

Recherche d'information multimédia : Apport de la fouille de données et des ontologies

Marie-Aude Aufaure *, Marinette Bouet **

* Supélec, Plateau du Moulon, Département Informatique,
F-91192 Gif-sur Yvette Cedex, France
Marie-Aude.Aufaure@supelec.fr
www.supelec.fr/ecole/si/pages_perso/aufaure.html

** LIMOS, UMR 6158 CNRS – Université Blaise Pascal (Clermont-Ferrand II)
Campus des Cézeaux – 24, Avenue des Landais – 63173 AUBIERE Cedex – France
Marinette.Bouet@cust.univ-bpclermont.fr

Résumé. A ce jour, le média image est omniprésent dans de nombreuses applications. Un volume de données considérable est produit ce qui conduit à la nécessité de développer des outils permettant de retrouver efficacement de l'information pertinente. Les systèmes de recherche actuels montrent aujourd'hui leurs limites en raison de l'absence de sémantique. Une voie qui semble intéressante à explorer afin de combler le fossé existant entre les propriétés extraites et le contenu sémantique, est la fouille de données. C'est un domaine de recherche encore immature mais très prometteur. Cet article présente des travaux préliminaires sur la manière de définir de nouveaux descripteurs intégrant la sémantique. Le clustering et la caractérisation des classes obtenues sont utilisés pour réduire l'espace de recherche et produire une vue résumée de la base. La navigation basée sur une ontologie visuelle est un moyen puissant et convivial pour retrouver de l'information pertinente.

1 Introduction

Durant la dernière décennie, un volume considérable de données multimédia a été produit. Ces données sont par essence complexes, non structurées et volumineuses et les applications ayant besoin de rechercher des images pertinentes de manière efficace, de plus en plus nombreuses. Du fait qu'une image ne contient pas directement d'information interprétable de manière automatique, les méta-données vont jouer un rôle très important. L'étape de pré-traitement permet d'extraire un ensemble de méta-données comme : (1) les méta-données relatives au type de donnée multimédia, (2) les méta-données descriptives : nom de l'auteur, date, etc., (3) les méta-données relatives au contenu (sémantique, visuel, relations spatiales) : le contenu visuel est décrit en termes de couleur, forme et texture, et le contenu sémantique est une interprétation de l'image.

Le but est de pouvoir traiter les données du pixel à la connaissance puisque par le vocable « image », on entend image numérique c'est-à-dire une image qui se présente sous la forme d'une matrice de pixels. Il est à noter aussi que le média image ne concerne que les images fixes; les images animées étant dénotées par l'expression « média animation ». Au niveau pixel, des descripteurs visuels sont extraits et les requêtes sont basées sur le contenu. De nombreux travaux existent dans le domaine de la vision par ordinateur sur la partie descripteurs visuels. Dans ce cas, la recherche d'information consiste en une recherche par similarité (**C**ontent **B**ased **I**mage **R**etrieval) basée sur une distance entre les descripteurs visuels extraits des images (Venters et Cooper 2000). Le niveau d'abstraction suivant est

celui des objets ou régions extraits des images. L'utilisateur peut alors sélectionner des objets dans sa requête, de manière à ce que celle-ci soit plus précise. Les relations spatiales sont prises en compte à ce niveau. Le niveau sémantique est dédié à l'extraction et à la génération de méta-données sémantiques, et peut être réalisé à l'aide d'ontologies (Staab et Studer 2004). Enfin, le niveau connaissance utilise le niveau sémantique pour découvrir des relations cachées entre les objets, de la connaissance et pour résumer et caractériser une grande base d'images.

