

1089

E.N.S.S.I.B.

Ecole Nationale Supérieure
des Sciences de l'Information
et des Bibliothèques

D.E.A. "Sciences de l'Information et de la Communication"
option "Informatique Documentaire"

Note de synthèse

Les systèmes de recherche
de documents
multimédia

Christine MICHEL

sous la direction de M LAFOUGE

1994-1995

E.N.S.S.I.B.
Ecole Nationale Supérieure
des Sciences de l'Information
et des Bibliothèques

D.E.A. "Sciences de l'Information et de la Communication"
option "Informatique Documentaire"

Note de synthèse

**Les systèmes de recherche
de documents
multimédia**

Christine MICHEL

sous la direction de M LAFOUGE

1994-1995

SOMMAIRE

Méthodologie.....	1
Introduction.....	2
Index des abréviations.....	4
<u>1- Les premiers systèmes de recherche d'images et limites.....</u>	5
1-1- Les problèmes liés à l'indexation de documents multimédia	5
1-1-1- La subjectivité.....	5
1-1-2- Le facteur temps	6
1-2- Les problèmes liés à la gestion de documents multimédia.....	6
1-2-1- SGBD conventionnels utilisés comme IRS (système de recherche d'image).....	6
1-2-2- Systèmes graphiques ayant des fonctionnalités de bases de données.....	6
1-2-3- Bases de données étendues.....	7
1-2-4- Bases de données extensibles.....	7
1-3- Limites et ouvertures.....	7
<u>2- L'analyse de l'image fixe.....</u>	9
2-1- Le processus de compression d'images.....	9
2-2- La compression "intelligente"	10
2-2-1- La multirésolution.....	10
2-2-2- La compression sémantique.....	11
2-3- Une méthode d'indexation automatique présentée par Rabitti [RAB91].....	12
<u>3- L'analyse de l'image animée.....</u>	14
3-1- La segmentation.....	14
3-2- Le facteur temps	15
<u>4- Les nouveaux systèmes.....</u>	16
4-1- Amélioration de la requête QBD (Query By subjective Description).....	16
4-2- Amélioration de l'interface visuelle. De nouveaux types de requêtes.....	18
4-2-1- QPE (Query by Pictural Exemple).....	18
4-2-2- QVE (Query by Visual Exemple).....	18
4-2-3- Le projet QBIC (Query By Image Content).....	19
4-2-4- L'effet de flou.....	20
4-3- Le futur les systèmes orientés objets	20

Conclusion	23
Bibliographie	24

Méthodologie

Le sujet initial de ma note de synthèse portait tout d'abord sur l'évaluation des multimédias. Je devais chercher s'il n'y avait pas eu des modèles mathématiques pour évaluer la pertinence de la recherche, dans des systèmes de recherche multimédia. J'ai questionné diverses bases comme LISA, ou INSPECT par l'intermédiaire du serveur DIALOG avec comme mots clé "evaluation", "information retrieval" et "multimédia". Le système renvoyait un nombre faramineux de références. Le terme "evaluation" n'était pas assez discriminant, je l'ai remplacé par "pertinence" et "relevance". Beaucoup trop de références étaient encore proposées, un tri a été fait par la date de parution, entre 1991 et 1995.

Sur la centaine de références sélectionnées, aucun article ne présentait d'évaluation par des modèles mathématiques, les auteurs entendaient par "evaluation" ou "pertinence" un jugement de valeur sur le travail qu'ils présentaient. De plus, le terme multimédia, très diversifié, était souvent placé dans le titre, alors que le travail ne présentait pas vraiment un système typiquement multimédia.

J'ai refait une interrogation sous DIALOG avec des mots clés plus précis. J'ai choisi "mathematics models", "user's relevance", "user's friendliness", et bien sûr "multimédia", "information retrieval" et "evaluation". Seulement sept références étaient sélectionnées, et aucune ne traitait du sujet.

Devant le peu de succès obtenu avec l'interrogation en ligne, j'ai choisi de parcourir directement les magazines et livres à ma disposition. En effet, les résumés proposés par les banques de données semblaient très attractifs, cependant, à la lecture, les articles s'avéraient traiter les sujets proposés de manière totalement différente. Ces différentes lectures m'ont montré que les chercheurs n'avaient pas encore mis en place de modèles mathématiques pour l'évaluation de tels systèmes, et pour cause, les modèles de systèmes de recherche présentés étaient encore complètement différents au niveau de la structure de stockage, de la gestion des données, de l'indexation ou de l'interface utilisateurs. Il m'a donc paru intéressant de présenter un état de l'art des recherches actuelles sur les systèmes de recherche manipulant des objets de type multimédia.

Introduction

Multimédia has been the buzzword in the 1990s. The myth promised a truly fantastic combination of images, text, sound and video that would bring about a dramatic change in the way computers are used today.

Voilà comment Trevor Wing commence son article sur les possibilités qu'offre actuellement un système multimédia [WIN94]. Les ouvertures industrielles de Microsoft ou d'Apple sont certes énormes, mais ce n'est pas exactement les "software packages" que nous avons choisi de présenter dans cette étude.

En effet, un gros travail de recherche est actuellement mené concernant les systèmes de recherche de documents de types multimédia et notamment ceux manipulant des images, les **IRS** (Images Retrieval Systems). Pour cela, nous avons choisi de présenter une étude des systèmes de recherche de documents de type images (fixes ou animées), nous avons volontairement ignoré le son.

Les avancées techniques au niveau machine nous permettent maintenant de stocker et manipuler beaucoup plus facilement des données images ou vidéos. La plupart des IRS actuels fonctionnent sur le modèle des systèmes classiques de recherche d'information: l'image est décrite par des mots clés, elle est ensuite manipulée comme n'importe quel fichier textuel. On utilise pour cela les SGBD classiques (Systèmes de Gestion de Bases de Données ¹). Ils sont répertoriés selon trois familles: les SGBD hiérarchiques, relationnels et réseaux. Cependant, pour les données picturales, la place mémoire nécessaire pour le stockage et la manipulation est énorme, de plus une description et une indexation correcte par des mots est beaucoup trop coûteuse en temps. En effet, la quantité d'informations sémantiques contenue dans une image est très diverse. En plus du sujet présenté, l'image peut avoir une valeur symbolique ou poétique, difficilement exprimable par des mots.

