

REPUBLIC ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE
SCIENTIFIQUE
INSTITUT NATIONAL DE FORMATION EN INFORMATIQUE

Mémoire

Pour l'obtention du diplôme
d'ingénieur d'état en informatique
Option systèmes informatiques

THEME

Indexation de la vidéo par le contenu

Organisme d'accueil
CERIST

Proposé par
M.MAREDJ AZEDDINE

Présenté par
M.MEDDOUR DJAMEL EDDINE
M.SAADI FARID

Promotion 1999-2000

TH1.5722

Résumé

Notre étude porte sur l'indexation de la vidéo par le contenu dans le but d'offrir la possibilité de recherche et d'accès à des bases de données vidéo basée sur des critères de contenu (mouvement, intensité, couleur). Dans un premier temps, nous avons présentons différentes techniques de traitement d'image servant à la structuration du contenu de la vidéo. L'approche du découpage de la vidéo en plans élémentaires est basée sur l'utilisation des histogrammes globaux d'intensité et sur une technique de détection du changement à deux seuils. L'extraction des objets mobiles de la vidéo a été menée grâce à une technique de segmentation spatio-temporelle non supervisée à deux phases, la première consiste à obtenir une partition spatiale de l'image par l'utilisation d'un algorithme d'étiquetage statistique de graphe, la deuxième phase permet d'estimer le mouvement apparent sur les différentes régions spatiales de l'image. Cette opération permet d'extraire les différents descripteurs de la vidéo qui sont indexés dans une base de données pour faciliter la recherche des vidéos ayant des descripteurs similaires.

L'étude a aboutit à une conception et réalisation d'un prototype mettant en œuvre les techniques et les approches retenues le long de l'étude.

Mots-Clefs : indexation par le contenu – analyse du mouvement – document multimédia – techniques multigrille – champs markoviens – segmentation – estimateurs robustes – découpage en plans.

Introduction générale

La vidéo et les images occupent une place prépondérante dans les documents multimédias. Films, vidéos, peintures et dessins numériques sont de nos jours omniprésents, et mis à la disposition des professionnels et du grand public. La démocratisation de l'informatique et des télécommunications, le coût désormais accessible des caméras et mémoires de masse, et bien évidemment, la quantité d'information que transporte l'image par rapport au texte, sont responsables de ce développement.

Les Bases de Données vidéo peuvent de nos jours contenir plusieurs milliers de documents et occuper plusieurs téra-octets. Elles concernent essentiellement les applications professionnelles qui les utilisent depuis longtemps : archives audiovisuelles pour les émissions télévisées, bandes vidéo de surveillance de trafic routier, imagerie médicale et météorologique....

Il est donc indispensable de développer des outils permettant de sélectionner les vidéos les plus pertinentes par leur contenu, comme c'est le cas pour les systèmes de recherche d'information. Le problème majeur qui se pose, alors, est qu'à l'inverse des données alphanumériques, les images ne sont pas décrites par un alphabet, et ne peuvent donc être répertoriés dans un dictionnaire.

Traditionnellement, des attributs textuels de l'image tels que le nom de fichier, la légende et les mots clés ont été utilisés pour indexer les bases de vidéo. On a donc déplacé la problématique de la recherche dans une base vidéo en une problématique de recherche de mots, que l'on sait bien résoudre. Ce type d'approche est parfois efficace mais s'avère limité. Il nécessite l'intervention manuelle d'un opérateur dont la sémantique est souvent inadéquate, approximative ou subjective. De plus, les mots sont confrontés à la barrière linguistique, alors que les images ont une signification universelle.

Depuis le début de la décennie, les chercheurs se posent le problème de l'indexation automatique des images fondée sur leur contenu. Ce thème de recherche se situe à mi-chemin entre la vision par ordinateur et la recherche d'information. Au vu de la variété des besoins dans les domaines concernés par *la recherche dans une Base de Données vidéo par le contenu* (applications scientifiques, images satellitaires, la médecine, télécommunications, télésurveillance, les médias...), il est évident que le problème de la recherche de vidéo par

leur contenu est très complexe, et ne connaît sans doute pas de solution générique. La difficulté du problème vient du contenu même du document vidéo qui peut être très varié, et dont l'analyse est très difficile, voire impossible dans le cas d'images bruitées ou contenant un mouvement très important (mouvement de foule, ou d'arbres...). L'intérêt se porte alors vers des vidéos dont le contenu est généralement plus maîtrisé, et qui ne contiennent pas d'éléments qui perturbent trop leur analyse.

Construire un système permettant d'interroger des bases d'images ou de vidéo par le contenu sollicite des compétences que l'on trouve dans le domaine du traitement d'images et dans celui des bases de données. Du domaine de l'image sont tirées les techniques d'analyse du signal qui permettent d'extraire de l'image les informations caractéristiques. Ces informations doivent permettre de reconnaître le contenu de l'image même en présence de certaines variations : illuminations, mouvements... Ces informations extraites se présentent généralement sous la forme d'un ou de plusieurs ensembles de valeurs numériques réelles.