La fouille dans les images est un domaine de recherche récent (Zhang et al. 2001, Simoff et al. 2002, Djeraba 2002) mais n'est pas encore très développé du fait que l'extraction de connaissances à partir des images reste une tâche difficile. Les techniques classiques de fouille de données (Han et Kamber 2001) ont été largement utilisées pour des données alphanumériques. Cependant, dans un contexte multimédia, les bases de données contiennent un volume important de données numériques (les descripteurs) et de données sémantiques (les annotations). Les techniques classiques de fouille de données ne peuvent donc pas être directement appliquées aux images du fait de leur nature non structurée et de leur grande dimensionnalité (Berrani et al. 2002, Oria et al. 2004). Les images (et plus généralement les données multimédia) représentent de nouveaux défis pour l'apprentissage et la découverte de connaissances. Parmi les techniques de fouille de données, les plus utilisées sont le clustering (Jain et al. 1999), les règles d'association (Agrawal et al. 1994) et les réseaux de neurones (Dreyfus et al. 2002). La plupart des approches de fouille d'images se basent uniquement sur les descripteurs visuels. Notre approche vise à combiner le visuel et le textuel en utilisant des méthodes de clustering et de caractérisation. L'idée est de permettre à l'utilisateur de naviguer du textuel au visuel à travers une ontologie visuelle spécifiquement dédiée à l'application considérée.

L'organisation de cet article est la suivante : le paragraphe 2 décrit la recherche par le contenu en distinguant les phases d'indexation logique et de recherche ; la section 3 décrit l'architecture que nous proposons de mettre en place et d'expérimenter sur des corpus d'images annotées. Enfin, nous concluons sur nos travaux futurs en section 4.

2 Recherche par le contenu

Cette section donne un rapide aperçu de l'état de l'art en matière de recherche d'images. La recherche d'images repose sur deux phases à savoir l'indexation logique et la recherche d'images à proprement parler. La phase d'indexation logique consiste à extraire et à modéliser les méta-données (descripteurs textuels et visuels) associées aux images et à les stocker dans une base de données. La phase de recherche permet à un utilisateur final de retrouver rapidement, facilement et efficacement des images « pertinentes ».

2.1 Indexation logique

La recherche dans une base d'images s'effectue en général à partir de descriptions textuelles (mots-clés, annotations, texte, etc.) et/ou de descriptions visuelles (couleur, forme, texture, relations spatiales). Ces descripteurs doivent donc être modélisés de telle sorte que la recherche d'images pertinentes soit efficace tant sur des bases d'images généralistes (pas de domaine d'application particulier) que spécifiques (visages, empreintes). L'extraction des descripteurs est une étape réalisée en amont de la recherche du fait des traitements et des temps d'exécution qu'elle nécessite. Les descripteurs ainsi modélisés et extraits sont stockés dans la base de données en vue d'une exploitation lors du processus d'interrogation.

En ce qui concerne les descripteurs textuels, ils peuvent être des mots-clés représentant des méta-données sur le contenu ou un modèle de base de données. La recherche par mots-clés est un processus très fortement limité du fait que les relations sémantiques entre les mots-clés ne sont pas prises en considération, comme par exemple le fait qu'un cheval est un animal. Si le mot-clé de la requête est cheval alors les images de chevaux annotées avec le mot-clé animal ne seront pas retournées. Un moyen de pallier cette limitation est d'utiliser une ontologie qui permettra de représenter les liens sémantiques entre les différents objets. D'un point de vue bases de données, de nombreux modèles ont été proposés ainsi que des langages de requêtes. Dans la plupart des cas, la description de la sémantique des images est réalisée manuellement. Le modèle DISIMA (Oria et al. 2001) est basé sur une base de données orientée objets et ICDM (Meharga et Monties 2001) sur une base de données relationnelle objets. Dans le premier prototype, une hiérarchie de classes est définie par l'utilisateur. Une image est composée d'un OID, d'un ensemble de représentations physiques (raster ou vecteur) et d'un contenu (relations spatiales, objets pertinents). Le lien entre un objet détecté dans une image et un objet de la classe des objets pertinents est établi manuellement. Dans le second prototype, le modèle peut être divisé en quatre niveaux d'abstraction : (1) le niveau image contenant les propriétés globales d'une image, (2) le niveau syntaxique qui extrait des caractéristiques locales, (3) le niveau contenu dans lequel les objets syntaxiques sont regroupés, et enfin, (4) le niveau sémantique qui définit des hiérarchies sémantiques. Le principal intérêt de ce modèle réside dans le niveau sémantique qui utilise des relations sémantiques comme la synonymie, l'hyponymie, etc. Les modèles semi-structurés sont également bien adaptés à la modélisation de bases d'images. MPEG-7 (Chang et al. 2001) est un standard de description de contenus multimédias. Cette description est écrite en XML et correspond à une approche semi-structurée.