De plus, les images ont des caractéristiques propres qu'un texte n'a pas, et qu'il serait dommage d'occulter.

Les deux grands axes de recherche en IRS concernent:

- la gestion physique des données par la création de nouveaux systèmes de base de données multimédia (de type orienté objet principalement)

¹ i.e. DBMS

- la manipulation de l'image elle-même. Faite par des chercheurs en imagerie, le but est de trouver de nouveaux "query langages" et moyens d'accès aux images.

Dans le premier chapitre de ce travail, j'ai essayé de faire le point sur les systèmes actuels de recherche d'images. Je présente divers systèmes de gestion de documents multimédia, les problèmes rencontrés lors de l'indexation par la "translittération"² du contenu de l'image et la prise en compte de la subjectivité de l'indexeur et de l'utilisateur.

Dans la seconde partie, je présente les traitements possibles sur l'image elle-même en vue d'améliorer les IRS, en particulier divers types de compressions et une méthode d'indexation automatique. De la même manière, je présente des travaux liés à la vidéo. Les chercheurs essaient d'extraire automatiquement les points forts d'une séquence, ou de créer des systèmes dans lesquels la gestion du facteur temps permet une manipulation facile et rapide des séquences vidéos (avant, arrière, défilement lent, accéléré ...).

La dernière partie est consacrée à présenter de nouveaux systèmes, les premiers cherchent à améliorer la requête QBD, les autres, par une interface graphiques testent de nouveaux types de requêtes où la recherche se fait directement sur le contenu de l'image.

² traduction du contenu d'une image par des mots

Index des abréviations

SGBD: Systèmes de Gestion de Bases de Données

DBMS: Database management system (nominatif, relationnel, hiérarchique, réseau)

IDM: Image data model (représente les champs de description des images)

IDBMS: Image database management system

IRM: Image retrieval model

IRS: (Images Retrieval Systems)

QBE: Query By Exemple

QBD: Query By subjective Description

QPE: Query by Pictural Exemple

QVE: Query by Visual Exemple

VLSI: Very Large Scale Integration

1- Les premiers systèmes de recherche d'images et limites

Les premiers IRS (*Images Retrieval Systems*) sont totalement calqués sur les systèmes de recherche de documents textuels autant au niveau de la gestion des données que de l'indexation. En effet, le document multimédia est décrit textuellement par un opérateur humain, des mots-clés sont extraits et constituent les index d'accès, l'utilisateur formule textuellement une requête de type QBD (Query By subjective Description) pour obtenir le document souhaité. De plus, la recherche et la gestion du document se fait grâce à des systèmes proches des SGBD classiques.

Comme nous allons le voir, divers problèmes en découlent. Nous verrons tout d'abord que ce type d'indexation appauvrit considérablement le contenu sémantique du document multimédia, et ensuite que les SGBD classiques ne sont pas adaptés à la gestion de documents multimédia.

1-1- Les problèmes liés à l'indexation de documents multimédia

1-1-1- La subjectivité

"*A picture is worth a thousand words*". Voilà la phrase clé quand on parle des problèmes de recherche de documents visuels, elle concerne plus particulièrement les problèmes liés à l'indexation.

"La partie la plus délicate de la description de l'image est bien entendu ce qu'on peut regrouper sous le terme d'indexation et qui concerne la "translittération"³ du contenu de l'image. Cette opération se heurte en effet à l'impossibilité théorique de traduire par des mots et l'imperfection inhérente à cette traduction". [MEL88]

Pour J Chaumier, la traduction de façon régulière et répétitive d'une image en langage symbolique est impossible. En effet le problème essentiel réside dans la finesse de l'analyse; comment ne pas aller trop loin dans la description en étant sûr de ne rien oublier d'important.

Les premières analyses traitaient uniquement le "dénoté", présentation du contenu objectif ou anecdotique de l'image (qui, quoi, quand, où, comment). On a tenté ensuite de faire une analyse "connotée", c'est à dire en essayant d'extraire le message symbolique, affectif ou poétique de

³ voir introduction

l'image, l'impression qu'elle dégage. Que se soit pour l'une ou l'autre des deux écoles, le problème général de l'indexation de documents multimédia concerne donc la subjectivité, à la fois de l'indexeur, mais aussi de l'utilisateur [RES89].

Les recherches s'orientent donc vers des systèmes permettant de "faire voir" les images (ou "faire entendre" la bande son) à la machine.

1-1-2- Le facteur temps

Pour coordonner la synchronisation des séquences vidéos, il est important d'inclure un attribut temporel dans la définition des caractéristiques. Le challenge technique n'est pas vraiment difficile notamment avec les nouveaux systèmes orientés objets. Nous en montrerons un exemple extrait de [OZK94].

1-2- Les problèmes liés à la gestion de documents multimédia

Nous allons tout d'abord faire un bref descriptif des premiers systèmes de gestion de documents multimédia. Comme pour des SGBD classiques, ces systèmes comportent un LDD (langage de description des données), un LMD (langage de manipulation des données), un SGF (système de gestion de fichiers) et éventuellement un module de télétraitement pour les manipulations en ligne.

1-2-1- SGBD conventionnels utilisés comme IRS (système de recherche d'image)

Ce type d'approche très populaire a été développé par les chercheurs en informatique orientés vers les bases de données.

Les données images sont stockées sous forme d'une table relationnelle (on dit que les données sont alors **formatées**). Chaque image y est représentée par des attributs ou mots clés. Ils permettent de faire les requêtes usuelles des SGBD (requêtes SQL le plus souvent).

2- Systèmes graphiques ayant des fonctionnalités de bases de données

Les renseignements à propos de l'image (date d'acquisition...) sont stockés en début de fichier, ainsi que l'interprétation qui sera donnée à

l'utilisateur (texte bref qui décrit l'image). Les images seront manipulées comme des données formatées ou non.

Le type de questions formulées pour des données formatées est similaire à celui décrit précédemment (on interroge sur les attributs stockés en début de fichier). Il est aussi possible de questionner en recherchant des données non formatées. Une série de commandes est accessible, elle permet d'effectuer des recherches par reconnaissance de formes, ou type d'images utilisées...

