Towards Sign language recognition system in Human-Computer interactions

Maher JEBALI[#1], Patrice DALLE[*2], Mohamed JEMNI[#3]

#Research Lab. LaTICE - ESSTT
University of Tunis - Tunisia
[1]maher.jbeli@gmail.com
[3]mohamed.jemni@fst.rnu.tn

*Research Lab. IRIT
Univ. of Toulouse3 - France
[2]patrice.dalle@irit.fr

Abstract— **With the efforts to improve the use of Human-Computer-Interaction, there has been an important interest in trying to integrate human gestures into human–computer interface. This paper presents a modelling of Sign language recognition system, which is summarized in a dialogue between deaf people and signing avatar. With this modelling, the system can be configurable: we can keep the general modelling and only we change the scenario and the vocabulary. We have included to these modelling two important elements, which are context and prediction, to improve the reliability of sign language recognition system compared to the classic systems, which don't use semantic concept.**

***Keywords*— French sign language recognition, HCI, avatar**

I. INTRODUCTION

In several disciplines, many researchers have been interested in the field of gesture based Human-Computer interaction (HCI) and gesture recognition. Among these disciplines we quote computer vision, natural language processing, pattern recognition, HCI and linguistics. This multi-disciplinary research field can develop useful applications such as robotics control, emotion analysis, psychological behaviour and sign language recognition.

Incessantly, HCI is defining new communication modalities and new machines interacting ways.

Gesture can transmit information for which other aspects are not suitable or efficient. In spontaneous interaction, gestures can be used as a single modality or in combination with multi-modal interaction programs involving textual media, speech or facial expression. Sign language constitutes the multi-aspects interactions where different manual or non-manual components may occur simultaneously.

Most French deaf people do not understand perfectly French language, which explain the difficulties that they encounter to communicate with the computer and new technologies.

Sign language recognition is an application area of HCI, which tends to facilitate interaction between deaf person and technologies.

The grammars of sign languages are as highly complex as the grammars of voice languages and share with them many universal features, despite the difference in modality between sign languages (which use the visual channel) and voice languages (which use the auditory channel). Yet, sign languages also differ from spoken languages in radical ways: morphological information in SL is often conveyed simultaneously and from the beginning of signing; moreover, certain aspects of their phonological, syntactic and semantic structures are not commonly found in voice languages. These differences raise an interesting challenge both for including context and handle large vocabulary.

This paper is structured as follows. Section 2 gives an overview of related works in this study. Section 3 gives the modelling system and the benefits of prediction. Finally, we present the conclusion and some perspectives.

II. RELATED WORKS

The automatic recognition of sign language is almost 20 years behind speech recognition for multiple reasons.

Classification and processing of one-dimensional audio signals are easier than two-dimensional video signals. Also, sign language processing is by far not completely explored yet. Understanding sign language requires better linguistic knowledge, but until now there is no general rules that define the signing from a linguistic point of view.

The first scientific publications in the field of sign language recognition has become in the beginning of the 90s. Most applications presented in previous works don't operate in real-time and need up to 20 seconds after the sign production to complete the processing. There is a rarely published work, which gives details on camera hardware and resolution, suggesting that professional hardware, optimal camera placement, low noise and high resolution were used.

The data acquisition method constitutes the first feature, which classifies the different works. The most simple, exact and reliable techniques are intrusive. Put magnetic or optical markers on hands

and face facilitate the determination of manual configuration and facial expression. However, this is restrictive and unnatural for the user. Furthermore, data gloves, which measure the flexion of the finger joints, are undesirable for practical systems due to their high cost.
Furthermore, most existing applications do not exploit non-manual features [4].
Many work deal only the notion of signer-dependent where every signer is required to train the system before being able to use it.
The use of the notion of signer-independent requires a suitable features normalization from the first step of processing to rid of features dependencies on the distance of the signer's from the camera, his position in the image and other morphological rules.
Many researchers are focused on isolated signs like the speech recognition in their early days. Some existing systems process continuous production of signs but their vocabulary is not large. To improve the recognition rate, the exploitation of grammar and context is necessary.
The described system's feature and the several important works are listed in Table 1. In contrast to speech recognition, we cannot compare the indicated performances, due to the absence of a standardized benchmark for sign language recognition.