Les descripteurs visuels quant à eux, résument l'information photométrique de l'image. Dans la mesure où le processus de recherche repose sur ces descripteurs, une grande attention est portée sur leur extraction et modélisation. Dans ce contexte des systèmes visuels, une modélisation est « intéressante » si non seulement elle est fiable, mais si elle est aussi compacte et précise. Un tel objectif ne passe que par une véritable synergie entre les domaines du traitement numérique d'images, du traitement du signal et des mathématiques.

D'une manière générale, ils sont représentés par un vecteur numérique; un ou plusieurs descripteurs pouvant être associés à une image. Deux approches se dessinent pour appréhender le contenu des images. La première consiste à modéliser les propriétés visuelles selon des modélisations ayant une correspondance directe avec des critères psycho-visuels humains. Par exemple, le descripteur texture se décline en termes de critères qualitatifs tels que le contraste, la granularité, la régularité, etc. Pour chaque propriété de nombreuses propositions de modélisations ont été faites tant dans le domaine de l'imagerie que dans celui de la recherche par le contenu. Il faut tout de même noter que le descripteur classique et naturel forme est plus sujet à discussion que les autres. En effet, une étape primordiale et préalable à la modélisation de la forme est la segmentation dont dépend essentiellement la qualité de la modélisation de la forme. Il existe de nombreux algorithmes de segmentation automatique qui donnent de bons résultats avec des images peu complexes ou des images complexes d'un domaine bien particulier sur lequel on a de la connaissance à priori. En revanche la segmentation dans le cadre d'images hétérogènes n'est pas toujours fiable dans le sens où les objets extraits ne correspondent pas forcément à des objets sémantiques du monde réel. C'est pourquoi, certains systèmes privilégient les techniques de segmentation semi-automatiques en vue d'obtenir des régions sémantiques, tandis que d'autres préfèrent avoir recours à des techniques entièrement automatisées afin de privilégier le traitement de gros volumes (au détriment des zones sémantiques). Dans ce dernier cas, une méthode

classique consiste à décomposer l'image en petites zones homogènes (blocs) en termes de texture et/ou couleur afin de déterminer « grossièrement » le contour d'un objet. Chaque bloc « cohérent » est alors caractérisé par sa position, sa forme, sa couleur et sa texture. Pour de plus amples informations sur la modélisation de ces propriétés, on pourra se référer pour la couleur à (Swain et Ballard 1991, Smith et Chang 1996), pour la forme et pour la texture à (Gonzalez et Woods 2002). La seconde approche quant à elle, a recours à des modélisations qui n'ont pas de correspondance directe avec des critères psycho-visuels humains. De ces modélisations résultent une signature de l'image importante et des recherches intéressantes. Par exemple, tandis que dans (Nastar et al. 1998) ces caractéristiques sont calculées au moyen des transformées de Fourier, des ondelettes, etc. dans (Schmid et al. 1998), les caractéristiques extraites appréhendent plutôt l'information photométrique locale de l'image. Enfin, il est à noter que l'estimation de la ressemblance entre deux propriétés est dépendante de la modélisation retenue et qu'elle se fait au travers de distances. Par exemple, le calcul de la similarité entre les descripteurs modélisés sous forme de « signature » est effectué au moyen de la distance de Mahalanobis puisque la distance euclidienne et la distance pondérée ne tiennent pas compte des différentes incertitudes et corrélations. De plus, chaque distance a ses avantages et limites : la distance quadratique apprécie particulièrement bien la similarité entre couleurs, en revanche elle nécessite des temps de calculs non négligeables. Pour de plus amples informations sur l'estimation de la similarité entre propriétés, on pourra se référer à (Niblack et al. 1998, Stricker et al. 1995, Venters et Cooper 2000, Oria et al. 2004).

