Excellent. The obtained score shares the acceptable range as well as many other applications described in [15], such as Excel, Gmail, among others.

5 Conclusions and Future Work

A HGR model based on EMG and IMU signals together with a VR application has been presented. The structure of the developed HGR model is based on the stages of data acquisition, preprocessing, feature extraction and classification.

The EMG-IMU-EPN-100+ was used to evaluate the model, and the results show a classification accuracy of $80.04\% \pm 13.66\%$ and a recognition accuracy of $66.12\% \pm 18.30\%$. The response time of the model is below 0.22 seconds which, according to the literature, validates the real-time performance of the model. Similarly, the VR application was successfully integrated with the HGR model. A global average of 83 in SUS scale demonstrate the acceptable range of the system.

Future work on the HGR model includes testing with a stride between consecutive windows of less than 1 second. Similarly, the use of parallel processing for the classification step is also a potential improvement. For its part, the 3D Model Viewer could implement wireless virtual reality glasses, improving the user's mobility. The application could also provide model management tasks from its graphical interface.

Acknowledgments. The authors gratefully acknowledge the financial support provided by the Escuela Politécnica Nacional (EPN) for the development of the research project "PIGR-22-09 Avances para el desarrollo de un prototipo de prótesis mioeléctrica de mano y control avanzado de su operación usando Inteligencia Artificial".

References

1. Ahmed, S., Kallu, K. D., Ahmed, S., Cho, S. H.: Hand gestures recognition using radar sensors for human-computer-interaction: A review. Remote Sensing, vol. 13, no. 3, pp. 527 (2021) doi: 10.3390/rs13030527
2. Artificial Intelligence and Computer Vision Research Lab Alan Turing: Dataset EMG-IMU-EPN-100+ (2022)
3. Artificial Intelligence and Computer Vision Research Lab Alan Turing: Source code of the design and application of a recognition model of 11 hand gestures using EMG, IMU, DTW and KNN (2022)
4. Artificial Intelligence and Computer Vision Research Lab Alan Turing: Video demo of the design and application of a recognition model of 11 hand gestures using EMG, IMU, DTW and KNN (2022)
5. Arun-Faisal, I., Waluyo-Purboyo, T., Raharjo-Ansori, A. S.: A review of accelerometer sensor and gyroscope sensor in IMU sensors on motion capture. Journal of Engineering and Applied Sciences, vol. 15, no. 3, pp. 826–829 (2019) doi: 10.36478/jeasci.2020.826.829
6. Barona-López, L. I., Valdivieso-Caraguay, A. L., Vimos, V. H., Zea, J. A., Vásconez, J. P., Álvarez, M., Benalcázar, M. E.: An energy-based method for orientation correction of EMG bracelet sensors in hand gesture recognition systems. Sensors, vol. 20, no. 21, pp. 6327 (2020) doi: 10.3390/s20216327
7. Benalcázar, M. E., Jaramillo, A. G., Jonathan, Zea, A., Páez, A., Andaluz, V. H.: Hand gesture recognition using machine learning and the Myo armband. In: 25th European Signal Processing Conference (EUSIPCO), pp. 1040–1044 (2017) doi: 10.23919/EUSIPCO.2017.8081366

8. Can, C., Kaya, Y., Kılıç, F.: A deep convolutional neural network model for hand gesture recognition in 2d near-infrared images. Biomedical Physics and Engineering Express, vol. 7, no. 5, pp. 55005 (2021) doi: 10.1088/2057-1976/ac0d91
9. Chirinos-Delfino, Y.: La realidad virtual como mediadora de aprendizajes: Desarrollo de una aplicación móvil de realidad virtual orientada a niños. Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología, vol. 1, no. 30, pp. 136–137 (2021) doi: 10.24215 /18509959.30.e16
10. De-Paolis, L. T., De-Luca, V.: The impact of the input interface in a virtual environment: The Vive controller and the Myo armband. Virtual Reality, vol. 24, no. 3, pp. 483–502 (2020) doi: 10.1007/s10055-019-00409-6
11. Djemal, A., Hellara, H., Barioul, R., Atitallah, B. B., Ramalingame, R., Fricke, E., Kanoun, O.: Real-time model for dynamic hand gestures classification based on inertial sensor. In: IEEE 9th International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications, pp. 1–6 (2022) doi: 10.1109/CIVEMSA53371 .2022.9853648
12. Jaramillo-Yánez, A., Benalcázar, M. E., Mena-Maldonado, E.: Real-time hand gesture recognition using surface electromyography and machine learning: A systematic literature review. Sensors, vol. 20, no. 9, pp. 2467 (2020) doi: 10.3390/s20092467
13. Khushaba, R. N., Nazarpour, K.: Decoding HD-EMG signals for myoelectric control - how small can the analysis window size be? IEEE Robotics and Automation Letters, vol. 6, no. 4, pp. 8569–8574 (2021) doi: 10.1109/LRA.2021.3111850
14. Koch, P., Dreier, M., Maass, M., Böhme, M., Phan, H., Mertins, A.: A recurrent neural network for hand gesture recognition based on accelerometer data. In: 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 5088–5091 (2019) doi: 10.1109/EMBC.2019.8856844
15. Kortum, P. T., Bangor, A.: Usability ratings for everyday products measured with the system usability scale. International Journal of Human-Computer Interaction, vol. 29, no. 2, pp. 67–76 (2013) doi: 10.1080/10447318.2012.681221
16. Lee, Y. S., Sohn, B. S.: Immersive gesture interfaces for navigation of 3D maps in HMD-based mobile virtual environments. Mobile Information Systems, vol. 2018 (2018) doi: 10.1155/2018/2585797
17. Lian, K. Y., Chiu, C. C., Hong, Y. J., Sung, W. T.: Wearable armband for real time hand gesture recognition. In: IEEE International Conference on Systems, Man, and Cybernetics, pp. 2992–2995 (2017) doi: 10.1109/SMC.2017.8123083
18. Maragliulo, S., Lopes, P. F. A., Osório, L. B., De Almeida, A. T., Tavakoli, M.: Foot gesture recognition through dual channel wearable EMG system. IEEE Sensors Journal, vol. 19, no. 22, pp. 10187–10197 (2019) doi: 10.1109/JSEN.2019.2931715
19. Nandwana, B., Tazi, S., Trivedi, S., Kumar, D., Vipparthi, S. K.: A survey paper on hand gesture recognition. In: 7th International Conference on Communication Systems and Network Technologies, pp. 147–152 (2017) doi: 10.1109/CSNT.2017.8418527
20. Rautaray, S. S., Agrawal, A.: Interaction with virtual game through hand gesture recognition. In: International Conference on Multimedia, Signal Processing and Communication Technologies, pp. 244–247 (2011) doi: 10.1109/MSPCT.2011.6150485
21. Rawat, S., Vats, S., Kumar, P.: Evaluating and exploring the MYO ARMBAND. In: International Conference System Modeling and Advancement in Research Trends, pp. 115–120 (2016) doi: 10.1109/SYSMART.2016.7894501
22. Riaz, M. M., Zhang, Z.: Surface EMG real-time chinese language recognition using artificial neural networks. In: Intelligent Life System Modelling, Image Processing and Analysis, Springer Singapore, pp. 114–122 (2021) doi: 10.1007/978-981-16-7207-1_12

