

TOWARD AN ANNOTATION SOFTWARE FOR VIDEO OF SIGN LANGUAGE, INCLUDING IMAGE PROCESSING TOOLS AND SIGNING SPACE MODELLING

A Braffort¹, A. Choisier¹, C. Collet¹, P. Dalle², F. Gianni², F. Lenseigne², J. Segouat¹

¹LIMSI/CNRS – BP 133 – F-91 403 Orsay cedex – France

²IRIT-UPS – 118 route de Narbonne – F-31 062 Toulouse cedex 4 – France

¹{braffort, choisier, collet, segouat}@limsi.fr – ²{dalle, gianni, lenseigne}@irit.fr

Abstract

Several French national projects have been achieved last years, allowing linguist and computer scientists working on French Sign Language (FSL) to establish a permanent collaboration.

During these projects, several video FSL corpora have been realised. One of the research orientations induced by these projects relates to multi-disciplinary annotation.

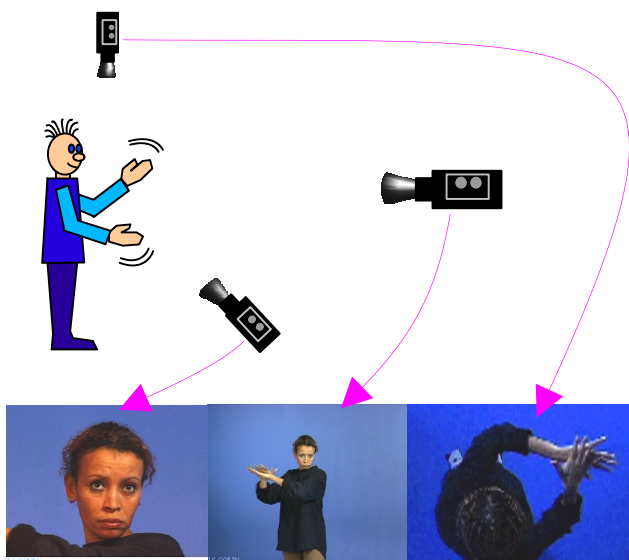
From the computer scientist's side, the aim is to develop an annotation software integrating different kind of tools based on image processing and 3d modelling that will be used to automatically annotate both lexical and syntactic information.

This article presents several of these components. A first version of the annotation software integrating these components is under development.

1. INTRODUCTION

The study presented here has been initiated during one of the different French national multi-disciplinary projects dedicated to French Sign Language (FSL) that have been achieved last years, where both linguists and computer scientists were involved.

One of them is named LS-COLIN (Braffort et al., 2001). During this project, a video database of FSL was build, with the double aim to provide data for linguists who want to highlight the iconicity of the FSL, and to provide good quality pictures for computer scientists working on image processing (see Figure 1). In the same way than Quek and Mc Neill (1999), linguists and computer scientists are working together on the same video in order to perform complementary analysis.



Close-up shot , frontal shot and upper view

Figure 1: LS-COLIN corpus; the three shots.

One of the research orientations induced by this project relates to annotation. The annotation is seen, there too, according to multi-disciplinary point of views:

- From the linguistic point of view, the structure and the granularity of the annotated data are studied, with new representation possibilities.
- From the data-processing point of view, our aim is to study processing components based on multidisciplinary models (vision, bio-mechanics, linguistics...).

For that, we are developing together a new annotation software dedicated to video Sign Language that will integrate different kind of processing components, which are not proposed for the moment in annotation software like Elan (Wittenburg, 2002) or ANVIL (Kipp, 2001).

These tools are based on image processing as in SignStream 3 project (Neidle, 2003), but also on 3d modelling. They will be used to automatically annotate both lexical and syntactic information.

This objective is approached in an iterative way. At this moment, we are designing various specialized components. Each component is evaluated independently of the others. When a satisfying level of robustness is reach, the component is integrated within an annotation tool. The principal users of this platform will be linguists, and the components must be simple to use and reliable.

This article presents the design of the image processing components (section 2) and the design of a component which aim is to represent the spatial context of the utterances (section 3).

2. IMAGE PROCESSING COMPONENTS

We have studied information that could be extracted from the video sequence by image processing. Four families of treatment were identified: extraction of information, tracking of parts of the body, reconstruction of 3d information, characterization, and recognition. Several components were implemented, either by adapting known techniques, or by developing original methods. They are listed in the following.

Extraction of areas of interest:

- Extraction of the signer, by comparison with an reference image of the background, filtering and morphological processing (see Figure 2);



Figure 2: Extraction of the signer

- Extraction of the entities hands and head, from a signature of the characteristic colour of the skin, and identification based on criteria of size and shape (see Figure 3).



Figure 3: Extraction of the hands and head

Tracking:

- Labelling and tracking of the entities in the image sequence, by overlapping of bounding boxes;
- Tracking of characteristic patterns: tracking of the index tip (see Figure 4), using the Radial Cumulative Similarity method (Darrel, 1998), tracking of the face (see Figure 5), using the Mean Shift method (Comaniciu, 2000).

