

GLOVE-BASED GESTURE RECOGNITION SYSTEM

MARIA EUGENIA CABRERA^{*}, JUAN MANUEL BOGADO, LEONARDO FERMIN,
RAUL ACUÑA, DIMITAR RALEV

*Mechatronics Research Group, Simon Bolivar University, ELE Bldg 3rd Floor, Lab 302.
Valle de Sartenejas
Caracas, Miranda 1080-A, Venezuela*

Among the research topics developing within the Mechatronics Group at Simon Bolivar University, there's a field corresponding to Human-Computer Interaction (HCI), which is the center of this paper, regarding the design and implementation of a hand gesture recognition system, based on the information gathered with a data glove. The data glove sensors convey the degree of flexion in each finger of the hand. Additionally, an accelerometer sensor was placed on top of the glove to acquire the orientation of the hand in three axes. With this information, the system would begin a decision making process based on artificial intelligence tools, like neural networks, to recognize the American Sign Language (ASL) signs for fingerspelling. The whole system is executed by interfacing with a user, to collect the data and display the result of its classification.

1. Introduction

The use of gestures among humans, both in the form of pantomime or by using sign language, is closely linked to speech and represents an effective way of communication, used even prior to talking.

The formality of the set of rules chosen in each case is related to the validity of the performed gestures, which means that a pantomime gesture could be complimenting speech in a spontaneously and colloquial manner, or the same gesture be considered a part of a sign language and posses a particular meaning, whether it is to perform an action or convey some type of information (1).

There's an approach within HCI, relating to the communication structures between man and machines, and studying the physiological, psychophysical and psychological aspects associated to the brain and its abilities to interpret information. HCI generally focuses on the design, implementation and evaluation of computer systems, through which a human operator performs some kind of information exchange (2). Considering that the hands are used as a means of communication, one could think of a system where their movement could be interpreted as the input, to be analyzed and classified, using a determined set of rules that would pass along the information.

In order to decide which input device to acquire the user's hand movements, there are a few items to consider, regarding effectiveness, comfort of the user, and the selection also needs to be suitable to the working environment to fulfill the tasks set for the device within the system(3). In the topic of hand gestures recognition, the input device must capture the position of each finger on the hand, as well as the hand's orientation in its rotation axis. The most commonly used devices reported in the literature are data gloves for motion capture and cameras for artificial vision. A more detailed revision of the state of the art will appear in the following subsection.

The general idea of this particular project is based on the design of a system that allows the recognition of a set of established gestures, such as the alphabet in American Sign Language, performed by a user with a single hand, using a data glove as an input device. In that system, artificial intelligence tools will be used, like artificial neural networks, to achieve an interaction between a computer and the user, where the first classifies and recognizes the gestures executed by the latter. Within the alphabet of ASL, there are static and dynamic gestures: static gestures can be accomplished by maintaining a pose of the hand, while dynamic gestures join a specific pose of the hand with a movement.

1.1. *State of the Art*

Every research work starts by reviewing the literature on the subject, in order to get a sense of what people are doing all around the world and to know the possible applications and solutions presented for each topic specifically.

On the state of the art regarding the subject of gesture recognition, there are many teams in many countries approaching the matter with two particular goals; at least those two represent the most common among the reviewed literature. One goal, seeks for gesture recognition to aid hearing impaired people, giving them a way to interact with people who aren't fluid in such language, through the computer. The other goal relates to controlling robots, by the performed gestures of its operator.

Another feature reviewed on the literature, was about the tools used to achieve such said goals and the proposed solutions in each case. The work published by Parton (4), reviews a number of approaches within the field of Artificial Intelligence (AI) to effectively recognize and translate sign language: from Artificial Neural Networks (ANN), Hidden Markov Models (HMM), Artificial Vision, Virtual Reality, and motion capture using data gloves or electromyographic (EMG) sensors. Which leads to believe that there are two main approaches to recognizing hand gestures: one is based on motion capture

of the user's hand, while the other bases on artificial vision to determine the gestures made in a specific time. Both approaches have their pros and cons.

Vision-based gesture recognition relies on image capture by one or more cameras and applying algorithms to extract visual features, in order to recognize the gesture made. The fact that the person making the gesture is totally detached from the system is one of the most important benefits in working with visual recognition. On the other hand, the system can be very susceptible to changes in the environment, like lighting or background figures and colors that could affect the process.

