

Descripteurs de mouvement pour l'annotation semi-automatique de vidéos de langue des signes



Encadrantes : Michèle Gouiffès (michele.gouiffes@limsi.fr) et Annelies Braffort (annelies.braffort@limsi.fr)

Dates : 5 mois dès que possible à définir avec le stagiaire.

Lieu : Laboratoire [LIMSI-CNRS](#) , Rue John Von Neumann Bât. 508, 91403 ORSAY.

Rémunération : Indemnité légale, gratification de stage de 546 €/

Mots clé : Annotation automatique de vidéos, langue des signes, analyse de vidéos, interaction homme-machine, analyse du mouvement, reconnaissance de gestes.

Contexte

Le **LIMSI** (Laboratoire d'Informatique et de Mécanique pour les Sciences de L'ingénieur), unité propre du CNRS, abrite des thématiques de recherche variées réparties en deux départements. Ce stage se déroulera dans le département CHM (pour Communication Homme Machine) qui travaille sur le traitement de la langue parlée, écrite et signée, le traitement du signal audio-accoustique, l'interaction homme-machine, la réalité augmentée. Le stage fait l'objet d'une collaboration entre le thème [M&TALS](#) du groupe [ILES](#), et le thème [image](#) du groupe [AMI](#).

M&TALS travaille sur l'élaboration, l'analyse et l'annotation de lexiques et de corpus de Langues de Signes (LS) qui permettent de mener des études sur les LS en sciences du mouvement et en linguistique. De plus, à l'aide de ces ressources, l'équipe propose des modélisations mathématiques et informatiques des LS qui peuvent être implémentées dans des applications de génération automatique par l'animation d'avatars 3D ou d'aide à la traduction de l'écrit vers la LS et ainsi faciliter l'accès aux informations pour les personnes sourdes.

Le thème **Image** quant à lui travaille sur le développement de méthodes d'analyse d'images et de vidéos avec des applications en interaction homme-machine, réalité augmentée spatiale et art-science. Il s'intéresse notamment à l'extraction de primitives visuelles (couleur, texture, mouvement etc) et aux modèles de représentation des objets et des activités en vue de leur reconnaissance. Dans le cadre du projet de recherche MAPOCA financé par Digiteo, nous proposons de nouvelles méthodologies d'analyse de vidéos pour la reconnaissance d'activités et de gestes, fondées sur les caractéristiques cinématiques de trajectoires et sur des descripteurs définis par des matrices de covariance de caractéristiques dynamiques.

Ce stage va permettre d'établir un pont entre les thématiques de recherche [des](#) deux équipes.

Objectifs du stage

Le sujet de stage porte sur **l'annotation semi-automatique de vidéos de langue des signes par analyse vidéo**. L'objectif du stage est de développer des méthodes d'analyse automatique de vidéos de LS dans le but de fournir à l'expert une première annotation qu'il devra valider/modifier et qui lui permettra de concentrer une énergie plus importante sur l'analyse fine de la vidéo. Cette annotation automatique de la vidéo se fera de manière **incrémentale** (ou hiérarchique) en commençant par une classification **globale** (ou grossière) du mouvement puis en analysant plus localement (ou plus finement) les mouvements dans la vidéo, c'est-à-dire en dissociant le mouvement des mains, du visage, des épaules. À cet effet, une approche par arbre de décision pourra être employée.

Le stage se focalisera sur la première étape de classification (globale) qui devra permettre d'extraire des informations de structure du mouvement gestuel telles que celles illustrées sur la figure 1 : la dépendance (Fig. 1 (a)), la symétrie (Fig. 1 (b)), l'alternance, la répétition (voir Fig. 1 (c)), la ou les orientation(s) du mouvement.

