Sign Language Recognition Model Combining Non-manual Markers and Handshapes

Luis Quesada[1,2(✉)], Gabriela Marín[1,2], and Luis A. Guerrero[1,2]

[1] Escuela de Ciencias de la Computación e Informática, Universidad de Costa Rica,
San Pedro, Costa Rica
{luis.quesada,gabriela.marin,luis.guerreroblanco}@ucr.ac.cr
[2] Centro de Investigaciones en Tecnologías de La Información y Comunicación,
Universidad de Costa Rica, San Pedro, Costa Rica

Abstract. People with disabilities have fewer opportunities. Technological developments should be used to help these people to have more opportunities. In this paper we present partial results of a research project which aims to help people with disabilities, specifically deaf and hard of hearing. We present a sign language recognition model. The model takes advantage of the natural user interfaces (NUI) and a classification algorithm (support vector machines). Moreover, we combine handshapes (signs) and non-manual markers (associated to emotions and face gestures) in the recognition process to enhance the sign language expressivity recognition. Additionally, non-manual markers representation is proposed. A model evaluation is also reported.

Keywords: Sign language recognition · Handshapes recognition · Non-manual markers recognition · Intel RealSense

1 Introduction

In our human computer interaction (HCI) research group, we believe that technological advances should help diminishing the gap between people living with and without disabilities. To bridge that gap, it is essential to provide inexpensive tools.

Our work integrates natural user interfaces (NUI) and automatic sign language recognition (SLR) algorithms. There are more than 100 sign languages worldwide which are spoke mostly by deaf people [1]. Some researchers have achieved SLR using recent technological devices, i.e. Microsoft Kinect and Leap Motion. However, many of these works are still under development or their contributions were the evaluation of techniques and algorithms used in the context of some sign languages.

Sign language recognition is a complex task because includes several characteristics [2]. Within a dialogue, any of the characteristics can change the phrase's meaning. By example, the interlocutor's eyebrows position (a non-manual marker) makes the difference between an informative sentence and an interrogative sentence.

NUI use 3D cameras, which by infrared technology and depth sensors, recognize handshapes, face configuration and body position allowing recognition in a given moment of time. The device used for experimentation was the Intel RealSense.

C.R. García et al. (Eds.): UCAmI 2016, Part I, LNCS 10069, pp. 400–405, 2016.
DOI: 10.1007/978-3-319-48746-5_41

The goal of this paper is to show a holistic model to achieve sign language recognition. This model recognizes sign language parameters: handshapes and non-manual markers using NUI and support vector machines (SVM). The main contribution of the paper is scalable SLR model using NUI (Intel RealSense). Also, this work aims to promote the development of tools to take advantage of new technologies for the benefit of the Deaf Community around the world.

In the next section we refer to sign language recognitions main components. In Sect. 3, the sign recognition model is proposed. In Sect. 4 the assessment system is presented. Finally, the discussion and the conclusions are presented in the final section.

2 Sign Language Recognition

This section describes sign languages main features: levels and parameters. Moreover, related work is described.

2.1 Sign Language Levels and Parameters

In Costa Rica, *Centro Nacional de Recursos para la Educación* (CENAREC) developed a project to describe the *Lengua de Señas Costarricense* (LESCO). LESCO description project 5 parameters: (1) handshape, (2) non-manual markers, (3) movement, (4) palm orientation and (5) location. These parameters described most of the sign languages [3].

This work considered only 2 parameters: handshapes and non-manual markers. The handshapes associate hand postures and concepts (words). Non-manual markers are changes in eyebrows position, facial expressions and head gestures uses to add grammatical information [4].

2.2 Related Work

Research related to sign language recognition include a variety of devices, i.e. 3D cameras, gloves and specialized hardware. This paper will focus its literature review only on the Microsoft Kinect and the Intel RealSense.

Microsoft Kinect has been tested recognizing small groups of signs, including ASL [5] and Japanese Sign Language (JSL) [6]. These works applied mixed recognition techniques with promising results. The techniques include Artificial Neural Networks (ANN), Hidden Markov Models (HMM) and SVM.