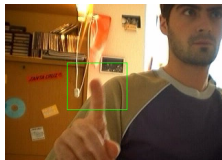


Figure 4: Index tracking

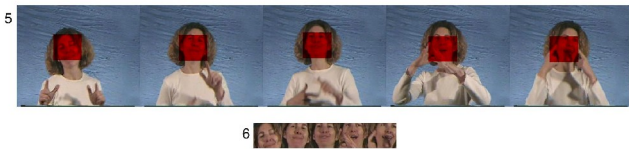


Figure 5: Face tracking

Computation of characteristics:

- Synthetic visualization of the arm movements (see Figure 6), using the Time Motion History Image method (Davis, 1997).

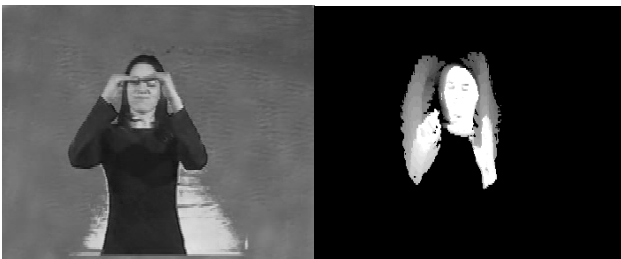


Figure 6: Arm movements

Reconstruction:

- Determination of the arm posture (see Figure 7) from a single calibrated camera (Lenseigne, 2004).

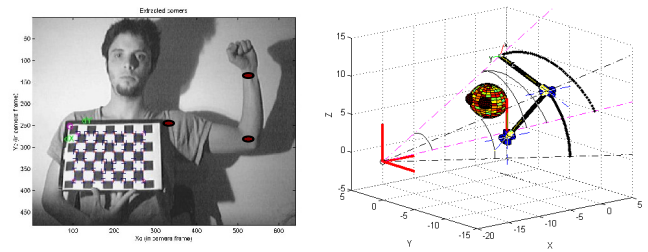


Figure 7: Determination of the arm posture

- Study of the zones pointed by the signer in the signing space, illustrated in Figures 8 and 9 (see definition in the following section) (Risser, 02).



Figure 8: Pointing direction computation

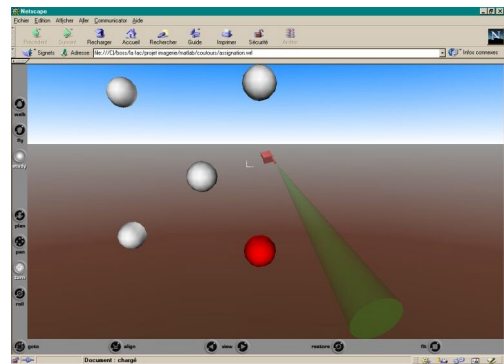


Figure 9: Visualisation of the pointing direction

Interpretation:

- Study of a formalism and a software architecture for the interpretation of the facial expressions in SL (Mercier, 2003)
- Study of an interpretation system of video sequences of SL (Dalle, 2001): Definition of the multilevel architecture of an interactive system of concept construction of SL

These various components will be used within the annotation software in order to detect automatically some visual information, on the position, the direction or the speed of the articulators, and also higher-level information, such as primitives of movement, face expressions.

Image processing tools can also provide information on more high-level information, such as the spatial references used in the discourse. Annotation of spatial information is

particularly interesting for our linguistic colleagues working on FSL syntactic structures (Cuxac 2000; Sallandre & Cuxac 2002; Sallandre 2003). This is why we plan also to include 3d tools in our annotation software allowing us to represent the signing space.

3. SIGNING SPACE MODELLING

In addition to the annotation of traditional information (parameters, lexicon, syntax and semantics), we want to annotate information on the spatial context of SL utterances. Indeed, the SL are articulated in a space located in front of the signer, and named the signing space (Cuxac 2000). This space is employed in an intensive way, for example to specify the context, to locate the entities of the discourse, to conjugate directional verbs, to describe forms or actions. The annotation of 3d information related to this signing space is very useful for the linguists.

In our annotation software, we want to be able to build this space in two manners: by explicitly indicating the zones used to position entities, or by determining these zones by image processing (see the section "reconstruction"). The first method can be used to refine or complete the second method.

For that, we need to design a component dedicated to signing space modelling. At this moment, the first sketching out of such a component, carried out during a study on the automatic recognition of FLS utterances, allows us to represent information related to directional verbs (Braffort, 1996).

The modelling of signing space raises various problems, especially the segmentation of this space and the choice of the 3d references.

Segmentation:

We have to propose several levels of segmentations of the signing space, due to the various levels of granularities encountered: For example, the hands can be located in a defined zone, which has at least the size of the hand, but the finger can also indicates a precise point of the space; the gaze can indicate a precise point or a more global zone.

