
Journal of Computer Science and Technology
ISSN: 1666-6046
journal@lidi.info.edu.ar
Universidad Nacional de La Plata
Argentina

Ronchetti, Franco
Dynamic Gesture Recognition and its Application to Sign Language
Journal of Computer Science and Technology, vol. 17, núm. 02, octubre, 2017, pp. 154-155
Universidad Nacional de La Plata
La Plata, Argentina

# Thesis Overview:

## Dynamic Gesture Recognition and its Application to Sign Language

Franco Ronchetti
School of Computer Science, National University of La Plata, Argentina
PhD Thesis in Computer Science[1]
Advisors: Laura Lanzarini, Alejandro Rosete
{fronchetti, laulal}@lidi.info.unlp.edu.ar, rosete@ceis.cujae.edu.cu

The automatic recognition of human gestures is a complex multidisciplinary problem that has not yet been completely solved. Since the advent of digital video capture technologies, there have been attempts to recognize dynamic gestures for different purposes. In the recent years, new technologies such as depth sensors or high-resolution cameras were incorporated as well as the high processing capacity of the current devices emerged, allowing the new technologies development capable of detecting different movements and acting in real time. Unlike the recognition of the spoken voice, which has been researched for more than forty years, the topic of this thesis is relatively new in the scientific area and it evolves rapidly as new devices appear as well as new computer vision algorithms.

It is necessary to tackle many different tasks to be able to use an automatic sign language recognition system to translate the interpreter gestures. First, there are different approaches depending on the sensing device to use. Once the gesture is captured, several pre-processing stages are required to identify regions of interest such as the hands and face of the interpreter, and then identify the different trajectories of the performed gesture.

The sign language presents a huge variability in the different postures or configurations that a hand can have, which makes this discipline a particularly complex problem. To deal with this, a correct generation of the static and dynamic descriptors is necessary. In addition, because each region has specific language grammars, it is required the provision of an Argentine Sign Language (LSA) database, which has not been available yet. Based on the reasons mentioned above, this thesis aims to develop a complete process of interpretation and translation of the Argentinian Sign Language through videos obtained with an RGB camera.

First, a state of the art study about the gesture recognition was carried out. Intelligent techniques for image and video processing as well as the different descriptors types were researched. As a preliminary work, a strategy capable of processing human actions captured with an MS Kinect device [4] was developed. This strategy implements a probabilistic SOM neural network (ProbSOM) with a descriptor specifically designed to retain temporal information. This work allowed to overcome the existing results so far for two recognized databases.

As a result of this thesis, two main contributions in the sign language field were made. In the first place, a specific database for the recognition of the Argentinian Sign Language was developed. This included an image database with the 16 configurations most used in the language [3], along with a database of high-resolution videos with 64 different signs, with a total of 3200 videos [2]. These databases were recorded with 10 different interpreters and several repetitions, allowing their use with classic techniques of machine learning. In addition, in these databases, the interpreters have worn colored gloves in the form of a marker. This has been done to facilitate the segmentation task of the hands of the images/videos and thus to can advance with the rest of the classification stages. In this way, other researchers will be able to evaluate other recognition algorithms without the need to worry about this segmentation stage.

Secondly, two methods of sign classification were designed and implemented, which were satisfactorily evaluated in the databases. The first method is dedicated to the hand configurations classification (static gestures) [3]. Here a probabilistic clustering to correctly classify the 16 possible configurations of the database was used, achieving a simple and powerful recognizer. The second classification model allows the classification of segmented signs in videos [1]. This consists of a probabilistic system based on the information which has been captured from the two hands, where each one evaluates three main components: the position, the configuration

**Citation:** F. Ronchetti. "*Thesis Overview: Dynamic Gesture Recognition and its Application to Sign Language",* Journal of Computer Science & Technology, vol. 17, no. 2, pp. 154-155, 2017


[1]Full text available at http://sedici.unlp.edu.ar/handle/10915/59330

and the movement of the hands. This discrimination allows to have a modular system, with different sub-classifiers capable of being interchanged and independently evaluated. To obtain the adequate descriptors for these subsystems, it is necessary to perform a processing that involves the correct segmentation and tracking of the hand of the interpreter, the different configurations classification and a proper representation of the movement information. Figure 1 graphically summarizes the strategies used for the sign classification, including the stage of segmentation, stage of getting descriptors, sub-classifiers, and final classification.