Intel RealSense has been involved in work in progress applications. Huang et al. [7] use this device to recognize fingerspelling alphabet. They predict the handshapes using SVM and Deep Neural Networks (DNN). Results are promising [7].

Moreover, automatic recognition systems for non-manual markers have been explored. Using multi-scale and spatial-temporal analysis, Liu et al. [8] proposed a system to recognize raised and lowered eyebrows, head nods, and head shakes [8].

There are few automatic SLR systems recognizing handshapes and non-manual markers at same time. These works are prior to the NUIs. Hence, they use conventional cameras [9]. NUI represents a new trend in the use of technologies to support HCI.

3 Sign Language Recognition Model

This section proposes a sign recognition model using the Intel RealSense and SVM. The device of Intel is a set of cameras and development libraries that allow users to use gestures to interact with computers. SVM predicts performed handshapes and non-manual markers. Figure 1 shows the proposed model. The combination handshape – non-manual markers enhance the SLR semantic.

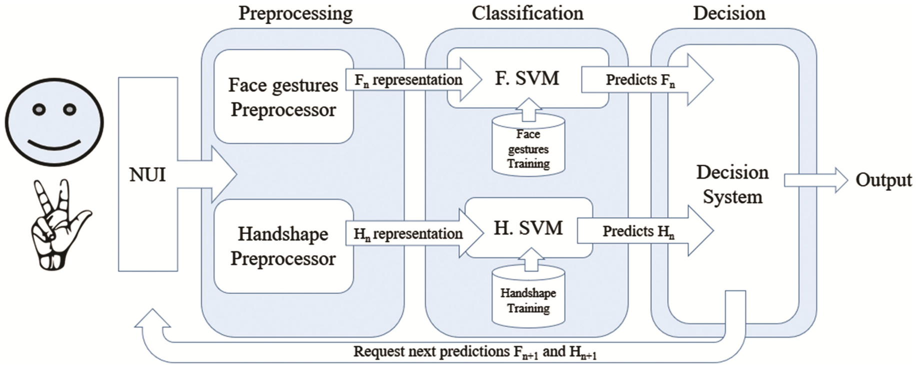


Fig. 1. SLR model.

Main components of the SLR model are 3D Camera (NUI), preprocessing module, classification module and decision module. This model assumes prior training of each SVM. Once the training is complete, the system is prepared to predict concepts performed by a signer using the trained sign language.

The NUI module includes hardware and software (libraries) provided by the device manufacturer. The hardware recognizes the position of each finger and the software provides the key points describing the performed sign. The hardware also recognizes the non-manual markers (eyebrows, eyes, etc.)

The preprocessing module transforms the raw data gathered from the device. The transformation allows the data be used by SVMs. The data correspond to non-manual markers and handshapes. Non-manual markers include key features: eyebrows configuration (position), gaze and mouth configuration. Each feature was represented by an integer number between 0 and 100.

Therefore, F representation flowing between preprocessing module and classification module (see Fig. 1) is a list of 7 integer numbers. These numbers represent a face gesture associated to an emotion or a discursive modifier. Different non-manual markers configurations enhance the handshape significance [10].

Handshapes representation contains fingertips coordinates <x, y, z> and direction vectors between phalanges. These data are gathered for each finger. This handshape representation allows the distinction between several [11]. Consequently, H representation (see Fig. 1) is a list of real numbers corresponding to fingertips coordinates and vectors between phalanges.

Classification module receives preprocessed data from previous module. Two SVMs (previously trained) predict a face gesture and a handshape. SVMs are supervised learning models that use training observations to recognize patterns [12].

Based on information provided by the SVMs, the decision module chooses which sign is performing in the device vision range. Every 6 milliseconds the SVM predict two values: a face gesture prediction and a handshape prediction.