En résumé, on peut dire que dans de nombreux cas, la modélisation correspond à un vecteur de valeurs numériques résumant en fait sous une autre forme l'information photométrique contenue dans l'image. Ce vecteur qui présente généralement une dimension non négligeable, n'est pas sans poser des problèmes d'indexation physique dans les bases de données puisque les caractéristiques extraites s'avèrent être d'excellentes candidates au support de cette indexation. Malgré son importance, l'indexation physique n'est pas abordée dans cet article (Böhm et al. 2001, Oria et al. 2004).. De plus, le choix de la modélisation à retenir est délicat et est étroitement lié au domaine d'application considéré ainsi qu'aux objectifs visés. C'est pourquoi le groupe MPEG a élaboré MPEG-7 (Chang et al. 2001), un standard de représentation du contenu pour le filtrage, la gestion, le traitement et la recherche d'information multimédia. Des liens évidents existent entre cette interface de description du contenu et les systèmes de recherche par le contenu. Cependant, même si MPEG-7 décrit le contenu des images, il ne spécifie pas comment les caractéristiques sont extraites et comment doit être effectuée la recherche sur ces dernières.

2.2 Recherche

Les systèmes de recherche incluent généralement des outils visuels de recherche permettant aux utilisateurs de définir une requête en dessinant, en sélectionnant des couleurs, des formes, etc. ou encore en sélectionnant une ou plusieurs images d'intérêt dans la base. A travers une interface conviviale et intuitive, les utilisateurs peuvent donc formuler leurs requêtes en exploitant à la fois les descriptions textuelles et visuelles (extraites durant la phase d'indexation logique et stockées dans la base). Ces deux types de métadonnées sont nécessaires pour parvenir à une recherche efficace ; en effet, l'utilisation seule du texte ou de l'information visuelle n'est pas suffisante pour décrire le contenu sémantique des images (la puissance d'expression de chaque descripteur est intrinsèquement limitée). Par exemple, les couleurs et les formes sont bien adaptées pour décrire les aspects visuels d'une région mais ne permettent pas d'exprimer des concepts de haut niveau, contrairement aux annotations textuelles qui permettent ces descriptions de haut niveau mais qui sont faibles pour

représenter le contenu visuel. Ces descripteurs, pris séparément, sont incomplets et inefficaces.

Le processus de recherche s'appuie alors sur une fonction de distance entre les descripteurs et calcule la similarité entre la requête utilisateur et la base d'images. Les images sont ensuite affichées par ordre de similarité décroissante. Du fait de l'imprécision des descripteurs extraits des images, le processus de recherche s'appuie sur l'interrogation probabiliste et le contrôle de pertinence. Ceci signifie que l'utilisateur obtient une liste de résultats ordonnés, et, s'il n'est pas satisfait de ce résultat, peut raffiner sa requête en choisissant, parmi les images retournées, des exemples positifs et négatifs. De nouveaux résultats sont alors obtenus et le processus peut être itéré jusqu'à ce que l'utilisateur soit satisfait du résultat. Le lecteur peut se référer à (Del Bimbo 1999) pour un état de l'art de la recherche par le contenu visuel.

D'un point de vue bases de données, les requêtes sont exprimées en utilisant des extensions de SQL ou OQL. De nouveaux prédicats comme *contains* et *similar* sont introduits. Ces requêtes peuvent être qualifiées de requêtes exactes. Des prédicats flous peuvent être introduits (Dubois et al. 2001) pour exprimer des requêtes imprécises.

La combinaison de descripteurs textuels et visuels s'avère cependant insuffisante, notamment lorsque l'interrogation sémantique prédomine, c'est-à-dire que l'image et son contexte sont nécessaires (comme par exemple la recherche de séquences audiovisuelles traitant du chômage). Cette limitation est connue comme le fossé sémantique entre l'apparence visuelle d'une image et l'idée que l'utilisateur se fait de l'information qu'il recherche, incluant bien évidemment une forte composante sémantique. La recherche par le contenu souffre d'un manque de puissance d'expression du fait que la sémantique n'est pas suffisamment prise en compte. C'est la raison pour laquelle de nombreux travaux de recherche sont actuellement menés sur la sémantique des images

Les approches actuelles visent essentiellement à propager des annotations à partir de bases d'images partiellement annotées. Dans notre approche, présentée au paragraphe suivant, nous souhaitons construire une ontologie dédiée à l'application à partir des annotations et l'utiliser dans une phase d'exploration de la base d'images.

3 Une architecture intégrant la fouille de données et les ontologies

En vue de supporter des systèmes de recherche d'images plus puissants, une nouvelle architecture est proposée en figure 1. En ce qui nous concerne, nous souhaitons exploiter la synergie de deux approches à savoir la fouille de données et une ontologie visuelle afin de permettre aux utilisateurs d'explorer et d'exploiter au mieux la base d'images.

