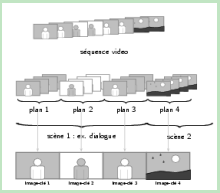
A.C.I. Masse de données NavGraphe

Navigation dans les contenus multimédia indexés basée sur les graphes

Introduction

Les besoins croissants dans le contexte de la recherche, la récupération et la manipulation de l'information multimédia rendent nécessaires de nouveaux modes de visualisation du contenu multimédia dans les interfaces homme-machine afin d'assurer des fonctionnalités de navigation à l'intérieur du contenu visuel. Cette navigation peut être considérée à deux niveaux : (i) la navigation à l'intérieur de grandes bases de données contenant des documents multimédia et (ii) la navigation à l'intérieur de chaque unité de contenu multimédia qui peut représenter des objets très complexes tels que des programmes TV ou des documentaires sous forme digitalisée. Dans les deux cas les données multimédia doivent être **indexées par le contenu** afin de permettre une **visualisation intuitive basée sur la similitude entre les descripteurs extraits**. Le standard multimédia **MPEG7** [1] propose un ensemble de descripteurs et de composition de descripteurs permettant l'indexation normalisée des documents multimédia. Ces descripteurs contiennent non seulement des index textuels mais également des **descripteurs de bas niveau** calculés directement à partir des signaux audio et visuels tels que les descripteurs basés sur la texture, la couleur, les formes ou le mouvement de caméra, pour la vidéo et le rythme, la hauteur ou d'autres caractéristiques acoustiques pour le son.Ces descripteurs peuvent être utilisés dans de la cadre d'interfaces de navigation intuitives.
Dans le cas des contenus vidéo, les interfaces de navigation habituelles proposent généralement une représentation linéaire du contenu sous la forme de storyboards. De telles interfaces de navigation linéraires ne montrent pas la **structure globale d'un document visuel** et ne permettent pas une identification facile des plans similaires qui sont éloignés dans le temps.
Les interfaces basées sur les graphes représentent aujourd'hui un champ très actif de recherches. La raison principale est qu'à partir de toutes les structures de données on peut facilement extraire une représentation sous forme de graphe. Il a été montré que notre perception visuelle est très efficace et que quand un algorithme ne peut pas relier des objets entre eux, l'oeil le peut si une bonne représentation est donnée à l'utilisateur. Notre approche est clairement dans le domaine de la visualisation de l'information et plus précisément dans **le paradigme de Schneiderman** [2]: *"vue d'ensemble d'abord, zoom et filtrage, puis détails à la demande"*.

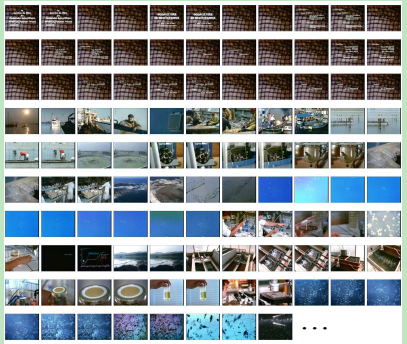
Description des contenus vidéo



Une **séquence vidéo** peut être décomposée en une hiérarchie de **scènes** et de **plans**. Un plan vidéo représente une **prise de caméra continue**. Les plans qui **partagent la même sémantique** peuvent être groupés en scènes. Ici, on suppose que la sémantique des plans vidéo est suffisament bien exprimée par des **descripteurs de bas-niveau**.

Chaque plan vidéo peut être caractérisé par une ou plusieurs **images-clé** extraites du plan vidéo.

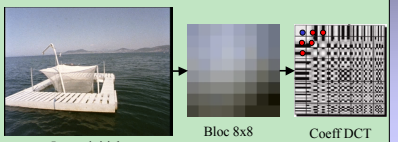
L'ensemble des images-clé issues d'un document vidéo peut être visualisé sous forme de **story-board**.
On a représenté ici les images clés issues des chacun des plans du documentaire intitulé « Aquaculture en Méditerranée » (SFRS).

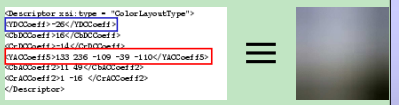


Indexation couleur

Chaque descripteur MPEG7 est un résumé des caractéristiques de l'image. On peut notament extraire des descripteurs de couleur de chaque image-clé et utiliser la mesure de similarité associé pour comparer des paires d'images selon les caractéristiques représentées par le descripteur.