The work by Ricco & Tomasi (5), looks for recognizing features not only in the images where there is a gesture but they also consider in their algorithm the frames of transition to recognize gestures; while Gui *et.al* (6) focus on intrinsic characteristics of the gestures to recognize them. Another method reported by Malima *et.al* is to obtain the image centroid to recognize positions of the hand to command a robot to move. (7)

The approach based on capturing the user's motions, physically links the user to the system but in return, it detaches it from the changes in the environment giving the opportunity to use the system in different places and obtaining the same results. The problem with incorporating the user to the acquisition process is that it causes a discomfort on the user, since their hand is no longer as free as it would be in a normal scenario. The most common device used in the literature is the data glove, to acquire the flexion degree of each finger in the hand. Another system (8) bases its input device on EMG sensors placed on the forearm of an operator to control a robot with grasp recognition.

On most of the research work reviewed where a data glove was used the problem to be solved was the translation of sign language to written or spoken. As mentioned previously, most are based on a pre-established sign language to recognize a subset of letters on the alphabet in each case, commonly the static gesture letters. The work of Parvini *et.al* compares the result of three configurations for ANN classifying the static gestures of ASL acquired by a dataglove, one with neural networks alone, one with multi-layer networks working with template matching, and finally a gesture recognition system using bio-mechanical characteristics. The best results were obtained from the first and last configuration (9). In other papers, the system architecture is a bit more elaborate, like the case of Fels & Hinton, where they use two data gloves and a voice synthesizer with intonation control using a pedal. (10)

2. Architecture System Description

The proposed system for recognizing hand gestures begins and ends with the user. In between there are five stages: input, acquisition, preprocessing, classification and results presentation.

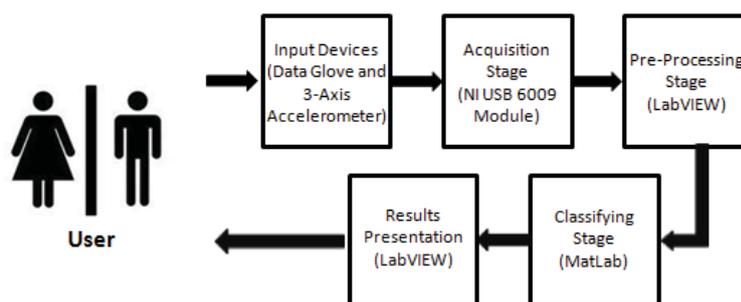


Figure 1. Diagram for the System Architecture.

As presented in Figure 1, the user places the input devices on his right hand; the input devices consist on the 5DT Data Glove 5 Ultra to gather information of the flexion degree of each finger of the hand and a 3-axis accelerometer place on the back of the hand to know its orientation. The data from the accelerometer goes through the acquisition stage, where an ADC converts the values of each axis to 14-bit numbers and passes them along to the computer for the preprocessing stage. In that stage, the information of the data glove, that communicates directly to the computer and the acquired data of the accelerometer go through a calibration and standardization process and then it is stored to use it as inputs in the classifying stage. Said stage is about using an artificial intelligence tool, such as artificial neural networks, to recognize gestures based on the experience of the previously collected data. With a trained network, an interface is developed to present the results of the gesture classification back to the users.

The proposed system innovates comparing to previous works by adding the accelerometer to gather information of wrist rotation, which allows identifying some ASL gestures for letters such as H and U that hold an identical position for all fingers in the hand and differ on the wrist rotation, another examples are K and P, also G and Q. Previous works, such as (9), select only one of the similar letters to input the system and be recognized by it. On figure 2, the similarities between the previously named letters are shown.

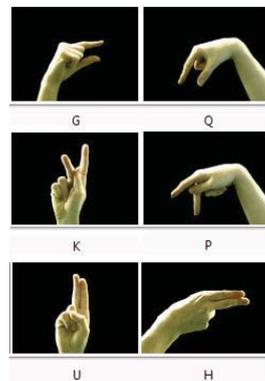


Figure 2. Similarities between ASL gestures for specific letters.

The final interface is made in National Instruments Labview® Software, as shown on Figure 3; the software is also used to gather the sensors information and store the data to be used in Matlab® for the ANN configuration and training.

