



Sujet :

Transcription de parole pour l'indexation multimédia

Niveau : Bac + 5 - 3^{ème} 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.

La recherche par le contenu des vidéos nécessite une étape de transcription de la parole permettant d'utiliser les outils « classiques » des moteurs de recherche sur le texte [1,2]. Ces algorithmes font appel à divers techniques permettant de rendre compte de la morphologie, de la syntaxe et même partiellement de la sémantique des textes. Néanmoins cela suppose d'avoir un texte rédigé « correctement » à la base. Or les algorithmes de transcription de la parole ont un rendu imparfait et le texte résultant comporte de nombreuses coquilles induisant les moteurs de recherche en erreur

Objectifs du stage

L'objectif de ce stage est d'évaluer des algorithmes de transcription de parole [3,4,5] dans le cadre de la recherche vidéo [6], puis de proposer et mettre en œuvre des solutions pour améliorer les performances. L'évaluation sera conduite sur des corpus de données assez larges, permettant de se comparer aisément avec l'état de l'art. L'ensemble des travaux seront intégrés dans un système de traitement des vidéos complet existant au laboratoire. Les travaux 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 LIST
- définition et implémentation de l'outil
- amélioration de l'outil pour la recherche de vidéos
- intégration dans un moteur de recherche multimédia

Bibliographie :

- [1] www.google.com
- [2] www.exalead.com
- [3] J.-L. Gauvain, G. Adda, M. Adda-Decker, A. Allauzen, V. Gendner, L. Lamel et H. Schwenk. « Where are we in Transcribing French Broadcast News? ». Dans Proc. of the 9th European Conference on Speech Communication and Technology (Eurospeech), Lisbonne, Portugal, 2005.
- [4] D. S. Pallett. « A Look at NIST's Benchmark ASR Tests: Past, Present, and Future ». Dans Proc. of the IEEE Workshop Automatic Speech Recognition and Understanding, St. Thomas, îles Vierges, Etats-Unis, 2003.
- [5] G. Linarès, P. Nocera, D. Massonié, and D. Matrouf, The LIA speech recognition system: from 10xRT to 1xRT, rapport technique, 2007.

Domaines de spécialité requis :

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

Moyens informatiques mis en œuvre :

Langages : C++, matlab, scripts

Logiciels : kdevelop, matlab



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

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