La fouille de données est un ensemble de méthodes visant à extraire de la connaissance dans un but exploratoire ou décisionnel. Dans notre approche, nous nous situons dans un contexte exploratoire puisque nous cherchons à déterminer un ensemble de clusters et de règles à partir de descripteurs visuels et textuels (métadonnées associées à notre base).

La sémantique peut être exprimée de manière plus ou moins riche, allant de simples taxonomies aux ontologies (Guarino 1995, Staab et Studer 2004). Une taxonomie est un vocabulaire contrôlé organisé sous forme hiérarchique. Un thésaurus est organisé dans un ordre structuré et connu, de telle manière que les relations d'équivalence, homographiques, hiérarchiques et associatives soient clairement identifiées. Wordnet (Miller 1995) en est un exemple, et organise les noms, verbes, adjectifs et adverbes de la langue anglaise en ensembles de synonymes. Une ontologie est un modèle abstrait représentant une

compréhension commune et partagée d'un domaine. Une ontologie est décrite par un ensemble de concepts, de relations entre ces concepts et de propriétés. Elles peuvent être définies de manière plus ou moins formelle, du langage naturel aux logiques terminologiques. Le langage OWL (Web Ontology Language) appartient à cette dernière catégorie.

Dans l'architecture de la figure 1, on retrouve les deux processus que sont l'extraction et la recherche. Le premier processus a pour objectif de créer un résumé de la base d'images. Après extraction et stockage des caractéristiques visuelles et textuelles des images, le système résume la base d'images au moyen de méthodes issues de la fouille de données. Cette étape intitulée « extraction de connaissances » et détaillée dans la figure 2, s'avère être le noyau de notre architecture ; noyau autour duquel s'articulent les processus d'extraction et de recherche. Il nécessite plusieurs méthodes comme le clustering et la caractérisation des clusters sous forme de règles. Alors que le clustering (Jain et al. 1999) est utilisé pour réduire l'espace de recherche, les règles elles, ont pour finalité de caractériser chaque cluster et de permettre la classification automatique de toute nouvelle image dans les clusters auxquels elle doit appartenir. Du fait de leur nature intrinsèque différente, les descriptions textuelles et les descriptions visuelles sont traitées séparément selon des techniques pour lesquelles les mesures et distances sont appropriées à leur spécificité. Ainsi, à partir de chaque ensemble de caractéristiques (comme l'ensemble des couleurs, l'ensemble des mots-clés, etc.), le système regroupe automatiquement ensemble les images partageant des propriétés similaires en s'appuyant sur le principe des cartes de Kohonen (Kohonen 1995, Dreyfus et al. 2002). Cette étape qui consiste à simplement grouper les objets similaires ensembles est loin d'être simple. La qualité des clusters obtenus dépend souvent du choix des paramètres initiaux ce qui est un problème en soi. De plus, nous souhaitons caractériser chaque cluster au travers d'une représentation plus appropriée que celle du centroïde, à savoir sous forme de règle. Ces règles sont déterminées soit à partir de tous les points du cluster afin d'obtenir les motifs les plus fréquents, soit à partir d'une agrégation des données comme un histogramme médian dans le contexte des clusters couleur (ce qui est représentatif du contenu des clusters). Les règles sont de la forme antécédent → conséquent avec une certaine précision où antécédent et conséquent correspondent respectivement à une valeur de caractéristique et à un cluster. La précision quant à elle est fondamentale puisque son rôle est de permettre l'estimation de la qualité des règles induites. Elle repose en fait sur des mesures statistiques. Des méthodes telles que le marquage symbolique (Diday et al. 2000), la découverte de règles d'association (Agrawal et al. 1994, Han et Kamber 2001) etc. sont envisagées.