In order to evaluate the developed models, several tests on the built databases were performed. First, cross-validation tests were carried out using a percentage of the data as training set and the rest of the data for testing set. Besides, an assessment of how robust the system is was made by incorporating new interpreters, unknowns for the model. Thus, 9 of the 10 individuals in the database were used as system input data, evaluating with the remaining individual. All these experiments showed excellent results, with an error rate of less than 5%. On the other hand, to evaluate the effectiveness of the implemented model, some of the sub-classifiers were changed by known techniques in the literature as Markov Models or Feed-Forward Neural Networks, showing solidity in the proposed strategies in this thesis.

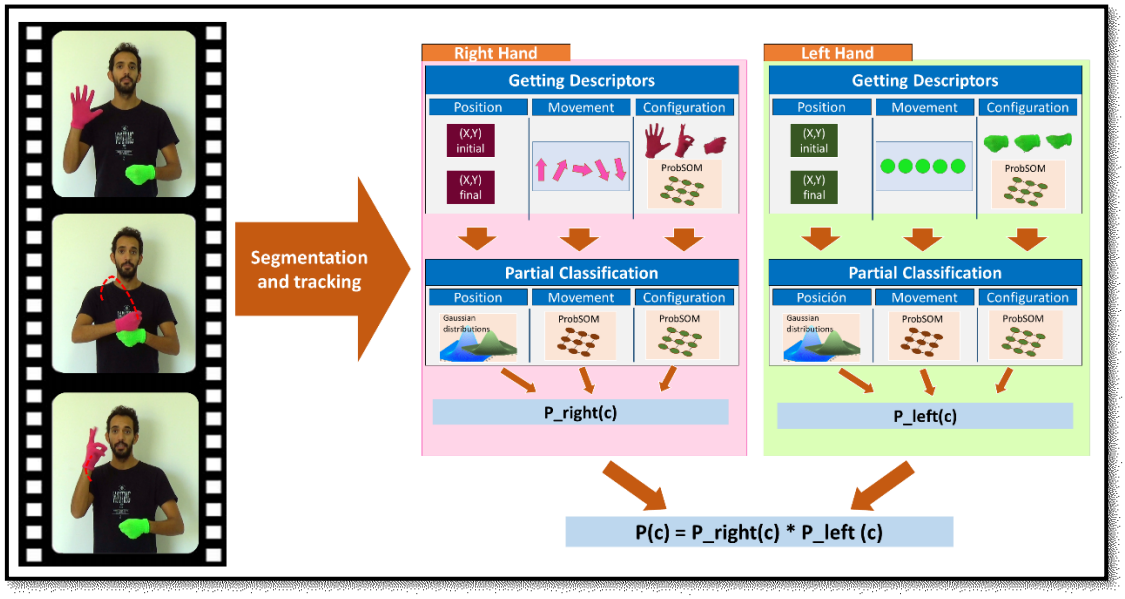


Figure 1. Classification model

## References

[1] F. Ronchetti, F. Quiroga, C. Estrebou, L. Lanzarini, A. Rosete. "*Sign Language Recognition without frame-sequencing constraints: A proof of concept on the Argentinian Sign Language"*. Advances in Artificial Intelligence - IBERAMIA 2016: 15th Ibero-American Conference on AI, San José, Costa Rica, November 23-25, 2016, Proceedings. pp338-349. Springer International Publishing. 2016.

[2] F. Ronchetti, F. Quiroga, C. Estrebou, L. Lanzarini, A. Rosete. "*LSA64: An Argentinian Sign Language Dataset*". XXII Congreso Argentino de Ciencias de la Computación. CACIC 2016. San Luis. Argentina. pp794-803. October 2016.

[3] F. Ronchetti, F. Quiroga, C. Estrebou, L. Lanzarini. "*Handshape recognition for Argentinian Sign Language using ProbSom*". Journal of Computer Science & Technology. ISSN 1666-6038. ISTEC – RedUNCI. Vol. 16, Num. 1, pp01-05. April 2016.

[4] F. Ronchetti, F. Quiroga, L. Lanzarini, C. Estrebou. "*Distribution of Action Movements (DAM): A Descriptor for Human Action Recognition*". Frontiers of Computer Science. ISSN 2095-2236. Springer, Higher Education Press. v9. pp956-965. December 2015.