23. Slater, M., Wilbur, S.: A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. Presence: Teleoperators and Virtual Environments, vol. 6, no. 6, pp. 603–616 (1997) doi: 10.1162/pres.1997.6.6.603
24. Sommerville, I.: Software Engineering. Pearson (2011)
25. Tavakoli, M., Benussi, C., Lopes, P. A., Osorio, L. B., de Almeida, A. T.: Robust hand gesture recognition with a double channel surface EMG wearable armband and SVM classifier. Biomedical Signal Processing and Control, vol. 46, pp. 121–130 (2018) doi: 10.1016/j.bspc.2018.07.010
26. Vaitkevičius, A., Taroza, M., Blažauskas, T., Damaševičius, R., Maskeliūnas, R., Woźniak, M.: Recognition of american sign language gestures in a virtual reality using leap motion. Applied Sciences, vol. 9, no. 3, pp. 445 (2019) doi: 10.3390/app9030445
27. Vatavu, R. D., Pentiuc, S. G.: Multi-level representation of gesture as command for human computer interaction. Computing and Informatics, vol. 27, no. 6, pp. 837–851 (2008)
28. Wen, F., Sun, Z., He, T., Shi, Q., Zhu, M., Zhang, Z., Li, L., Zhang, T., Lee, C.: Machine learning glove using self-powered conductive superhydrophobic triboelectric textile for gesture recognition in VR/AR applications. Advanced Science, vol. 7, no. 14 (2020) doi: 10.1002/advs.202000261
29. Xu, D.: A neural network approach for hand gesture recognition in virtual reality driving training system of SPG. In: 18th International Conference on Pattern Recognition, vol. 3, pp. 519–522 (2006) doi: 10.1109/ICPR.2006.109
30. Yeo, H. S., Lee, B. G., Lim, H.: Hand tracking and gesture recognition system for human-computer interaction using low-cost hardware. Multimedia Tools and Applications, vol. 74, no. 8, pp. 2687–2715 (2015) doi: 10.1007/s11042-013-1501-1
31. Young, A. J., Smith, L. H., Rouse, E. J., Hargrove, L. J.: Classification of simultaneous movements using surface EMG pattern recognition. IEEE Transactions on Bio-Medical Engineering, vol. 60, no. 5, pp. 1250–1258 (2013) doi: 10.1109/tbme.2012.2232293
32. Young, G., Milne, H., Griffiths, D., Padfield, E., Blenkinsopp, R., Georgiou, O.: Designing mid-air haptic gesture controlled user interfaces for cars. In: Proceedings of the ACM on Human-Computer Interaction, vol. 4, pp. 1–23 (2020) doi: 10.1145/3397869
33. Zaman-Khan, R., Ibraheem, N.: Hand gesture recognition: A literature review. International Journal of Artificial Intelligence and Applications, vol. 3, no. 4, pp. 161–174 (2012) doi: 10.5121/ijaia.2012.3412
34. Zhou, C., Yang, L., Liao, H., Liang, B., Ye, X.: Ankle foot motion recognition based on wireless wearable sEMG and acceleration sensors for smart AFO. Sensors and Actuators A: Physical, vol. 331, pp. 113025 (2021) doi: 10.1016/j.sna.2021.113025