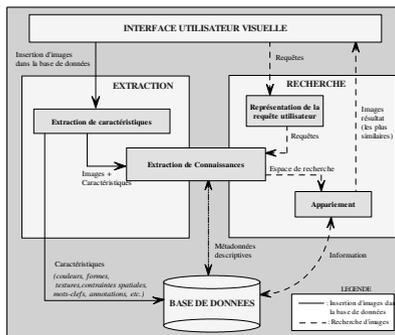


FIG. 1 - Architecture proposée pour les systèmes de recherche d'images

En ce qui concerne le processus de traitement des descriptions textuelles, il se différencie de celui de traitement des descriptions visuelles par le fait qu'il nécessite une phase de pré-traitement afin de réduire le nombre de mots-clefs. Ensuite, le clustering peut être réalisé selon des techniques de clustering conceptuel comme Cobweb ou bien encore en appliquant les techniques de nuées dynamiques après transformation des données textuelles en données numériques. L'ensemble des concepts extraits à partir des clusters est alors hiérarchisé à l'aide de connaissances à priori du domaine, d'un expert ou par des méthodes de classification hiérarchique. Dans le cas où nous disposons uniquement de mots-clés, l'ontologie du domaine est primordiale pour traduire les relations sémantiques existant entre les informations textuelles considérées. Dans le cas où l'aspect textuel est plus important et se traduit par des documents associés aux images (comme des pages web par exemple), il est possible d'extraire automatiquement des relations entre les concepts.

De plus, suite au calcul du résumé d'une base, le système doit être en mesure de classer automatiquement de nouvelles images dans les clusters les plus appropriés au moyen des règles de caractérisation. Cette classification d'images dans les « bons » clusters n'est envisageable que si les règles extraites au préalable sont globalement respectées. Dans la négative, cela soulève un problème crucial nécessitant de plus amples travaux : plusieurs solutions peuvent être envisagées (1) la génération d'un cluster « bruit », (2) la prise en considération de cette nouvelle image et surtout de son impact sur le clustering et leur caractérisation, etc. La suppression d'images dans la base de données n'est pas abordée puisqu'en fait elle soulève les mêmes questions que l'insertion de nouvelles images. Enfin, après avoir effectué le clustering pour réduire l'espace de recherche et caractérisé les clusters au moyen de règles, le système sauvegarde les métadonnées descriptives dans la base de données. Ces métadonnées correspondent en fait aux caractéristiques découvertes et partagées par les images appartenant aux mêmes clusters. Elles ont un rôle important puisqu'elles doivent permettre à l'utilisateur de passer du monde « textuel » au monde « visuel » et inversement, c'est-à-dire lui permettre d'explorer et d'exploiter au mieux la base d'images. La navigation se fera au moyen d'une ontologie visuelle : à partir des concepts extraits des mots-clés, une hiérarchie de concepts va être construite et, pour chaque concept, seront associées des images représentatives issues des clusters visuels.

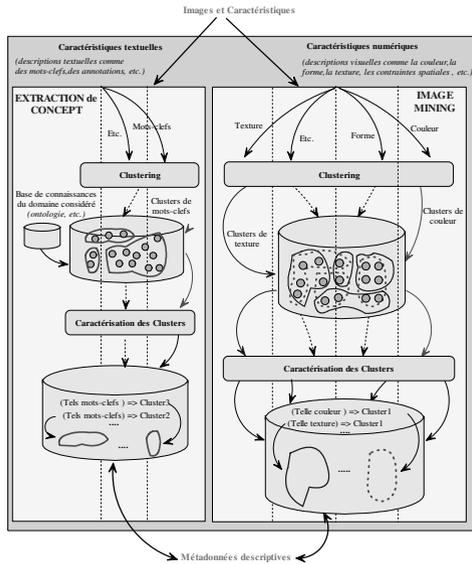


FIG. 2 - Processus « Extraction de Connaissances »

Des premiers résultats de clustering obtenus au moyen des cartes de Kohonen sur une base d'images texturées nous incitent à poursuivre dans cette voie. Ces cartes auto-organisatrices, initialement introduites par Kohonen en 1981, s'attachent à représenter des données multidimensionnelles sur de grands volumes de données. Elles permettent de projeter des données représentées dans un espace de grande dimension dans un espace de faible dimension. De nombreuses applications utilisent les cartes de Kohonen ou des extensions comme par exemple la classification de documents web, la robotique, la recherche d'images par le contenu (Oja et Kaski 2003), etc. Concernant l'aspect textuel, une expérience de découverte d'une ontologie a été réalisée sur un corpus de pages web relatives au domaine du tourisme. Les résultats obtenus sont présentés dans (Karoui et al. 2004). Les clusters obtenus constituent une étape vers la recherche d'images (voire multimédia) plus « intelligente » et cette architecture devrait donner toute son ampleur à la recherche d'information dans les bases d'images voire sur Internet.