Le descripteur **MPEG7 ColorLayout** d'une image est obtenu en sous échantillonant cette image pour obtenir un bloc de 8x8 pixels. L'information des 3 canaux de couleur du bloc est traitée séparément afin d'obtenir une décomposition en signaux de base par application de la DCT. Ces signaux de base caractérisent la **composition spatiale** du signal. Les coefficients situés dans la partie supérieure gauche de la matrice correspondent aux **signaux de basse fréquence** et composent le descripteur CLD car c'est l'information **la plus significative visuellement.**

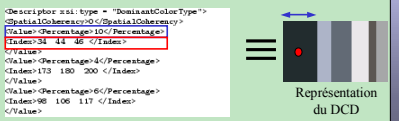




Représentation XML MPEG7 du descripteur ColorLayout

Représentation du CLD

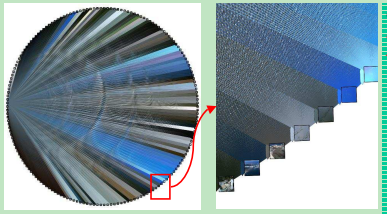
Le descripteur **MPEG7 DominantColor** est obtenu par quantification vectorielle des couleurs présentes dans l'image en classes. Le pourcentage des pixels associés à chaque classe ainsi que la valeur des 3 composantes couleur de chaque classes constituent le DCD.



Représentation XML MPEG7 du descripteur DominantColor

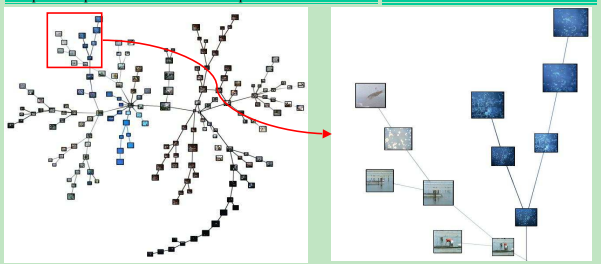
Représentation du DCD

Visualisation basée sur les graphes



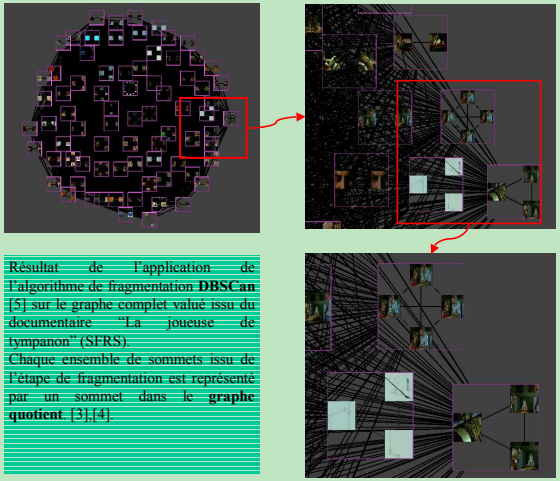
Construction du graphe complet
S : ensemble des images-clé
A_e : mesure de similarité entre images basée sur le descripteur e
Le graphe complet valué, $G=(V,E)$ est défini par :
- $V = S$
- $E = \{(s, s') \; s/s' \wedge s \in S, s' \in S\}$
- $f : E \rightarrow R$
  $e \in E, e = (s_i, s_j)$
  $f(e) = A_e(s_i, s_j)$

Graphe complet du documentaire "Aquaculture".



Arbre couvrant de poids minimum du graphe complet. La mesure de similarité utilisée est celle du descripteur **ColorLayout**.

Fragmentation et détection de scènes visuelles



Résultat de l'application de l'algorithme de fragmentation **DBSCan** [5] sur le graphe complet valué issu du documentaire "La joueuse de tympanon" (SFRS).
Chaque ensemble de sommets issu de l'étape de fragmentation est représenté par un sommet dans le **graphe quotient**. [3],[4].

Perspectives

Navigation locale : utilisation de structures de graphes de type grille pour représenter un ensemble d'images indexées dans son espace de description. Utilisation d'agents de recherche utilisant le routage glouton pour la résolution de requêtes par l'exemple (en cours).
Descripteurs sémantiques et seuils de discrimination : étude de descripteurs sémantiques en dimension 1 et caractérisation des seuils de discrimination associés (en cours).
Zoom sémantique et méthodes d'échantillonage : utilisation de 2 méthodes d'échantillonage efficaces d'un ensemble de points en haute dimension pour proposer un outil de navigation dans une collection d'images indexées mettant en oeuvre le zoom sémantique.
Détection de communautés : en collaboration avec l'équipe du Pr. Guy Melançon du LIRMM. Utilisation de techniques de détection de communautés pour le partitionnement du graphe d'indexation.

Références

[1] ISO/IEC JTC 1/SC 29/WG 11/M6156, *MPEG-7 Multimedia Description Schemes WD (Version 3.1)*, Beijing, July 2000
[2] "*The eyes have it: A Task by Data Type Taxonomy for Information Visualization*," B. Shneiderman, IEEE Conference on visual languages, Boulder, 1996, 336-343.
[3] "Intuitive color-based visualization of multimedia content as large graphs", Maylis Delest, Anthony Don, Jenny Benois-Pineau, in Visualization and Data Analysis 2004, San Jose, California;June 4, 2004; p. 65-74
[4] "DAG-Based visual interfaces for navigation in indexed video content", Maylis Delest, Anthony Don, Jenny Benois-Pineau, accepted in Multimedia Tools and Applications.
[5] "A density-based algorithm for discovering clusters in large spatial databases with noise", Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiowei Xu, in KDD'