TABLE I

CLASSIFIER CHARACTERISTICS FOR SIGN LANGUAGE RECOGNITION

Author	Features	Interface	Vocab	Language Level	Recog. rate in %
[3]	M	Optical markers	22	Word	95.5
[8]	M	Video	40	Word	98.1
[2]	M	Data glove	203	S	92.1
[6]	M	Video	40	S	97.8
[7]	M	Video	22	S	91.8
[4]	M	Video	39	S	92.0
[1]	M	Video	164	SB	74.3
[5]	M	Video	961	Word	82.0
M : Manual, S: Sentence, SB: Subunits					

All recognition rates are valid only for the test examples. Also, we observed that when the vocabulary size increases, the recognition rate decreases sharply and becomes insufficient.
In summary, we can judge that the existing systems not meet the requirements for a robust real word system. In the following sections, we describe a framework for experimentation that takes into account the context and it more closely matches the real world.

III. MODELLING INTERACTION ANALYSIS

The interpreters are capable to convert voice or text to sign, despite the difficult conditions (simultaneity and the non-equivalence between source language and target language (omissions, additions, substitutions…) because they exploit the context and semantics. On the other side, existing systems have poor performance because they are not interested by the context and semantics.
We have chosen to deal with situations of dialogue led by the system because they can exploit the sense and control the context. (illustrated in figure 1).

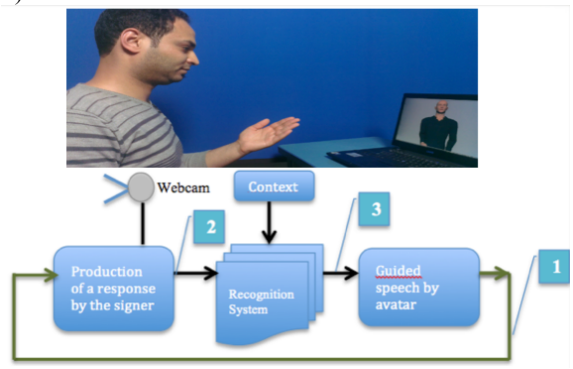

Fig. 1: Genaral schema of dialogue (recognition system is controlled by context)

We model the scenarios of a diaogue that can occur between a real interocutor and other virtual, knowing that this dialogue is driven by the avatar. The different stages of the dialogue are presented in the following subsection.

A. Description of the dialogue algorithm

1- In the beginning, Avatar lances a message of welcome
2- Avatar begins the scenario :
- Explanation of the rules (in this case, the avatar is the master)
- asks the interlocutor to explain his request

2 (a) - if the interlocutor has understood the request and he haven't the answer (hesitation, long inactivity, sign ...), the system (avatar) intervenes and reformulates the request.
2 (b) - if the interlocutor has understood the demand, he responds to the request.
3- The system analyzes the signs produced by the interlocutor
3 (a) - If the system do not recognize a sign or all the statement, it generates a message in avatar language, thereafter it passes the information to the interlocutor. The interlocutor produces again the request
3 (b) - If the system recognize the request, it generates a message in avatar language, thereafter it passes the information to the interlocutor
4- whether there are any other iteration, we repeat the same process from 2, otherwise we end the dialogue.

B. Exploitation

In all language analysis, context is very important. It shows why some sign or word is used in a certain situation.

We can not talk about the optimality of sign language recognition system without adding the context as an input to the system, because the same concept may appear in a variety of contexts and its appearance can be very different depending on these context.
In French sign language, there are similarities between several signs but with different meaning, thus adding the context in each step of the dialogue allows us to refine the reliability of the recognition system.

C. Interest of dialogue led by avatar

In addition to the context, prediction is one of the most essential issues that need to be explored for sign language recognition.
Such as in human behaviour, it is possible – in sign language recognition system – to predict the future outcome, rather than to simply provide backward-looking data about past interactions and to do these predictions in real-time.
The wide use of signs characterization data, whether on 2D or 3D data, enable sometimes to improve the recognition rate. But these approaches require huge processing time, since it attempts to define signs detection based methods.
Accurate prediction of location information is also crucial in processing location-dependent queries.