3- Bases de données étendues

On a imaginé des extensions d'une base de donnée de type relationnel pour les images contenant des objets géométriques, par exemple les images produites industriellement grâce à la PAO ou CAO. Le langage de requête et la stratégie de recherche est de type DBMS (SQL), mais le pouvoir d'action est largement augmenté par la possibilité d'attribuer le type d'objet géométrique recherché.

4- Bases de données extensibles

Une base de données conventionnelle contient seulement quelques types de données. L'idée est de permettre à l'utilisateur de créer lui même ses propres types de données. Les données peuvent être représentées sous des formes formatées (voir plus haut), non formatées, structurées (liens hiérarchiques) ou complexes. Le langage est étendu pour accepter ces divers types. Beaucoup de chercheurs travaillent actuellement sur les stratégies de recherches concernant ces nouvelles caractéristiques.

1-3- Limites et ouvertures

Un des gros problème engendré par cette méthode (comme pour un système classique indexé manuellement) concerne la représentation sémantique du document. Bien souvent l'utilisateur ne récupère pas un document pertinent car l'indexation est incorrecte. Nous verrons dans les chapitres suivants les recherches qui sont en cours sur le traitement et l'analyse d'images ou de vidéos. Certains chercheurs travaillent d'ailleurs sur des applications de reconnaissance de formes ([PET94], [PEN94], [GUD94]), ainsi que sur l'indexation automatique d'images [RAB91].

De plus, à l'usage, on s'est aperçu que de tels systèmes n'étaient pas ergonomiques du tout. En effet, Belkin [KRA94] considère le modèle standard comme inapproprié et suggère une nouvelle approche centrée sur l'utilisateur: *"The user must be considered the central component of*

the IR systems, and interaction [...] is the central process of IR". Fox [KRA94] considère qu'il faut créer un nouveau domaine d'activité de recherche: l'accès à l'information via un réseau multimédia. De récentes publications montrent que le sujet "classique" de recherche sur l'évaluation et les méthodes statistiques est remplacé petit à petit par des recherches sur des systèmes basés sur l'utilisateur, la facilité des questionnements, des systèmes d'aide intelligents ou encore l'ergonomie des logiciels et les recherches multimodales.

L'un des principaux axes de recherche actuel concerne le questionnement par le contenu. Divers types de requêtes allant dans ce sens ont été imaginées pour créer un système plus convivial.

De plus Gudivana considère que les langages de requêtes de type SQL ou QBE (Query By Exemple) ne sont pas naturels pour manipuler des images et la pertinence de la recherche s'en ressent. Il a imaginé un système, AIRS (Adaptative Image Retrieval System), basé sur la représentation logique, qui résout le problème de l'interprétation des données. [GUD94] page 424.

Un autre problème est soulevé: l'implémentation au coup par coup d'applications en fonction de besoins spécifiques. Cela conduit à des IRS spécifiques à certains domaines et n'ayant aucune portabilité dans des domaines plus généraux.

Il était donc nécessaire de travailler sur des systèmes plus généraux ayant une architecture compatible avec les caractéristiques propres des images et une utilisation simplifiée pour l'utilisateur, Nous verrons que les chercheurs se sont orientés vers les systèmes orientés objets [OSK94].

2- L'analyse de l'image fixe

Dès lors que des documents comportent des images ils représentent des fichiers extrêmement volumineux, longs à transférer et difficiles à traiter. "Pour donner un ordre de grandeur, une image 1024 x 1024 pixels codée sur 24 bits nécessite 3 Mbits d'espace d'archivage et 7 mn de transmission sur un canal à 64 bits par seconde. Si cette image peut être compressée selon un facteur de 10, le fichier ne représente alors plus que 300 Kbits ce qui devient plus acceptable et le temps de transmission chute à 6 secondes" [DUP94]. Nous voyons bien que les documents multimédia nécessitent d'être compressés. Plusieurs travaux s'orientent vers une compression que l'on pourrait qualifier "d'intelligente" car, au contraire de la compression classique, il est encore possible d'effectuer des traitements (analyses comparatives) sur l'image compressée.

D'autres travaux, qui n'ont aboutit pour l'instant qu'à des prototypes, sont présentés. Le premier décrit un système d'indexation automatique des images, les suivants décrivent des systèmes qui améliorent la requête sur image, en optimisant le requête usuelle de type QBD, ou en imaginant d'autres types de requêtes propres à l'image.

2-1- Le processus de compression d'images

Deux ensembles de techniques sont utilisées [DUP94]:

- *compression sans perte d'information*. La technique la plus utilisée est celle qui consiste à analyser une chaîne de pixels ou de caractères pour trouver les répétitions et les coder. Cette technique ne donne pas des taux de compression très élevés. Une autre technique de codage dite statistique (code de Huffman), consiste à établir un dictionnaire hiérarchique des occurrences rencontrées, selon leur fréquence, pour leur faire correspondre une représentation moins consommatrice de bits (les occurrences les plus élevées ont les codages les plus courts). Les taux de compression varient de 1 pour 1,2 à 1 pour 2,5.

- *compression avec perte d'information*. On observe trois techniques de compression. Les deux premières ont conduit à la normalisation JPEG. Il s'agit d'analyser la corrélation et les fréquences par des traitements de type FFT ⁴, DCT ⁵ ou KLT ⁶ (projections sur

4 Fast Fourier Transform

5 Discret Cosine Transform

6Kahrunen Loewe Transform

d'autres espaces de représentation). La seconde technique découlant de la première consiste à coder l'image suivant différentes zones (i. e. l'image est divisée en blocs) qui sont analysées chacune selon leur nature par telles ou telles transformations décrites ci-dessus. La dernière technique s'appuie sur les **fractales**. Elle consiste à coder des formes élémentaires répétitives et récursives, tout en fixant la profondeur des itérations et donc la taille des formes élémentaires. Les taux de compression sont très élevés (ils passent de 1 pour 10 à 1 pour 100). Les temps de calcul étant beaucoup plus lents pour la compression que pour la décompression, cet algorithme est essentiellement utilisé pour la réalisation de CD-ROM.

2-2- La compression "intelligente"

Sous le terme de compression intelligente nous rassemblerons des techniques qui permettent de travailler sur l'image compressée.