Spatial reference:

A style of story telling usually employed in SL relates to the "personal transfer" (Cuxac, 2000). The signer takes the role of one of the story entities to specify his characteristics or his actions. Various articulators are employed to specify when the signer switch from one role to another one. It can be a pointing carried out with the hand towards the entity location in the signing space, a gaze towards this zone, or a chest movement towards this zone.

Thus, several signing spaces can coexist, each owning its own spatial reference: the narrator space, and a space for each entity.

These different points must be taking into account while designing the modelling of signing space.

4. CONCLUSION

Currently, the different components have been implemented independently. A first version of the annotation software is under development. It includes two windows: one for video and one for annotation. We plan to carry out the first integration of all the components in the next months. The software will then propose one or more video windows, an annotation window, and a window for the representation of the signing space. Our linguist colleagues will validate this platform progressively.

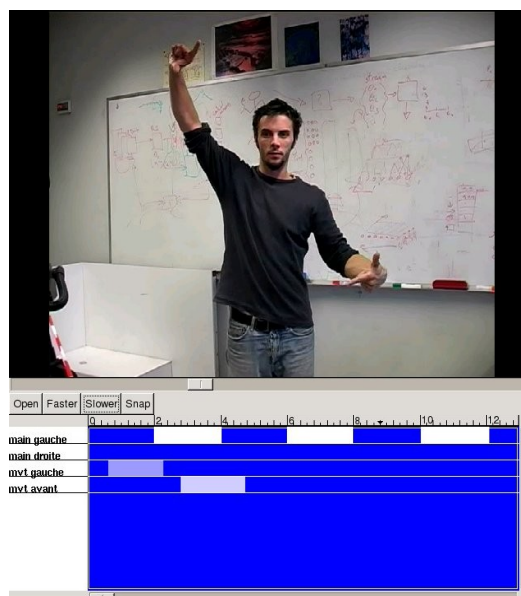


Figure 10: First version of the annotation software

5. REFERENCES

- Braffort A., Choisier A., Collet C., Cuxac C., Dalle P. et al. (2001). Projet LS-COLIN. Quel outil de notation pour quelle analyse de la LS ? In proceedings of RLSF'01 : Recherches sur les langues des signes.
- Braffort A.(1996). ARGo: An architecture for sign language recognition and interpretation. In Progress in Gestural Interaction. P.A. Harling and A.D.N. Edwards Eds. Springer Pub.
- Coanicu D., Ramesh V., Meer P. (2000). Real-Time Tracking of non rigid Objects using Mean Shift, In proceedings of Computer Vision and Pattern recognition.
- Cuxac C. (2000). La Langue des Signes Française (LSF) – Les voies de l'iconicité. In Faits de Langues 15-16, Ophrys.
- Dalle P., Lenseigne B., Hudelot C. (2001). Apport d'un système d'analyse d'images pour l'étude de la langue des signes, In proceedings of RLSF'01 : Recherches sur les langues des signes.
- Darrel T. (1998). A radial cumulative similarity transform for robust image correspondence, In proceedings of Computer Vision and Pattern Recognition.

- Davis J., Bobick A. (1997). The representation and recognition of human movements using temporal templates. In proceedings of Computer Vision and Pattern recognition.
- Kipp M. (2001). Anvil - A Generic Annotation Tool for Multimodal Dialogue. In proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech).
- Lainé J. (2002). Interaction homme-machine temps réel en vision par ordinateur. Application au suivi du doigt. Student report, UPS Toulouse.
- Lenseigne B., Gianni F., Dalle P. (2004). Estimation mono-vue de la posture du bras, méthode et évaluation. Proceedings of RFIA (Pattern recognition and Artificial Intelligence).
- Mercier H. (2003). Analyse automatique des expressions du visage. Application à la langue des signes. Master report, UPS Toulouse.
- Neidle, C. (2002). SignStream™: A Database Tool for Research on Visual-Gestural Language. In B. Bergman, P. Boyes-Braem, T. Hanke, and E. Pizzuto, eds., Sign Transcription and Database Storage of Sign Information, a special issue of Sign Language and Linguistics 4:1/2, 203-214.
- Quek F., McNeill D., Ansari R., et al. (1999). Gesture Cues for Conversational Interaction in Monocular Video. Proceedings of ICCV'99, International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems (RATFG-RTS'99)
- Risser L., Laval D., (2002). Analyse et reconstruction d'un geste de signation. Student report, UPS Toulouse.
- Sallandre M.-A. (2003). Les unités du discours en Langue des Signes Française. Tentative de catégorisation dans le cadre d'une grammaire de l'iconicité. PhD thesis, Paris 8 University.
- Sallandre M.-A. & Cuxac C. (2002) : Iconicity in Sign Language : a theoretical and methodological point of view. In Wachsmuth I. & Sowa T. (eds.) : LNAI 2298, Springer-Verlag, Berlin (2002).
- Wittenburg P. Brugman, H.; Levinson, St.; Kita; S. (2002). Multimodal Annotations in Gesture and Sign Language Studies. In proceedings of the LREC 2002 Conference.