



**Sujet :**

**Création de résumé vidéo par une approche multimodale**

**Niveau** : Bac + 5 - 3<sup>ème</sup> année d'école d'ingénieur – Master

**Durée du stage** : 4 à 6 mois

Stage pouvant se poursuivre en thèse :  OUI  NON

**Contexte**

La constitution et l'accumulation de bases de vidéos numériques, que ce soit en milieu professionnel avec les archives audio-visuelles (voir la plateforme de l'INA), sur internet (les sites de type YouTube) et bientôt au niveau domestique avec les capacités de stockage numérique qui seront associées à la prochaine génération de téléviseurs, posent le problème de leur consultation et de leur exploitation. Parmi les modes d'indexation de vidéos par le contenu, la création automatique de résumés vidéo s'avère très pertinente et encore très peu explorée.

En effet, il est aujourd'hui possible de créer de façon indépendante un résumé textuel et un résumé visuel à partir des données extraites des vidéos. Cependant, l'aspect parallèle des deux traitements entraîne un certain nombre de limitations de la qualité des résultats obtenus. D'une part, les résumés visuels [1] sont très limités en termes de sémantique et ne contiennent donc que peu d'information pertinente sur les vidéos. D'autre part les résumés textuels [2] sont dissociés des caractéristiques visuelles et notamment de la structure des vidéos, ce qui limite leur cohérence.

**Objectifs du stage**

L'objectif de ce stage est de construire un outil de résumé des vidéos sémantique et multimédia en vue d'une présentation synthétique de leur contenu. Une compréhension profonde des mécanismes d'interrelations entre le flux de parole et visuel permettra le couplage des outils de résumé liés à chacun des médias. Le résumé visuel et textuel obtenu devra ainsi présenter une cohérence forte entre l'image diffusée et le texte l'accompagnant. Les technologies développées seront valorisées sous formes de publications scientifiques et s'inséreront dans les projets auxquels participent le laboratoire, à la fois au niveau national et européen, ainsi que ceux en collaboration directe avec des industriels.

Le stage comportera trois étapes classiques :

- état de l'art et appropriation des outils existant au CEA
- définition et implémentation de l'outil
- réalisation d'une interface de visualisation

**Bibliographie**

[1] M. Guironnet, « Méthodes de résumé vidéo à partir d'informations bas niveau, du mouvement de caméra ou de l'attention visuelle », thèse de l'université Grenoble I, 2006.

[2] Gaël de Chalendar, Romaric Besançon, Olivier Ferret, Gregory Grefenstette and Olivier Mesnard (2005)

« Crosslingual summarization with thematic extraction, syntactic sentence simplification, and bilingual generation ». Workshop on Crossing Barriers in Text Summarization Research, RANLP 2005 (Recent Advances in Natural Language Processing), Borovets, Bulgaria.

**Domaines de spécialité requis**

Informatique, Traitement du signal (image, audio ou vidéo) ou traitement du langage.

**Moyens informatiques mis en œuvre**

Langages : C++, matlab, scripts

Logiciels : kdevelop, matlab



COMMISSARIAT A L'ENERGIE ATOMIQUE  
DRT/DTSI/SRCI  
Relation avec les écoles : MARC MERGY  
Tél : 01 46 54 81 58  
E-Mail : marc.mergy@cea.fr

### Autres moyens mis en œuvre (expériences, méthodes d'analyses, autres...)

**Centre :** Fontenay-aux-Roses

**Pôle ou Direction :** DRT - LIST

**Dépt/Service/Labo :** DTSI/SRCI/LIC2M :

Laboratoire d'Ingénierie de la Connaissance Multimédia Multilingue

**Web :** <http://www-list.cea.fr/>

### Encadrement

**Nom de l'ingénieur responsable :** Bertrand Delezoide

**Tél. :** +33 1 46 54 86 53

**E-mail :** [bertrand.delezoide@cea.fr](mailto:bertrand.delezoide@cea.fr)

**Fax :** +33 1 46 54 75 80

**Secrétariat :** DTSI/SRCI

**Tél. :** +33 1 46 54 91 17

**Nom du chef de laboratoire :** Olivier Mesnard

**Tél. :** +33 1 46 54 73 38