1- La multirésolution

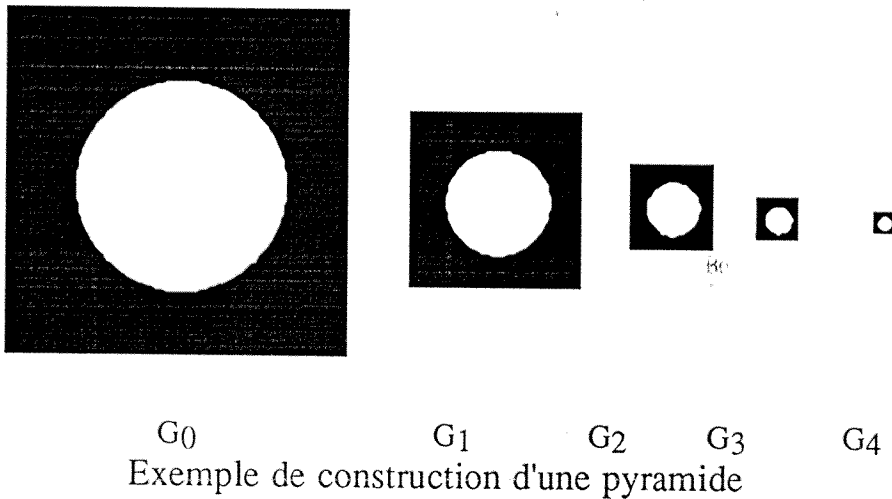
Les traitements de type multirésolution se sont inspirés des travaux sur la vision humaine [JOL94]. En effet, pour repérer et stocker une information visuelle, le cerveau humain la décompose préalablement en différentes images mentales suivant leur différentes fréquences ⁷. On peut mathématiquement décomposer une image en somme de composantes de fréquences différentes.

La théorie des pyramides

Jolion a imaginé un traitement multirésolution particulier, il va construire une pyramide (suite finie) d'images, constituée par la décomposition fréquentielle d'une image source. Le processus consiste à appliquer un filtre ⁸ sur l'image G_i . Ceci a pour effet de créer une image G_{i+1} de taille deux fois plus petite, qui est une concentration de l'information contenue dans l'image G_i ; les hautes fréquences (correspondant aux détails) sont petit à petit éliminées. La pyramide peut aller jusqu'à une image de 8x8 pixels mais on peut stopper le processus avant.

⁷La distribution fréquentielle donne un ensemble de valeurs qui correspondent à la contribution relative des différentes variations de contraste dans l'espace. Une image avec des variations progressives, un dégradé linéaire de couleurs par exemple, aura des composantes en basses fréquences. Une image avec beaucoup de différences lumineuses, par exemple un maillage serré, aura des composantes situées en hautes fréquences.

⁸ matrice 3x3 appelée "gaussienne" car les coefficients sont plus forts au centre



La multirésolution offre de nombreux avantages. Elle permet de manipuler l'information en utilisant différents niveaux d'analyse. Les traitements effectués peuvent être soit globaux (recherche d'une voiture dans un paysage), l'analyse se faisant sur les hauts niveaux de la pyramide, soit locaux (lecture de la plaque d'immatriculation), l'analyse portant alors sur les bas niveaux de la pyramide.

De plus, les algorithmes de calcul des pyramides sont d'une complexité raisonnable ce qui donne des temps de traitement courts.

Utilisation

La multirésolution est couramment utilisée pour le traitement et l'analyse d'images. Elle fournit divers outils rapides de rehaussement du contraste, de discrimination des textures...

Pour les systèmes de recherche d'images (IRS) elle peut se révéler très utile pour améliorer les temps de comparaison d'images quand on utilise des requêtes de types QPE, QVB ou QVE [KAT92](une explication plus détaillée est présentée dans le chapitre suivant). En effet, les informations pertinentes de l'image sont conservées dans les niveaux supérieurs de résolution. On peut donc imaginer que les traitements de comparaison se fassent sur des images de taille deux à quatre fois plus petite. Le gain de temps serait alors du même ordre.

2-2-2- La compression sémantique

Principes

Un des problèmes importants pour l'utilisation de systèmes multimédia est, comme nous venons de le voir, d'extraire et de représenter le contenu d'images ou de séquences vidéo, ceci en vue de les rassembler de manière cohérente et en permettre l'utilisation. Pour une grosse base il est impossible d'annoter chaque image comme il a été présenté dans [GUD94]. Pour répondre aux requêtes, la solution serait de "faire voir"

les images à l'ordinateur. Devant la taille impressionnante d'une base de données d'images, Pentland [PEN94] et son équipe ont eu l'idée de les compresser en se basant sur leurs contenus sémantiques. L'idée est de ne garder que "l'objet" le plus important de la scène. Pour ce faire, les caractéristiques les plus discriminantes sont mesurées: luminosité, contraste, textures (orientation, périodicité,...) mais aussi géométrie de la scène. Les différents objets sont identifiés et extraits. Ils sont stockés dans une image que l'on pourrait appeler "index", et sur laquelle se fait la recherche. Par exemple, sur une photo de portrait, seul le visage restera (information pertinente pour cette image), tout le fond sera rempli en noir. Dans une image plus complexe, il est possible d'extraire dans une première image, les objets en premier plan, et dans une seconde, ceux en second plan. L'image ainsi codée prend beaucoup moins de place que l'image initiale, et les recherches sont beaucoup plus rapides.

Utilisation

Cette technique est notamment utilisée pour repérer les *keyframes* (images les plus représentatives d'une séquence vidéo).

Une autre utilisation en est le butinage. Le système peut reconnaître des séquences similaires sémantiquement parlant et ainsi répondre à des requêtes telles que "montrez moi des images de ce type ci" ou "montrez moi des images qui ressemblent à celle-là". Il a été reconnu que le système était capable de reconnaître une personne à 95% et d'identifier des empreintes à 99,9%.

2-3- Une méthode d'indexation automatique présentée par Rabitti [RAB91]

Rabitti définit en premier lieu des classes d'images. En fonction de celles-ci, un processus automatique d'indexation est mis en route.