The answer of the SLR cannot be based on only one prediction. In an instant of time, the signer could be moving the hand to achieve the final position. The intermediate handshapes predictions could not be the final one.

Decision system main components are: a face gestures buffer (capacity: 5 predictions), a handshape buffer (capacity: 5 predictions), two buffer handlers (one handler per buffer) and a rules engine. Each prediction is evaluated by the respective buffer handler. Circular buffers are used. When new predictions F_n and H_n get in the decision system, the handler assigns them to a buffer space. Index *n* represents a frame of time. Therefore, each SVM prediction represents the sign performed in different frames of time. Each frame is separated by milliseconds between them.

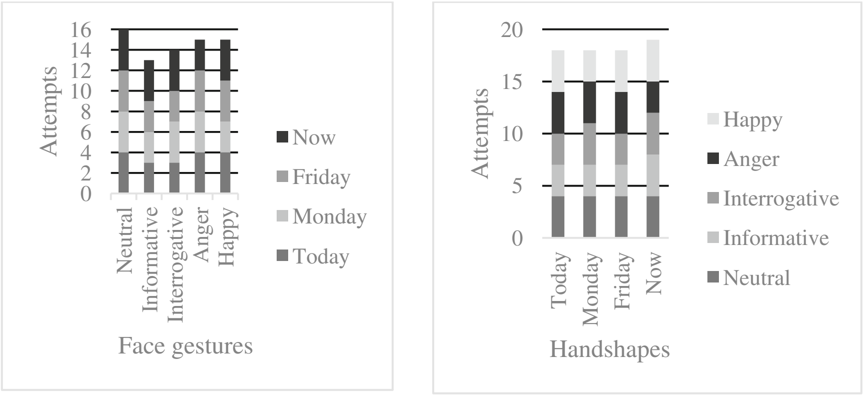
As soon as the 5 buffer spaces are equal (5 equal consecutive predictions), a recognized face gesture (F_i) or handshape (H_j) is send it to the rules engine. If a rule applied to the recognized face gesture and handshape, then the system outputs a recognized sign S_r. This solution was explored in [11] recognizing handshapes only.

4 Evaluation

We selected a list of signs (handshapes) and a list of face gestures. Rules were created for the handshapes and face gestures selected. Each face gesture is defined by non-manual markers configuration. Afterwards, the SVMs were trained. Each SVM was trained separately by one person. The trainer is acquainted with ASL and LESCO. Only one person was considered because the nature of the test. Each person express emotions differently. Further, physical features of each person are different (i.e. eyebrows position or neutral gaze).

When the SVMs were ready, they were tested individually. Then, we attempted to recognize simple signs according to the implemented rules. The handshapes selected were: “today”, “monday”, “friday”, and “now”. Face gestures selected were: “interrogative”, “informative”, “anger”, and “happy”. Handshapes are LESCO signs. Each sign is performed using one hand. Face gestures were selected from LESCO description [2] and basic emotions expressed in sign languages [13].

Twenty rules were defined combining handshapes and face gestures. Four performances by rule were requested. Therefore, 80 sign performances were tested (5 face gestures, 4 handshapes, 4 attempts per combination). The requested performances were randomly ordered. One person performed all the tests. Figure 2.a shows the results considering the attempts (performances in front of the NUI) per face gesture.



(a)

(b)

Fig. 2. (a) Accuracy results per face gesture, (b) Accuracy results per handshape.

Neutral face gesture achieved 100 % of gesture accuracy. Hence, all handshapes were recognized when the person expressed a neutral face. The worst recognition rate was 81.25 % (informative face gesture). While tests were performed, informative face gesture was misclassified as neutral face gesture; and anger gesture was misclassified as interrogative face.

No handshape was recognized with perfect accuracy. However, the signs were misclassified at most 2 times per handshape (total 20 attempts). Considering the 80 tests, 73 (91.25 %) were classified correctly. Only 3 of 7 misclassified signs do not match both features (face gesture and handshape). Remaining 4 misclassified signs do not match one of the features (see Fig. 2.b). Next section discusses the evaluation results.

