Chapitre 6

Extraction des apparences-clés

Je m'autoriserai des temps de pause.

Le niveau le plus élémentaire dans la structure hiérarchique d'un film vidéo, illustrée dans la figure 2.1, est celui des *images-clés*. Visuellement quelques images fixes choisies à différents instants de l'axe temporel, nommées "images-clés", peuvent décrire suffisamment le contenu (ou l'histoire) du plan vidéo. Lorsque des objets sont suivis dans les plans alors on introduit le niveau des *apparences-clés* .

Dans ce chapitre nous nous intéressons au problème de l'extraction automatique des apparences-clés d'un objet suivi. Notons qu'une relation forte existe entre apparence clé et image clé mais sans généralisation; par exemple le cas d'une scène avec un changement du fond (plusieurs images-clés) et sans variation des apparences des objets (une seule apparence clé).

La sélection des images-clés est un problème rencontré dans différents domaines d'applications : présentation, indexation, codage et transmission de la vidéo. Quelles sont les images les plus informatives à présenter à l'utilisateur de la vidéo structurée? afin de lui permettre une visualisation rapide du contenu de plan (les images-clés pointent sur les plans correspondants dans la vidéo, un clic sur une image clé permet par exemple de jouer le plan correspondant; idem pour la mosaïque d'objets [56]), et/ou une vue globale du contenu du film (balade dans la mosaïque de plans, voir figure 4.1 par exemple), tout en prenant en compte la capacité limitée des écrans d'ordinateurs et le nombre important de plans d'un film vidéo ($500 - 3000$ plans). En ce qui concerne l'indexation de la vidéo par le contenu, la similitude entre les plans d'une vidéo se fait souvent sur la base des images clés où l'image médiane (ou de début/fin) du plan est considérée comme image représentative [8] [92] [130] [53]. Cependant, ce choix simplificateur conduit à des taux de bons appariements des plans (ou objets) très faibles [57] pour les raisons que nous avons discutée dans le chapitre 4.

Un résumé de l'approche de sélection des images- apparences-clés que nous avons proposée est décrit dans la section suivante. Cette approche est expérimentée sur plusieurs objets suivis du film “Avengers” et sa mise en oeuvre dans l'application de la vidéo interactive est seulement exploitée pour un objectif de présentation (*browsing*) intelligente de l'hyper vidéo.

6.1 Résumé de “A Probabilistic Framework of Selecting Effective Key-Frames for Video Browsing and Indexing” - RISA‘2000

Cet article, publié à RISA‘2000, concerne la séléction des apparences intra-plan clés des objets suivis par une méthode de classification non-supervisée. Les outils de segmentation en plan et de suivi d'objets sont présentés dans le chapitre 2. L'idée de base de cette approche consiste dans un premier temps à regrouper les apparences d'un objet suivi qui sont proches dans l'espace de descripteurs de bas niveau en des classes homogènes. Pour ce fait le mélange gaussien est employé comme une méthode de classification automatique. Cette technique ainsi que le traitement des données expérimentales ont été présentés en grand détail dans le chapitre 3. Dans un second temps, un algorithme adaptatif permet de sélectionner les apparences-clés des classes d'apparences retenues dans la première étape. D'abord l'apparence médiane de chaque classe d'apparences est sélectionnée. Ensuite une phase de test de la compacité des classes d'apparences à l'aide de la distance de Mahalanobis permet d'étendre le nombre des apparences-clés par classe, et une dernière phase de vérification temporelle consiste à garder le nombre minimal des apparences-clés de l'objet suivi et qui soient suffisamment séparés dans le temps. Une description plus détaillée de cet algorithme est présentée dans les pages suivantes.