La première analyse appelée analyse de bas niveau, traite des images constituées de bit-map ou construites à partir de primitives graphiques. Elle permet d'extraire les objets principaux, définit leurs caractéristiques (géométriques) et les classe selon un graphe relationnel (un carré est une forme particulière de rectangle, qui est lui-même une forme particulière de quadrilatère). Les algorithmes utilisés ont une complexité raisonnable (complexité polynomiale) mais donnent des résultats pauvres au point de vue sémantique.

Une analyse plus poussée, analyse de haut niveau, est ensuite faite. Un algorithme de recherche récursif, travaillant avec les objets sélectionnés précédemment, permet d'analyser les objets les plus complexes.

Le résultat final pour une image donne les objets qui la composent ainsi que leurs degrés (haut ou bas niveau). Par ces objets on va pouvoir accéder à l'image, soit par l'adresse directe, soit par l'intermédiaire

d'index. Le processus de recherche se fait sur le contenu de l'image. Ce processus est applicable pour un corpus d'images simples et géométriques. Rabitti l'a testé avec une base d'images de plans d'appartements.

3- L'analyse de l'image animée

Les travaux sur l'image animée s'orientent vers la recherche de techniques visant à segmenter, c'est à dire à ne présenter à l'utilisateur que les "morceaux" importants d'une séquence.

Le problème du facteur temps quant à lui a été grandement simplifié par l'utilisation de bases de données objets.

3-1- La segmentation

1- Principe

La segmentation macroscopique d'un document en unités documentaires est essentielle pour la navigation d'un utilisateur. En effet, il est beaucoup trop long de visionner chacun des documents qui répondent à une question pour tester leur pertinence.

Beaucoup de chercheurs ont travaillé sur la reconnaissance de séquences (pour des vidéos) par la segmentation temporelle. Le but de ce type de segmentation est de séparer le contenu sémantique d'un enregistrement vidéo en divers enregistrements de taille inférieure, et ainsi d'en permettre l'extraction et la présentation interactive aux utilisateurs.

La segmentation peut se faire sur les deux composantes d'une vidéo: l'image et le son.

La majorité des travaux de segmentation sont orientés sur l'image et en particulier sur la recherche de la *sequence-shot*, unité fondamentale de segmentation d'une vidéo. On a d'abord cherché à identifier les séquences grâce à l'orientation des différentes textures présentes dans les images [AIG94]. Cela donne de bons résultats mais les calculs coûtent beaucoup trop cher. Les techniques à l'étude actuellement sont basées sur la reconnaissance de formes géométriques (géométrie projective).

La segmentation par la bande son est beaucoup moins significative pour la discrimination de séquences. Les études en cours portent sur l'identification de discours, musiques ou sons naturels. Les calculs sont peu onéreux mais la méthode nécessite un enregistrement sans bruits additionnels.

3-1-2- Utilisation

Aigrain [AIG94] a développé deux logiciels de traitement de l'image animée en se servant de la segmentation.

Création de *storyboards*

Une utilisation de la segmentation est la production de storyboards, courts résumés des séquences les plus importantes d'une vidéo. Créés à l'aide du logiciel *Videoboard*, ils peuvent servir à coder des catalogues ou être utilisés pour le butinage visuel.

Accès interactif

La segmentation permet aussi de faire un accès direct aux enregistrements vidéo par l'intermédiaire du logiciel *Videoline*. L'ensemble des vidéos est accessible graphiquement par une icône (image pertinente de la séquence). Six sequence-shots sont présentées à l'utilisateur lui permettant de ne visionner que la partie de la séquence qui l'intéresse.

3-2- *Le facteur temps*

Pour coordonner la synchronisation des séquences de données il est important d'inclure un attribut temporel dans la définition des caractéristiques.

Le modèle utilisé, *interval-based temporel data model*, garde une structure relationnelle [OZK94]. La structure objet/noeud est maintenue, les feuilles représentant les différents types de médias, les noeuds internes contenant des attributs de temps. Les relations NOEUD, SOUS-NOEUD, NOEUD-TERMINAL, nous montrent comment les données sont codifiées au niveau temporel.

NOEUD (No-Noeud, Type, Durée, Sujet) décrit l'objet (le noeud) dans son entier. La variable type va permettre d'assembler les sous-noeuds (film = type séquentiel, multitude d'images = type parallèle...).

SOUS-NOEUD (No-Noeud, No-ss-Noeud, Index-f, Index-r, Delta-f, Delta-r) décrit les entités primaires (ex: une image pour un film). Index donne l'ordre de classement des sous-noeuds (f = forward, r = reverse). Delta permet de définir le temps de défilement en avant ou arrière.

NOEUD-TERMINAL(No-Noeud, Media, Fichier) localise chacun des noeuds.

4- Les nouveaux systèmes

4-1- Amélioration de la requête QBD (Query By Subjective Description)

La requête de type QBD présentée dans le premier chapitre a beaucoup été dénigrée car elle simplifiait ou occultait la valeur sémantique d'une image au profit unique de sa valeur informationnelle. Des recherches ont donc été menées pour pallier à ce biais. Deux études sont présentées, la première inclue de nouveaux liens dans le système de gestion des images, la seconde prend en compte la subjectivité de l'utilisateur.

4-1-1- Introduction de liens sémantiques

Les premières études ont été faites à partir de dictionnaires. Tout a commencé en 1968 avec Quillian. Il considérait un système de recherche textuelle contenant la relation FAIT-PARTIE-DE (on avait par exemple "basset" FAIT-PARTIE-DE "chien"). Divers systèmes en ont découlé. On peut citer WordNet en 1990 qui travaillait plus sur les relations de synonymie, ou CYC qui s'intéressait au sens commun des mots (i.e. "*the factual and heuristic knowledge, much of it usually left unstated, that comprise "concensus reality": the things we assume everybody already knows*".)

Dès 1980 ce type de recherche s'est orienté vers l'image.

Le système de Chakravarthy [CHA94] fonctionne comme un IRS classique, à la différence près qu'il introduit dans le système des relations sémantiques ou de synonymie, et ainsi crée des liens sémantiques entre les images. Le processus est simple.