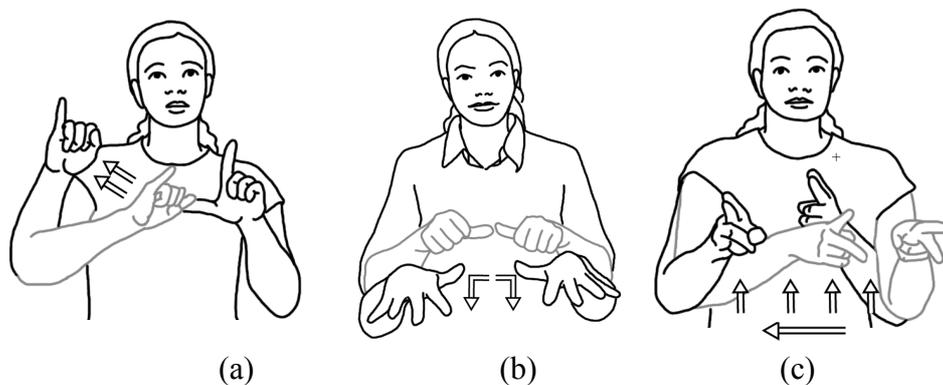


Figure 1. Exemples de signes de la LS française. (a) signe pour « LIMSI » qui montre une dépendance entre les deux mains ; (b) signe pour « réseau informatique » qui montre une symétrie des mouvements des mains ; (c) signe pour « segmentation », montrant une répétition.

La première phase du travail consistera à réaliser un état de l'art sur l'analyse de vidéo pour l'aide à l'annotation de LS. Les méthodes d'analyse vidéo qui seront développées se fonderont sur trois étapes : **(1)** l'extraction du mouvement par flot optique ; **(2)** le calcul de descripteurs globaux de mouvement ; **(3)** la mise en correspondance des descripteurs dans un schéma de classification. En termes de descripteurs, nous étudierons en priorité les matrices de covariance de caractéristiques [2,8] qui ont amplement prouvé leur pertinence pour la reconnaissance d'objets mais qui ont très peu été exploitées pour la reconnaissance d'activités ou de gestes. L'approche sera finalement validée sur des vidéos préalablement réalisées au LIMSI et annotées par les experts.

Perspectives.

Un sujet de thèse sera proposé dans la continuité de ce sujet de stage pour la rentrée 2016.

Profil recherché.

Étudiant en Master 2 recherche dans le domaine de l'informatique et du traitement d'images.

Comment candidater ?

Envoyer un CV et une lettre de motivation aux encadrantes.

Quelques références

- [1] M. Gonzales (thèse 2012). *Computer vision methods for unconstrained gesture recognition in the context of sign language annotation*. Univ Toulouse Paul Sabatier.
- [2] F. Laguzet A.Romero, M.Gouiffès, L.Lacassagne, D.Etiemble (2013) Color tracking for contextual switching: Real-time implementation on CPU. *Journal of Real-Time Image Processing* June 2015, Volume 10, Issue 2, pp 403-422
- [3] F. Lefebvre-Albaret (thèse 2010). *Traitement automatique de vidéos en LSF. Modélisation et exploitation des contraintes phonologiques du mouvement*. Univ Toulouse Paul Sabatier.
- [4] F. Martinez, A. Manzanera, M. Gouiffès and Annelies Braffort. A Gaussian mixture representation of gesture kinematics for on-line Sign Language video annotation ISVC 2015
- [5] E-J. Ong; O. Koller; N. Pugeault ; R. Bowden (2014). *Sign Spotting using Hierarchical Sequential Patterns with Temporal Intervals*. *IEEE Conference on Computer Vision and Pattern Recognition (2014)*, 1931--1938, 23-28.
- [6] S. Rautaray and A. Agrawal (2015). Vision based hand gesture recognition for human computer interaction: a survey. *Artificial Intelligence Review*, 43(1), 1-54.
- [7] S. Theodorakis, V. Pitsikalis, P. Maragos (2014). *Dynamic-static unsupervised sequentiality, statistical subunits and lexicon for sign language recognition*. *Image and Vision Computing* 32 (2014) 533-549
- [8] O. Tuzel F. Porikli, P. Meer (2006) *Region Covariance: A Fast Descriptor for Detection and Classification*. *European Conference on Computer Vision (ECCV 2006)*.