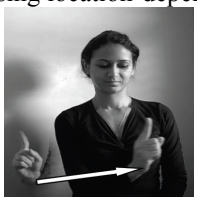

Fig. 2 Movement prediction

Each dialog step is controlled by the system in a known context, which is used to remove the ambiguities of similarity between the signs and to predict certain sign characteristics (hand location, head orientation…). Processing with prediction is easier than the bottom-up processing (segmentation, tracking, characterization, ...), because if there are errors in one of the different stages, the rest will be false. Also, measurements are more simple and easy to check with the predction processing.
For example, figure 2, shows how we can predict the hand position according to speed and movement of the hand. Also, we can check if the global shape of the hand has changed or not, instead of segmenting different fingers, which is costly in terms of time.
Dialogue is a particular case of Human-Machine Interactions where it is controlled by the machine. So, there is expectations on the response and rightly to exploit in order to make a feasible system.
In each interaction in dialogue, we predicted all condidats signs to be recognized. In the following phase (recognition), we are concerned with verifying the subsets characteristics which identifies a sign relative to another (table 2).

TABLE II

VERIFICATION OF SUBSET CHARACTERISTICS BASED ON PREDICTION

Sign	**Charac1**	**Charac2**	**Charac3**	**Charac4**	**…**	**CharacN**
S1	X	X	X	X		X
S2	X	X		X		X
S3		X	X	X		X

IV. CONCLUSIONS

Gestural interfaces can help deaf people to have more natural communication with computer. In this scope, we showed in the first time the main problematic of sign language recognition towards real world application.
After that, we detailed a modelling of HCI as a dialogue between a deaf person and signing avatar. This modelling is constrained by context concept and the prediction concept was proposed to handle the problem of large vocabulary complexity.
One of the most useful approaches for SL recognition is to use HMMs, a powerfull generative model. However, the observations of these generative models are conditionally independent, which allows us to focus may be on the discriminant models.

REFERENCES

[1] Cooper HM, Bowden R. Large Lexicon Detection of Sign Language. IEEE Workshop Human Computer Interaction, ICCV, Rio Brazil 07, LNCS 4796, Springer Verlag. pp88-97, 2007.

[2] Fang, G., Gao, W., Chen, X., Wang, C., Ma, J. Signer-independent continuous sign language recognition based on SRN/HMM. In: Revised Papers from the International Gesture Workshop on Gestures and Sign Languages in Human–Computer Interaction, pp. 76–85. Springer, Heidelberg, 2002.

[3] Holden, E.J., Owens, R.A. Visual sign language recognition. In: Proceedings of the 10th International Workshop on Theoretical Foundations of Computer Vision, pp. 270–288. Springer, Heidelberg, 2001.

[4] Parashar, A.S. Representation and interpretation of manual and non-manual information for automated American sign language recognition. Ph.D. Thesis, Department of Computer Science and Engineering, College of Engineering, University of South Florida, 2003.

[5] Pitsikalis V, Theodorakis S, Vogler C, Maragos P. Advances in Phonetics- based Sub-Unit Modeling for Transcription Alignment and Sign Language Recog-nition. IEEE CVPR Workshop on Gesture Recognition, Colorado Springs, USA, 2011.

[6] Starner, T., Weaver, J., Pentland, A. Real-time American sign language recognition using desk and wearable computer based video. IEEE Trans. Pattern Anal. Mach. Intell. 20(12), 1371– 1375, 1998.

[7] Vogler, C., Metaxas, D. Parallel hidden Markov models for American sign language recognition. In: Proceedings of the International Conference on Computer Vision, 1999.

[8] Yang, M., Ahuja, N., Tabb, M. Extraction of 2D motion trajectories and its application to hand gesture recognition.

IEEE Trans. Pattern Anal. Mach. Intell. 24, 1061–1074, 2002.