Pour Chakravarthy, il existe trois types de "desideratum" (ce que l'utilisateur recherche):

- l'objet comme desideratum

L'objet représente ici un sujet physiquement présent dans l'image. Diverses relations sémantiques peuvent être utilisées: FAIT-PARTIE DE, EST-MEMBRE-DE, JOUE-LE-ROLE-DE, EST-ASSOCIE-AVEC.

Exemple: trompe EST-ASSOCIE-AVEC éléphant.

- l'action comme desideratum

L'action est celle effectuée dans l'image. Elle est représentée par un verbe pour l'indexation.

Les relations utilisées sont EST-HERITE-DE, EST-CAUSE-DE, EST-UNE-ACTION-TYPIQUE-DE.

Exemple: lancer EST-CAUSE-DE tomber

- les relations sémantiques comme desideratum

On ne cherchera pas quelque chose d'explicite dans l'image mais quelque chose qui peut s'en déduire.

On peut ajouter aux relations précédentes: EST-LOCALISE-DANS, EST-UN-OBJET-DE, EST-UNE-PROPRIETE-DE.

Exemple: une image montrant quelqu'un en train de manger un bonbon engendre la relation: mangeable EST-UNE-PROPRIETE-DE bonbon.

Évaluation

Ce système a été évalué sur la base de 1566 photos. Un texte argumentatif a été enregistré pour chaque photo. WordNet a été utilisé pour créer les relations sémantiques. Ce processus a permis de créer 7180 index pour la recherche des photos. 200 de ces index ont été analysés, et ils ont permis de retrouver 930 photos. Sur les 930, 244 ont été jugées inadéquates, 481 ont été jugées proches mais incorrectes, et seulement 244 jugées correctes. Selon Chakravarthy, ce faible taux de réussite venait bien sûr de mauvaises relations sémantiques, mais aussi d'un problème d'ambiguïté dans la représentation.

4-1-2- Par évaluation de la subjectivité de l'utilisateur

Dans le système de Kato [KAT92], l'utilisateur n'a qu'à déterminer l'importance (pour son desideratum) de certains critères descriptifs de l'images, fixés initialement par le système, par des poids. Deux types de critères lui sont présentés:

- *des critères d'arrangement des couleurs*: En effet, une étude a montré que l'impression dominante d'une image était donnée par la couleur. Kato évalue les critères de couleurs selon la distribution RGB (arrangement global des couleurs) d'une part, et la corrélation RGB (combinaisons de couleurs).

- *des critères subjectifs*: d'impression que laisse une image (claire, douce, couleurs chaudes/froides, tableau moderne, romantique, japonisant, ...).

Pour palier le biais dû aux différences de représentations, le système évalue préalablement à toute réponse, les critères subjectifs de l'utilisateur. Pour ce faire, un échantillon d'images lui est présenté, il va devoir choisir les poids à attribuer aux différents critères.

La distribution, la corrélation et l'analyse des poids faites pour chaque image, vont permettre de construire un espace de caractéristiques unifiées (UF) donnant une pondération personnelle à chaque image (index personnels). La comparaison des critères de la requête et des images de la base se fera dans cet espace.

Une étude évaluative a été menée avec le logiciel ^{by}ART MUSEUM qui comprend une base de reproduction de peintures. Les résultats ont montré que l'espace UF gardait le sens de la couleur.

4-2- Amélioration de l'interface visuelle. De nouveaux types de requêtes

Nous allons décrire deux modèles de requêtes utilisant des interfaces graphiques. Dans les deux cas le système va comparer une image référence, donnée par l'utilisateur, avec les images stockées dans la base. L'image exemple va être considérée comme une clé picturale. Les deux modèles se différencient par l'obtention de cette image référence.

1- QPE (Query by Pictural Exemple)

L'image est indexée automatiquement suivant les caractéristiques géométriques du sujet présenté. Pour la requête, un ensemble de formes géométriques est proposé à l'utilisateur, il peut les assembler pour spécifier quel type d'image il souhaite avoir.

Ce système hypermédia code les données uniformément. De plus, il permet de faire une indexation subjective en associant les données à différents domaines. Pourtant il comporte un biais. L'utilisateur novice n'a pas la même représentation visuelle de ce qu'il souhaite avoir, que le système. Pour y pallier, le système devrait pouvoir évaluer le modèle mental de l'utilisateur. Des études sont en cours au niveau de l'interprétation et de l'expression des informations visuelles.

Une des nécessités pour l'interface visuelle était donc de construire des systèmes permettant une recherche basée sur la vision subjective de l'utilisateur sans effort d'indexation préalable.

Dans cette perspective Kato [KAT92] a travaillé sur l'interrogation de type QVE (Query by Visual Exemple).

2- QVE (Query by Visual Exemple)

La requête QVE se fait à partir d'une image exemple existante ou dessinée par l'utilisateur. Le système va chercher une image "subjectivement" similaire.

Kato [KAT92] a travaillé pour améliorer ce modèle. Comme il l'a fait pour la requête QBD, il a pensé créer un espace des représentations de l'utilisateur (i.e. redéfinitions des index de la base) dans lequel les images seraient comparées.

Il définit une image selon deux critères:

- *des critères graphiques*: la distribution spatiale, la distribution fréquentielle, la corrélation et le contraste.

Une étude expérimentale de recherche par des critères graphiques a été faite sur une base (TRADEMARK) de 2000 symboles graphiques. Le ratio de rappel a été de 100% (sic).

- *des critères subjectifs*: L'utilisateur doit choisir dans un tableau de symboles graphiques le type d'image souhaitée.

Pour éviter toute ambiguïté de définition, le système fait, préalablement à toute recherche, une évaluation des jugements de similarités de l'utilisateur. Comme pour la requête QBD, il va devoir classer un échantillon d'images dans divers tableaux de symboles graphiques. Une analyse selon les critères définis précédemment va permettre la construction d'un espace de subjectivité SF donnant une pondération personnelle à chaque symbole (index personnels). La recherche se fait alors par comparaison des distances entre symboles, dans l'espace SF. Le même type d'étude sur TRADEMARK, a montré un ratio de rappel de 98%.

Remarque

Bien que le taux de rappel soit très (trop) bon pour ces types de recherches, il ne faut pas oublier que la comparaison pixel par pixel est longue. Dans ces cas, l'analyse en multirésolution (présentée plus haut) se révèle être très utile.

