Virtual Sign Translator

Paula Escudeiro[1], Nuno Escudeiro[2], Rosa Reis[3], Maciel Barbosa[4] José Bidarra[5], Ana Bela Baltazar[6], Bruno Gouveia[7]

[1,2,3,6,7] Departamento de Engenharia Informática,Instituto Superior de Engenharia do Porto, Porto, Portugal
[4] Departamento de Engenharia Electrotécnica, Faculdade de Engenharia da Universidade do Porto, Porto, Portugal
[5]Departamento de Ciências Exactas e Tecnologia, Universidade Aberta, Lisboa Portugal
{pmo & nfe & rmr}@isep.ipp.pt; fmb@fe.up.pt; bidarra@univ.ab.pt; altazaranabela@sapo.pt;
brunogouveiarodrigues@gmail.com

Abstract - In this paper we present the overall study that includes the model developed (VS Model) and the experiences performed, with an automatic bidirectional sign language translator, between written and sign language, which is being supervised by the research group GILT (Graphics, interaction & learning technologies) under the frame of a national project called Virtual Sign (VS project). This project aims to develop and evaluate a model that facilitates access for the deaf and hearing impaired to digital content - in particular the educational content and learning objects - creating the conditions for greater social inclusion of deaf and hearing impaired people. Access to digital content will be supported by an automatic translator between Portuguese Writing (LEP) and Portuguese Sign Language (LGP) supported by an interaction model.

Index Term - Access for deaf to digital content, Bidirectional translator, Learning objects, Interaction model, Sign Language

1. Introduction

Disabled people face non-ending difficulties when they want to deal with the new technologies: the use of a computer, the access to Information, editing and printing a text, etc. Reading a document can be an extremely complex task, despite their simplicity for normal user [1]. Today, the extraordinary progress of the new technologies, bound to the data processing and the Internet, offers remarkable opportunities to bring a better quality of life to those who endure handicap and disabilities [2]. The target community addressed by the project has its own language, known as Sign language [3]. A sign language is a language that uses manual communication instead of sound to convey meaning - simultaneously combining hand shapes, orientation and movement of the hands, arms or body, and facial expressions to fluidly express a speaker's thoughts. The sign language remains nevertheless a fully-fledged language, with its own constructional method of the sentences. The Portuguese sign language (LGP) faces a low diffusion level among the deaf as well as the hearing communities. To be effective, the communication of a deaf person requires the knowledge of LGP, not only on the speaker's side, but also on the listener's. It is also essential that both speak the same sign language. Due to the shortage of information or lack of availability of the sign language, it is very important to improve it, allowing this group to have access to information in their first language, i.e., the sign language.

2. Principles of the Sign Language

Languages can be oral-auditives when written representations are not used but, instead, all communication is oral. As it happens with the Portuguese Language and all the oral languages. They can also be visual-spatial, with a realization almost natural of signs and visual reception. In this last case we could refer several sign languages used by many different deaf communities spread around the world. These, such as oral languages, have their own grammar that allows distinguishing between sign languages and oral languages.

The interest in the Portuguese Sign Language has been showing a remarkable growth, not only by the deaf community that today represents near 150.000 persons in Portugal but also by all the involving community, like family, educators, professors, etc. However, in Portugal there is not much done yet in this field to assist the deaf in their daily lives.

The sign language is performed in a three-dimensional space, therefore the support from the new technologies is indispensable because it is possible to reach to the total perception of the sign, including the movement, the hands location and orientation, the configuration and also facial and body expressions. The richness of this tridimensional language is not limited itself on the simple technical realization of the sign. It also involves a total dynamic of the communication that is natural to the human being. The hands configuration, location and orientation are as important as the facial and body expressions that follow the realization of the signs. These two aspects could be crucial to distinguish very similar signs.

Similarly to the oral language, sign language has a lexical, a “phonetic” (instead of articulated sounds it has articulated signs), a “phonology” (instead of phonemes, it has elements from different natures that accomplish the same differential function from the words visual form), a syntax, a semantic and a pragmatic of its own. Being characteristic from each country and culture, and not universal, allows describing all the reality that involves us, what we see, feel or think.

Considering all the basic principles of the Sign Language expressed above and the deaf community needs regarding the access to written sign language interpretation, we realize all the interest in the development of the Virtual Sign project. Its outcomes will assist the deaf community in accessing the written information throw Portuguese Sign Language and vice versa thus, contributing to reduce info-exclusion of disabled persons.

3. Automatic Sign Language recognition

Sign Language communication uses two-handed gestures and non-manual signals.