5 Discussion and Conclusion

We presented a sign language recognition model. The model recognizes two important parameters of sign languages: handshape and non-manual markers. Although we choose a specific technology (Intel RealSense) and a specific classification algorithm (SVMs), these components could be substituted by different devices or algorithms. By example, an ANN can predict signs or a new NUI can gather data more efficiently.

Furthermore, additional parameters can improve the automatic sign language recognition model. Remaining parameters are: movement, palm orientation and location. Adding these parameters require: (1) to define a parameter representation, (2) to select and evaluate a classification algorithm and (3) to create rules including the new parameters.

Non-manual markers representation was tested. More face gestures must be defined to express more emotions associated to the sign languages specific grammar. Recognizing more face gestures require more non-manual markers recognition. Evaluation combining handshapes and face gestures recognition was executed. More than 90 % of the tests were classified correctly. As shown, selected parameters recognition adds semantic richness to signed dialogues.

We hope that this proposal will encourage research and development of new tools that help deaf people. People living with disabilities in developing countries need to take advantage of technological advances to achieve a better quality of life.

Acknowledgments. This work was partially supported by the Escuela de Ciencias de la Computación e Informática at Universidad de Costa Rica grant No. 320-B5-291, by Centro de Investigaciones en Tecnologías de la Información y Comunicación de la Universidad de Costa Rica, and Consejo Nacional para Investigaciones Científicas y Tecnológicas (CONICIT) of the Government of Costa Rica.

References

1. Lewis, P., Simons, G., Fennig, C.: Ethnologue: Languages of the World. SIL International (2009)
2. Oviedo, A.: Descripción General Básica de la LESCO (2012)
3. Woodward, J.: Sign language varieties in Costa Rica. Sign Lang. Stud. **73**(1), 329–345 (1991)
4. Caridakis, G., Asteriadis, S., Karpouzis, K.: Non-manual cues in automatic sign language recognition. Pers. Ubiquit. Comput. **18**(1), 37–46 (2014)
5. Zafrulla, Z., Brashear, H., Starner, T., Hamilton, H., Presti, P.: American sign language recognition with the kinect. In: Proceedings of the 13th International Conference Multimodal Interfaces, pp. 279–286 (2011)
6. Agarwal, A., Thakur, M.K.: Sign language recognition using Microsoft Kinect. In: Sixth International Conference on Contemporary Computing, pp. 181–185 (2013)
7. Huang, J., Zhou, W., Li, W., Li, H.: Sign language recognition using real-sense. In: EEE China Summit and International Conference on Signal and Information Processing (ChinaSIP), pp. 166–170 (2015)
8. Liu, J., Liu, B., Zhang, S., Yang, F., Yang, P., Metaxas, D.N., Neidle, C.: Non-manual grammatical marker recognition based on multi-scale, spatio-temporal analysis of head pose and facial expressions. Image Vis. Comput. **32**(10), 671–681 (2014)
9. Yang, H.D., Lee, S.W.: Robust sign language recognition by combining manual and non-manual features based on conditional random field and support vector machine. Pattern Recognit. Lett. **34**(16), 2051–2056 (2013)
10. Nguyen, T.D., Ranganath, S.: Facial expressions in American sign language: tracking and recognition. Pattern Recognit. **45**(5), 1877–1891 (2012)
11. Quesada, L., López, G., Guerrero, L.A.: Sign language recognition using leap motion. In: García-Chamizo, J.M., et al. (eds.) UCAml 2015. LNCS, vol. 9454, pp. 277–288. Springer, Heidelberg (2015)
12. Cortes, C., Vapnik, V.: Support-vector networks. Mach. Learn. **20**, 273–297 (1995)
13. Antonakos, E., Roussos, A., Zafeiriou, S.: A survey on mouth modeling and analysis for Sign Language recognition. In: International Conference on Automatic Face and Gesture Recognition, pp. 1–7 (2015)