Construction et Présentation des Vidéos Interactives

L'arrivée de la norme MPEG-7 pour les vidéos exige la création de structures de haut niveau représentant leurs contenus. Le travail de cette thèse aborde l'automatisation de la fabrication d'une partie de ces structures. Comme point de départ, nous utilisons des outils de segmentation des objets en mouvements. Nos objectifs sont alors : retrouver des objets similaires dans la vidéo, utiliser les similarités entre plans caméras pour construire des regroupements de plans en scènes. Une fois ces structures construites, il est facile de fournir aux utilisateurs finaux des outils de visualisation de la vidéo permettant des navigations interactives : par exemple sauter au prochain plan ou scène contenant un personnage.

La difficulté principale réside dans la grande variabilité des objets observés : changements de points de vues, d'échelles, occultations, etc. La contribution principale de cette thèse est la modélisation de la variabilité des observations par un mélange de densités. La théorie du mélange gaussien est employée dans cette approche. Cette modélisation permet de capturer les différentes apparences intra-plan de l'objet suivi et de réduire considérablement le nombre des descripteurs de bas niveaux à indexer par objet suivi.

Autour de cette contribution se greffent des propositions qui peuvent être vues comme des mises en oeuvre de cette première pour différentes applications : mise en correspondance des objets suivis représentés par des mélanges gaussiens, fabrication initiale des catégories de tous les objets présents dans une vidéo par une technique de classification non supervisée, extraction de vues caractéristiques et utilisation de la détection d'objets similaires pour regrouper des plans en scènes.

Mots clefs : Vidéo hyperliée, MPEG-7, Reconnaissance et classification d'objets, Modélisation de la variabilité, Modèles de mélange gaussien, Navigation interactive, Structure de la vidéo.

Constructing and Browsing of Interactive Videos

The arrival of the MPEG-7 standard for videos requires the creation of high level structures representing their content. The work of this thesis approaches the automatic building of a part of these structures. As a starting point, we use the tools for segmentation of moving objects. Our objectives are then to find similar objects in the video and subsequently use the similarities between camera shots to group shots into video scenes. Once these structures have been built, it is easy to provide video visualization tools for the end users which permit interactive navigation like jumping to the next shot or scene containing a person.

The main difficulty lies in the great variability of observed objects: changes in point of view, scales, collusions, etc. The principal contribution of this thesis is the modeling of the variability of observations by a mixture of densities based on the Gaussian mixture theory. This modeling captures various intra-shot appearances of a tracked object and considerably reduces the number of low-level descriptors to be indexed by each tracked object.

The proposed formulation led to an implementation designed for different applications: matching of tracked object models represented by Gaussian mixtures, initial building of categories of all objects present in a video by a non-supervised classification technique, extraction of characteristic views and use of detected similar objects for grouping shots into scenes.

Keywords : Hyperlinked video, MPEG-7, Object recognition and classification, Variability modeling, Gaussian mixture models, Interactive video navigation, Video structure.

Thèse préparé dans le laboratoire GRAVIR–IMAG et INRIA Rhône-Alpes, au sein de l'équipe MOVI.
INRIA Rhône-Alpes, 655, Av. de l'Europe, ZIRST, 38330 Montbonnot Saint-Martin.