One of the biggest difficulties in recognizing sign language is that there are signs which involve global body movement while others involve only the configuration and orientation of the fingers of one hand. The sources of information that need to be acquired to decode sign language depend on the sign itself and are not always the same. The level of detail is also heterogeneous ranging from wide arm movements to confined finger configurations. This imposes conflicting requirements on the field of view; it must be large enough to capture the global motion, but at the same time, small local movements must not be lost. Moreover, both hands often touch or occlude each other when observed from a single viewpoint and, in some signs, the hands partially occlude the face. Occlusion handling is also an important consideration. Sign Language recognition includes tracking of the hands, face and body parts, feature extraction, modeling and recognition of time-varying signals, multimodal integration of information, etc. The use of various sensors for the wide range of features seems therefore inevitable.

A. The use of depth sensors

The development of a motion-sensing input device by Microsoft, the Kinect depth sensor, was a revolution in technology akin to those that shaped the most fundamental breakthroughs of the 20th Century. While this development may seem wide-ranging and diverse, it can be summarized simply: for the first time, computers can see [2].

The depth image outputted from a Kinect sensor is so important because it is much easier for a computer to interpret than a conventional color image. When processing a flat 2D image, pixels with similar colors that are near to each other might erroneously be assumed to belong to the same object. If you have 3D information then pixels that correspond to locations physically near to each other tend to belong to the same object, irrespective of their color. It has often been said that pattern recognition has been made artificially difficult because most systems rely on 2D data[3][4].

The use of depth sensors has been proven to be a more robust tracking method than vision based methods. Another big advantage of the Microsoft Kinect SDK is that it returns the position and orientations of the hand easily with a good accuracy.

B. Depth-based hand localization and tracking

To acquire hand gesture data we are using direct-measures outputted from a data glove with 14 sensors, (5DT Data Glove 5 Ultra). [finger abduction sensor have proven to be not accurate enough for joint angle measurement]]] The 5DT Data Glove SDK already provides basic gesture recognition of 15 hand gestures. [5].

C. Future of gesture recognition

The future of data acquisition of hand gestures is pointing to NUI (Natural User Interfaces) namely using depth sensors such as Kinect. The recently introduced Leap Motion Sensor should be an option for further development.

For the first part of the project, we focused on the letters of the alphabet, in which the amplitude of hand motion is very small. Alphabet recognition depends mainly on fingers configuration and orientation of the hand in a static position.

For vision-based approaches, the hand is generally restricted to palm facing the camera, against a uniform background.

D. Building the avatar hand coding the gestures

We are aiming to develop a bi-directional translator for Portuguese Sign Language. This requires, besides the recognition of signs based on body motion and their translation to written Portuguese, a way to simulate body motion representing written Portuguese. This direction, from written Portuguese to Portuguese Sign Language, will be assured by an avatar that will simulate a user using sign language. This avatar may be built on the common approach that hand-codes the categories of hand shape, hand orientation, hand location and movement type that make up each sign in the vocabulary, forming a lexicon of sign definitions. Classifying the sign label from component-level results is then performed by comparing the ideal lexicon categories with the corresponding recognized components.

E. Analysis of Non-manual Signals

In recent software like FaceReader [14], facial recognition is limited to the six basic emotions [14]: happiness, sadness, surprise, fear, anger, disgust and the neutral expression which is considered to be constrained. The Faceshift software [14] uses a depth sensor and is able to map with great precision a human face which allows for accurate feature extraction and the immediate animation of an avatar.

Other body movements and postures involved in non-manual signals generally consist of torso motion, such as, body leaning forwards/backwards or turning to the sides. This features can also be easily recognized with a depth sensor like Kinect.

New devices for gesture recognition are showing up in the market, namely Omek Gasp. Omek Gasp is a software able to track the hand gesture which includes responsive and accurate tracking of both hands and fingers, even with occlusions. A full 3D model of the hands and, most importantly, a Gesture Authoring Tool (GAT) allow us to quickly and easily generate custom gestures [8]. Another important development is the Leap Motion sensor. Leap Motion is a depth sensor specialized in tracking every feature of the hands [9].

4. VS Model

The VS project addresses the lack of adaptability of most educational organizations to deaf and hard of hearing people. This lack of adaptability generates adverse conditions to deaf and hard of hearing people, preventing them to have the same opportunities in education as the other citizens.

The main results of this project are:

- a model, that allows the deaf and hard of hearing people to improve their integration into mainstream education [5], as shown in Fig. 1.
- an animated virtual character, to be integrated in educational software, that translates text to sign language.
- a virtual reality environment to translate Portuguese sign language to Portuguese text.

As an engineering school, we intend to take the first steps for the integration of these individuals by developing a model to assist the translation of the educational content of the different course units of the undergraduate computer science degrees to the Portuguese Sign language.

This model may be applied in different fields, such as virtual museums, web pages, services, so on [4].