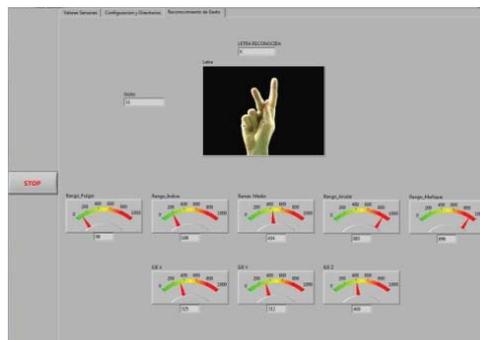


Figure 3. Interface in Labview® to present the final results of the Hand Gesture Recognition System

3. Classification Results

The Artificial Intelligence tool selected for this particular project was multilayer perceptron feed-forward neural networks with backpropagation as a training algorithm.

Several ANN were tried in order to obtain the best results in correctly classifying the data acquired and stored from the input devices. The total data used as input is 5 for each flexion sensor on the hand's fingers and 3 for each axis of the accelerometer. So the input layer of the system's network is set to 8. About the output layer, there are as many neurons as letters to classify and an

extra one for gestures that aren't letters, such as transition gestures or gestures not included in the ASL alphabet; the letters to be included in the process are the ones based on static gesture, excluding J and Z. Which leaves the output layer set to 25 neurons. Each input cluster will be feed-forwarded through the hidden layer and the output layer to activate one of the neurons in the output layer. The values of the output layer may vary from 0 to 1 with a sigmoid function.

The number of neurons in the hidden layer (H.L) was the parameter changed throughout the selection of the best performing network. The other parameters were kept the same in each test: 5300 training patterns, with an average of 200 per letter and 500 for transition gestures. A set of 1200 patterns were left apart for validation purposes. Each training process was made with a learning rate of 0.015, a maximum of 100 iterations and a mean squared error target of 1.10^{-5} . The results are shown on Table 1.

Table 1. Training Results and Validation Percentages for different ANN configurations, changing the number of neurons in Hidden Layer.

H.L Neurons	Training MSE	Training Iterations	Training Time	Val. Classification %
40	$5,25.10^{-3}$	61	18'16"	92,02
50	$4,16.10^{-3}$	83	33'36"	92,7
60	$4,10.10^{-3}$	62	34'23"	94,07
70	$3,71.10^{-3}$	64	47'21"	93,39

4. Conclusions

The performance of the Glove-Based Gesture Recognition System is directly related to the effectiveness of the classification stage within it. Based on the results given by different number of hidden layers in the neural network, the best validation percentage obtained was when there were 60 neurons in the hidden layer. Among the many benefits of using neural networks for classification duties, there is the ability to adapt and generalize when properly trained, and they can be easily retrained to adapt its synaptic weights to variations in the system.

References

1. **A. Mulder.** *Hand Gestures for HCI*. Simon Fraser University : NSERC Hand Centered Studies of Human Movement Project, 1996.
2. **Hewett & others.** *Curricula for Human-Computer Interaction*. Carnegie Mellon University : The Association for Computing Machinery Special Interest Group on Computer Human Interaction, 1992.
3. **J. Preece & others.** *Human Computer Interaction*. Wokingham, England : Addison-Wesley Publishing Company, 1994.

4. **B. Parton.** *Sign Language Recognition and Translation: a multidisciplinary approach from the field of Artificial Intelligence.* Journal of Deaf Studies and Deaf Education, Vol. 11, No. 11, págs. 94-101, 2006.
5. **S. Ricco, C. Tomasi.** *Fingerspelling Recognition through letter-to-letter transitions.* Ninth Asian Conference on Computer Vision, 2009
6. **L. Gui, J. Thiran, N. Paragios.** *Finger-spelling recognition within a collaborative segmentation/behavior inference framework.* 16th IEEE European Signal Processing Conference, 2008.
7. **A. Malima, E. Ozgur, M. Cetin.** *A fast algorithm for Vision-Based hand gesture recognition for robot control.* IEEE 14th Signal Processing and Communications Applications. págs. 1-4, 2006.
8. **S. Ferguson, G. Dunlop.** *Grasp Recognition from Myoelectric Signals.* Auckland : Proc. 2002 Australasian Conference on Robotic and Automation. págs. 83-87, 2002.
9. **F. Parvini & others.** *An approach to Glove-Based Gesture Recognition.* Lecture Notes in Computer Science. Proceedings of the 13th International Conference on Human-Computer Science Vol. 5611. págs. 236-245, 2009
10. **S. Fels, G. Hinton.** *Glove-Talk II - A neural-network interface which maps gestures to parallel formant speech synthesizer control.* IEEE Transactions on Neural Networks, Vol. 8, Issue 5, págs. 977-984, 1997.