

Cette thèse s'inscrit dans le cadre du développement d'un nouveau thème de recherche autour de l'analyse et indexation conjointe audio/vidéo au sein de l'équipe image et son du Labri. En effet, dans le domaine de l'analyse et de l'indexation vidéo, la plupart des méthodes existantes aujourd'hui se base soit uniquement sur les données images, soit uniquement sur les données audio. Dans la mesure où ces approches n'utilisent qu'une partie des données disponibles, ces approches sont fondamentalement sous-optimales. Cette situation s'explique largement par le fait que la mise en œuvre d'une approche conjointe audio/vidéo nécessite des compétences dans les deux domaines (analyse audio et vidéo), or cette double compétence n'existe que dans peu de laboratoires. Dans ce cadre, l'objectif de ce projet consiste à exploiter le fait que l'équipe Image et Son du Labri possède cette double compétence qu'il s'agit d'exploiter pour déployer efficacement de nouvelles méthodologies pour la structuration de films tel que décrit dans le sujet de thèse proposé ci-dessous.

Les applications envisagées sont nombreuses et reposent en particulier sur la détection de répétitions ou la détection de changements importants. Les algorithmes développés dans le cadre de ce doctorat doivent permettre la segmentation automatique de films en scènes, la classification des scènes pour faciliter la navigation (scènes d'action, romantiques, etc) et la génération automatique de résumés.

Sujet de thèse :

La problématique de l'indexation des vidéos est un sujet important en raison de l'augmentation spectaculaire du volume de données numériques disponible aujourd'hui. Pour obtenir une indexation réellement utile pour un utilisateur, il est nécessaire d'obtenir des informations de haut niveau sémantique sur le contenu et la structure des vidéos. Dans ce contexte, la difficulté essentielle consiste à définir des descripteurs « sémantique », c'est-à-dire capable de différencier dans une vidéo par exemple des scènes d'action, d'émotion, d'interview, de sport, d'intérieur ou d'extérieur... Pour être efficace, ces descripteurs doivent à la fois être identifiables dans les données via des techniques d'identification et de modélisation, et, d'autre part, être réellement représentatifs du type de scènes. La définition de tels descripteurs dans le domaine de la vidéo est difficile en raison de l'écart existant entre le niveau pixel dans les images et le niveau sémantique que l'on cherche à obtenir. Les techniques existantes aujourd'hui se basent souvent seulement sur les données images, ce qui ne permet pas d'obtenir une caractérisation d'un niveau sémantique suffisant. Pour résoudre ce problème, une approche possible consiste à exploiter l'ensemble des données disponibles. Dans ce contexte, l'objectif de cette thèse consistera à définir des descripteurs intégrant à la fois les données audio et vidéo, ainsi que les métadonnées disponibles. En particulier, les informations de type musicales seront tout particulièrement exploitées dans le cadre des films.

Cette thèse se déroulera au Labri dans le cadre d'une collaboration entre les thèmes « Modélisation du Son et de la Musique » et « Analyse et Indexation Vidéo » de l'équipe Image et Son.