3- Le projet QBIC (Query By Image Content)

Comme Gudivana et Raghavan [GUD94], l'équipe d'IBM [PET94] a cherché des moyens plus "naturels" de chercher des images dans une base de données (donner un exemple, image similaire, dessin approximatif,

formes géométriques, texture, couleur...). Ainsi est né le projet QBIC (Query By Image Content). Le contenu extrait est par exemple la couleur, le contraste, la taille, l'orientation, la position des objets. Ici, seules les caractéristiques structurelles de l'image, et non pas les caractéristiques sémantiques (comme chez Pentland [PEN94]) sont extraites. Ce projet vise non seulement une amélioration des recherches dans des systèmes de type multimédia, mais aussi à terme l'implémentation d'outils de navigation et d'indexation automatique d'images.

Le logiciel Ultimedia Manager, utilise la technologie QBIC alliée à une recherche usuelle par mots-clés. Ultimedia Manager fonctionne comme un IRS à la différence près qu'il inclut aussi un module d'indexation automatique (Batch Image Classifier), d'analyse d'image (Image Analyser) et de création de catalogue (Image Classifier). La classification sémantique (animal, maison...) se fait manuellement.

4- L'effet de flou

Pour remédier à l'effet de flou engendré par une question "large", Krause [KRA94] et son équipe ont développé un système, WING-IIR. Il combine une interface Windows où l'on questionne en langue naturelle, et des outils de recherche d'informations graphiques (présentation visuelle des données, possibilité de QBE). WING-IIR est un prolongement de WING-M2 auquel est apporté tout le côté graphique. WING-IIR pourvoit au manque de précision des systèmes usuels d'IRS. La recherche d'information se fait la plupart du temps de manière graphique ce qui optimise l'ergonomie du système, cependant la langue naturelle reste une bonne alternative dans le cas de questions floues.

Dans le cadre de WING-GRAPH (l'interface de recherche graphique) trois types de questions ont été proposées.

- la requête QBD qui définit la recherche sans faire référence à aucune donnée type QBD.
- la requête QVE qui est corrélée à l'image affichée.
- la requête directe (QPE) de l'utilisateur qui dessine le type de forme cherchée.

Un test mené avec 9 experts industriels a montré que la requête QVE était la meilleure amélioration apportée par l'interrogation graphique.

4-3- Le futur: les systèmes orientés objets

Ce type de système offre de nouvelles perspectives par l'amélioration des liens sémantiques.

Ozkarahan [OZK94] décrit un modèle conceptuel de base de données multimédia construit à partir d'un modèle Entité/Relation, qu'il a modifié en une base orientée objet. Les tests ont été réalisés sur la base de données

médicales d'un hôpital. Elle comprenait du texte (textes anatomiques et études de cas de malades), des enregistrements audio des consultations et bien sûr des images (radios, photos anatomiques ...).

Les enregistrements audio ayant été retranscrits sous forme de texte, le véritable challenge était donc le traitement des images.

L'**indexation** des images se fait en trois temps [OZK94]:

- l'image est convertie sous forme conceptuelle. Des mots-clés (*keywords*) sont choisis pour décrire ses caractéristiques et son contenu sémantique

- Les *keywords* sont collectés et triés. Ils vont permettre de définir les objets, leurs relations et leurs attributs à inclure dans la base orientée objet.

- les termes d'indexation sont choisis de la même manière que pour un texte.

Pour trouver le document répondant à la requête, trois solutions étaient envisagées:

- la recherche par index inverse

- la recherche sur un fichier de signature⁹

- la recherche avec pondération (le document est représenté par un vecteur)

Ozkarahan a choisi de faire une **recherche avec pondération** car elle permet une approche sémantique. On peut trouver une présentation plus approfondie de cette méthode dans l'appendice A page 128. Les poids des index ont été fixés à 1 s'ils apparaissaient dans le document, à 0 sinon.

La requête

La requête de l'utilisateur se fait par une extension de SQL, SQLX, plus adapté aux bases orientées objets. Une série de documents lui sont présentés, une recherche plus en détail de type QVE est possible ensuite.

SQLX possède des attributs de connexion, liens fonctionnels (i.e. n-1) ou "inverses fonctionnels", utilisés pour construire les différents chemins possibles d'un objet à un autre. L'utilisateur n'a plus à s'occuper des liens relationnels nécessaires auparavant dans SQL. De plus, SQLX donne la possibilité d'utiliser des données et procédures abstraites qui font références aux objets et à leurs méthodes. Par exemple il est possible d'effectuer une recherche suivant une méthode ("*invoke*").

SQLX est présenté plus en détail dans [OZK90].

Ozkarahan travaille actuellement à l'élaboration d'un langage de type SQLX plus orienté vers une recherche média [OZK94]. Il entend par là

⁹ Une signature électronique est un train de bits dépendant du contenu du document et d'un mot de passe secret, propre au signataire [DUP94].

une recherche avec un langage de requête spécifique aux objets média comme SQLX peut l'être pour la recherche sémantique. Ce type de recherche est, comme pour la recherche sémantique, intimement lié avec le stockage physique des données. Par exemple, il devrait être possible, sachant la place mémoire que prend une image, de faire des requêtes sur des images compressées. De plus la structure objet permet d'inclure des attributs temporels pour le stockage et la manipulation de séquences vidéo¹⁰.

¹⁰ Voir le chapitre sur l'image animée

Conclusion

Cet état des lieux des travaux sur les systèmes de recherche d'images nous montre que les chercheurs ont bien pris conscience de l'amélioration nécessaire quant à l'ergonomie de ces systèmes. Les IRS qui recherchent sur le contenu de l'image sont très prometteurs et très novateurs. Un gros travail a été fait au niveau de l'architecture et de la gestion des données. En effet, les recherches se sont orientées vers des systèmes propres à l'image et non plus adaptés à partir des systèmes de recherche textuelle. De plus, la programmation objet ouvre de grandes perspectives quant à l'organisation interne.

Signalons aussi l'importance des travaux sur la segmentation et la gestion du temps pour l'image animée qui se révèlent indispensables pour la recherche et la manipulation de la vidéo.

Il faut quand même rester prudent, l'ensemble des IRS présentés ne sont encore que des prototypes. Il faudra encore beaucoup d'améliorations à ces systèmes pour qu'ils puissent être commercialisés à grande échelle.