Les compétences en bases de données sont nécessaires dès que le volume de la base d'images devient trop important pour tenir en mémoire centrale, ce qui est souvent le cas. Il devient alors nécessaire de stocker sur disque les images et leurs descripteurs. Les techniques de base de données sont alors employées pour accroître la vitesse des recherches et minimiser le coût lié aux entrées-sorties. Les bases de données savent en effet organiser spécifiquement les données sur mémoire secondaire, offrent des techniques de structuration et d'indexation accélérant les recherches, et offrent également des garanties de fiabilité et de cohérence d'accès concurrents, garanties souhaitables au sein d'un système d'interrogation réel.

La présente étude consiste à concevoir et à réaliser un prototype d'indexation et de recherche automatiques de base de données vidéo par le contenu. Nous n'émettons aucune supposition sur le contenu des vidéos qui peut être quelconque. Mais, pour les besoins de l'analyse du mouvement, par exemple, le contenu doit être analysable au sens où une partition des images en objets doit être possible et avoir un sens.

Il s'agit pour nous d'analyser le contenu de chaque vidéo en utilisant des techniques de traitements d'images en vue de l'extraction des descripteurs de son contenu. Ces descripteurs sont, ensuite, stockés dans une base de descripteurs qui servira pour la recherche dans la base vidéo.

Les requêtes se font par l'exemple, dans ce cas l'utilisateur soumet une vidéo contenant des caractéristiques similaires au sens de la couleur, de l'intensité ou du mouvement aux documents qu'il cherche à obtenir. Le système doit alors analyser cette vidéo, en extraire les

attributs adéquats pour la recherche souhaitée et effectuer la recherche dans la base des descripteurs, et à la fin lui présenter les résultats.

Que se soit en traitements d'images ou en Base de Données, les techniques que nous allons utiliser doivent toutes viser deux objectifs :

- finesse des résultats : qui doivent être d'une bonne qualité vu le contexte de travail qui est l'indexation et la recherche basées sur la précision et la pertinence des informations manipulées.
- optimisation des temps de calcul : vu la nature et le volume des données à traiter les temps de calcul restent un facteur déterminant quant à la réussite ou non du système envisagé.

Ces deux objectifs sont contradictoires, d'un côté l'obtention des résultats précis nécessite un temps de calcul énorme (surtout en traitement d'images). De l'autre côté, l'optimisation du temps de calcul se fait au détriment des résultats, qui sont rarement acceptables, un compromis doit être alors trouvé entre ces deux objectifs. Ce compromis doit permettre d'obtenir des résultats acceptables pour un coût de calcul raisonnable. Nous avons essayé de respecter au maximum ce compromis dans le choix des techniques utilisées. Les techniques de segmentation et d'analyse du mouvement dans les images ont été améliorées par l'introduction d'approche multirésolution servant, d'une part, à accélérer la convergence et, d'autre part, à raffiner les solutions. L'accès aux Bases de Données a été accéléré en utilisant un SGBD.

Des fonctionnalités relatives à l'accès aux bases de données vidéo ont fait aussi l'objet de notre étude. Ces fonctionnalités sont relatives à l'amélioration de la consultation de la base et de la présentation du contenu de la vidéo qui est l'objectif essentiel de notre étude. Elles concernent essentiellement le partitionnement temporelle de la vidéo, ainsi que la génération de résumés visuels du contenu de chaque vidéo.

Le présent document est organisé de la manière suivante :

- Le premier chapitre constitue une introduction aux systèmes de recherche et d'indexation de la vidéo. Une comparaison entre les systèmes d'indexation manuelle textuelle et les systèmes d'indexation automatiques de la vidéo est présentée en montrant les limites des systèmes manuels textuels et en spécifiant les aspects fondamentaux de chaque type.
- Le découpage temporel de la vidéo en plans élémentaires est présenté dans le chapitre 2. Dans ce chapitre nous avons présenté différents aspects du montage vidéo, ainsi que des techniques servant à la structuration temporelle de la vidéo.
- Les techniques de segmentation spatiale par un critère d'intensité d'image sont détaillées dans le chapitre 3. La technique retenue qui est la segmentation par les champs markoviens est présentée, où la résolution du problème de segmentation revient à optimiser le champ de Markov défini sur l'image. Les algorithmes d'optimisation déterministe nécessitent l'emploi de techniques d'initialisation multigrille qui sont présentées vers la fin.
- Dans le chapitre 4 nous introduisons la segmentation spatio-temporelle qui est basée sur l'utilisation des modèles polynomiaux de description du mouvement apparent. Les méthodes d'estimation des paramètres de ces modèles sont détaillées. Et vers la fin, les différentes approches de segmentation au sens du mouvement sont présentées.
- Le chapitre 5 présente l'étude conceptuelle qui nous a mené à la réalisation du système VX (Vidéo eXplorer). Le schéma synoptique du système ainsi qu'une description des différentes parties qui le constituent y sont décrites.
- La mise en œuvre du système VX ainsi que les résultats expérimentaux obtenus sont présentés dans le chapitre 6.
- À la fin, nous présentons les conclusions et les perspectives qui se dégagent de cette étude.