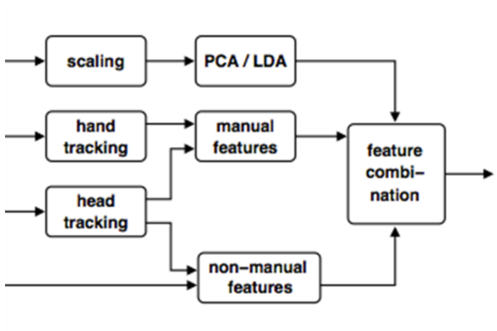

Fig. 1 Phases in the Creation of a Gesture.

5. Evaluation

To evaluate the performance of gesture recognition based on the input provided by both the Data Glove and Kinect we have collected data from seven different users. Each user, wearing the Data Glove and being tracked by Kinect skeletal tracking performed the Portuguese sign language symbols representing the alphabet and the ordinals from 1 to 9 in a total of 35 symbols. Each symbol was repeated 10 times. In total we have captured 2450 snapshots (seven users times 35 symbols times 10 repetitions). Since these symbols are represented only by the right hand and the right harm configurations we have extracted the 26 relevant features from all the sensors available through the Data Glove and Kinect. These features correspond to the 14 glove sensors plus 12 spatial coordinates provided by Kinect with regards to the right shoulder, right elbow, right wrist and right hand (x, y and z coordinates for each). This dataset was then used to evaluate the accuracy of an SVM classifier [13] using radial basis function (RBF) kernels. Error rates were estimated using 10-fold cross validation.

The performance of the classifier was evaluated on two distinct feature sets. We have started to evaluate the accuracy of the classifier based only on the input from the Data Glove. The estimated error in such a setting is 0.02. Then we have used all the data, including the input from both the Data Glove and the Kinect skeletal tracking (right arm). The error rate dropped to 0.01. The difference of the mean error observed in these two situations is statistically significant. The t-test for equal means that was performed yielded a p-value of 0.01344.

6. Conclusions

The selection of this target population arises due to the growing number of students with special needs who complete the elementary and high school and come to higher education. This situation demands for new means that allow these individuals to have easy access to educational content. Higher education institutions must host and provide appropriate conditions for students to get their degrees despite their disabilities. These institutions play an important role in the inclusive education, i.e., in the implementation of the educational system that includes everyone truly, (Brazil / MEC / SEESP, 1998).

As university teachers, we want to develop a methodology that allows the translation of textual educational content to sign language, and the other way round. This serves as a launching pad for the development of applications targeted for individuals with special needs. The first developments, focused on the use of a data glove and Kinect to recognize the Portuguese alphabet in a static setting using automatic classification are promising. We are now moving to the next step which is to recognize the signs in a dynamic setting.

7. Acknowledgment

The author would like to thank the Engineering Institute of Oporto and GILT (Graphics, Interaction and Learning Technologies) and FCT (Fundação para a ciência e Tecnologia)for making possible this work

8. References

[1] Reyes, I. 2005 “Comunicar a través del silencio :las posibilidades de la lengua de signos
[2] Herrero-Blanco, A., Salazar-Garcia, V. 2005 “Non-verbal predicability and copula support rule in Spanish Sign Language”. In: de Groot, Casper / Hengeveld, Kees (eds): Morphosyntactic.
[3] Annelies Braffort,2008 “ TRAITEMENT AUTOMATIQUE DE LA LANGUE DES SIGNES.
[4] VSigns v0.1a running on dbX v1.00 (c) 2004, Informatics & Telematics Institute HANIM models (c) Christian Babski (LIG/EPFL).
[5] SignWriting: um sistema de escrita para linguas de sinais. EFETA – Publicação Pastoral do Surdo, ano 5, n 32. Set/out. 1998
[6] Amaral, M. A., Coutinho, A., & Martins, M. R. (1994). Para Uma Gramática da Língua Gestual Portuguesa. Lisboa: Caminho.
[7] Baltazar, A.B. (2010). Dicionário da Língua Gestual Portuguesa. Porto: Porto Editora.
[8] Sylvie C.W. Ong and Surendra Ranganath, “Automatic Sign Language Analysis: A Survey and the Future beyond Lexical Meaning“ ,IEEE Transaction Pattern Analysis and Machine Intelligence, 27(6): 873-891, June 2005.
[9] Primesene, PrimeSense™ NITE Algorithms, Release 1.5. PrimeSense, 2011
[10] S. Kean; J. Hall; P. Kerry, Meet the Kinect: An Introduction to Programming Natural User Interfaces, Technology In Action, 2011.
[11] D. Catue, Programming with the Kinect for Windows: Software Development Kit, Microsoft, 2012
[12] 5DT, 5DT Data Glove Ultra SeriesUser's Manual, January 2011
[13] Campbell, C., Ying, Y., Learning with Support Vector Machines, Morgan & Claypool Publishers, 2011
[14] Noldus Information Technology bv Wageningen, The Netherlands