Bibliographie

[AIG94]: Aigrain P, Joly P. *Discrete visual manipulation user interface video*. Intelligent multimedia information retrieval systems and management. RIAO 94 conference proceedings volume 2, october 1994. p 13-17.

[ALL92]: Allen B. *Cognitive différence in end user searching of a CD-ROM index*. Proceeding of the fifteenth annual international ACM SIGIR conference on recherche and development in information retrieval. 1992. p 298-310.

[BOU93]: Boursier P, Taufour P. *La technologie multimédia*. Edition Hermes. 1993. 185 p. ISBN: 2-86601-344-1.

[CES91]: Cesarini F, Soda G. *A system for managing multimedia dossiers*. Intelligent text and image handling. RIAO 91 Conference proceedings vol 2, april 2-5 1991. p 86-103.

[CHA94]: Chakravarthy A.S. *Toward semantic retrieval of picture and video*. Intelligent multimedia information retrieval systems and management. RIAO 94 conference proceedings, october 11-13 1994. p 676-686.

[CHE91]: Cheiney J.P, Picouet P. *Digital prefix trees: an efficient structure for image data storage and manipulation*. Intelligent text and image handling. RIAO 91 Conference proceedings vol 2, april 2-5 1991. p 885-901.

[DAV91]: Davcev D, Cakmakov D. *An intelligent multimedia information handling system*. Intelligent text and image handling. RIAO 91 Conference proceedings vol 2, april 2-5 1991. p 121-136.

[DUP94]: Dupoirier G. *Technologie de la GED. L'édition électronique*. Edition Hermes. 1994. 223 p.

[LAR94]: Large A, Beheshti J, Breuleux A, Renaud A. *Multimedia and comprehension: A cognitive study*. Journal of the american society for information science. 45(7).Aout 94. p515-528.

[HAN89]: Hannafin M J, Rieber L P. *Psychological foundations of instructional design for emerging computer-based instructional technologies. Part II*. Educational technology reserch and development, 37, p 102-114. 1989.

[HAW91]: Hawamdeh S. Al., Ooi B.C. *Semantic-based query formulation in PAS*. Intelligent text and image handling. RIAO 91 Conference proceedings vol 2, april 2-5 1991. p 916-936.

[HOW92]: Howson B A, David H. *Enhancing comprehension with videodiscs*. Media and methods, 28, p 12-14. 1992.

[GUD91]: Gudivana V.N., Raghavan V.V. *Discovery of conceptual categorie in an image database*. Intelligent text and image handling. RIAO 91 Conference proceedings vol 2, april 2-5 1991. p 901-916.

[GUD94]: Gudivana V.N., Raghavan V.V. *A system for retrieving images by content*. Intelligent multimedia information retrival systems and management. RIAO 94 conference proceedings, october 11-13 1994. p 418-436.

[GUD94-2]: Gudivana V.N., Raghavan V.V. *Picture retrieval systems: A unified perspective*. Manuscript. Departement of computer science, Ohio University, Athen, OH 45701. 1994.

[JOL94]: Jolion J M. *A pyramid framework for early vision..* 1994

[KRA94]: Krause J, Wolff C, Womser-Hacker C. *Multimodality and vagueness in the context of a graphical, object-oriented material information system*. Intelligent multimedia information retrival systems and management. RIAO 94 conference proceedings, october 11-13 1994. p585-597.

[KAT92]: Kato T. *Database architechture for content-based image retrieval*. SPIE volume 1662. Image storage and retrieval systems. 1992. p 112-123.

[KRA94]: Krause J, Wolff C, Womser-Hacker C. *Multimodality and vagueness in the context of a graphical, object-oriented material information system*. Intelligent multimedia information retrival systems and management. RIAO 94 conference proceedings, october 11-13 1994. p585-597.

[KAT92]: Kato T. *Database architechture for content-based image retrieval*. SPIE volume 1662. Image storage and retrieval systems. 1992. p 112-123.

[KRA94]: Krause J, Wolff C, Womser-Hacker C. *Multimodality and vagueness in the context of a graphical, object-oriented material information system*. Intelligent multimedia information retrieval systems and management. RIAO 94 conference proceedings, october 11-13 1994. p585-597.

[KAT92]: Kato T. *Database architecture for content-based image retrieval*. SPIE volume 1662. Image storage and retrieval systems. 1992. p 112-123.

[MEL88]: Melot M. *Problématiques de l'image comme mode de connaissance*. Images et intelligence artificielle dans l'information scientifique et technique. Cours INRIA, 6-10 juin 1988, Bénodet, dir par C Bordes. Rocquencourt. 1988. p 15-30.

[OZK94]: Ozkarahan E. *Multimedia document retrieval*. Information processing and management, vol 31, No 1, 1995. p 113-131.

[OZK94]: Ozkarahan E. *Database management: Concepts, design, and practice*. Englewood Cliffs,NJ:Prentice-Hall. 1990.

[PAV86]: Paivio A. *Mentals representation: A dual coding approach..* New York: Oxford university press. 1986.

[PEN94]: Pentland A, Picard R, Davenport G, Haase K. *Video and images semantics: Advanced tools for telecommunications*. Intelligent multimedia information retrieval systems and management. RIAO 94 conference proceedings volume 2, october 19 1994. p 2-6.

[PET94]: Petkovic D, Yanker P, Niblack W, Flickner M, Hafner J, Lee D, Ashley J (IBM Almaden Research Center, San Jose, CA) and Tung F (IBM Santa Theresa Lab, San Jose, CA) . *Query by image content*. Intelligent multimedia information retrieval systems and management. RIAO 94 conference proceedings vol 2, october 19 1994. p 7-12.

[RAB91]: Rabitti F., Savarino P. *Automatic image indexation and retrieval*. Intelligent text and image handling. RIAO 91 Conference proceedings vol 2, april 2-5 1991. p 864-884.

[RES89]: Resche D. *L'indexation de l'image fixe*. Note de synthèse. Université Lyon I. DESS Informatique documentaire. 1989. 33 p.

[WIN94]: Wing T. *Multimedia - the reality of its possibilities today*. Information services and use. No14. 1994. p 37-40.