4 Conclusion

Ce papier se situe au croisement inévitable du traitement d'image, des bases de données, de la recherche d'informations et de la fouille de données. Il s'inscrit en fait dans la problématique de la recherche d'images au sein d'une base de données. Suite à une présentation des différentes méthodes employées dans les systèmes de recherche actuels et au constat du manque de puissance de ces méthodes pour retrouver efficacement de l'information pertinente comportant des concepts sémantiques, la fouille de données est avancée en vue de pallier leurs limitations en terme d'exploitabilité et d'explorabilité.

L'architecture proposée combine la fouille de données et l'ontologie. Du fait de leur nature intrinsèque différente, les descriptions textuelles et visuelles sont traitées séparément selon des techniques appropriées à leur spécificité. Tandis que nous avons recours au

clustering pour réduire l'espace de recherche, nous utilisons les règles de caractérisation pour décrire chaque cluster et classer une nouvelle image dans les « bons » clusters de la base. Ces techniques favorisent les performances de recherche puisque le système apparie uniquement les caractéristiques sélectionnées avec celles des « bons » clusters de la base. Elles favorisent également l'explorabilité et l'exploitabilité puisque les métadonnées descriptives découvertes et partagées par les images appartenant aux mêmes clusters doivent permettre à l'utilisateur de passer du monde « textuel » au monde « visuel » et inversement, c'est-à-dire naviguer au moyen d'une ontologie visuelle. Des premiers résultats de clustering obtenus au moyen des cartes de Kohonen sur des images texturées nous incitent à poursuivre dans cette voie. Des expérimentations complémentaires sur une base d'images plus complexes sont envisagées. Nous continuons à développer l'architecture proposée et étudions diverses méthodes pour déterminer des métadonnées descriptives et une ontologie visuelle appropriées.

Références

- Agrawal, R. et al (1994), Fast algorithms for mining association rules. International Conference Very Large Data Bases, Santiago, Chili, pp. 487-499.
- Berrani, S., Amsaleg, L. and Gros P. (2002), Recherche par similarité dans les bases de données multidimensionnelles: panorama des techniques d'indexation. RSTI, Ingénierie des systèmes d'information, bases de données et multimédia, 7(5/6), pp 9-44.
- Böhm, C. et al. (2001), Searching in high-dimensional spaces: index structures for improving the performance of multimedia databases. ACM Computing surveys, 33(3).
- Chang, S.F., Sikora, T. and Purl, A. (2001), Overview of the MPEG-7 Standard. IEEE Transactions on Circuits and Systems for Video Technology, special issue on MPEG-7, pp 688-695.
- Del Bimbo, A. (1999), Visual Information Retrieval. Morgan Kaufmann (eds).
- Diday, E., Kodratoff, Y., Brito, P. and Moulet, M. (2000), Induction symbolique numérique à partir de données, (chap. marquage symbolique), Cepadue Ed.
- Djeraba, C. (2002), Association and Content-Based Retrieval. IEEE Transaction on Knowledge and Data Engineering.
- Dreyfus, G., Martinez, J.M., Samuelides, M., Gordon, M.B., Badran, F., Thiria, S. and Héroult, L. (2002), Réseaux de neurones, méthodologie et applications, Ed. Eyrolles.
- Dubois, D., Prade, H., and Sedes, F. (2001), Fuzzy Logic Techniques in Multimedia Databases Querying: A Preliminary Investigation of the Potentials. IEEE TKDE 13(3), pp 383-392.
- Gonzalez, R.C. and Woods, R.E. (2002), Digital image processing. 2nd Ed., Prentice Hall.
- Guarino, N. (1995), Formal Ontology, Conceptual Analysis and Knowledge Representation. International Journal of Human and Computer Studies, 43(5/6), pp 625-640.
- Han, J. and Kamber, M. (2001), Data Mining: Concepts and Techniques, San Francisco, California, Morgan Kaufmann.
- Jain, A.K. and Murty, M.N. and Flynn, P.J. (1999), Data Clustering: A Review. ACM Computing Surveys, 31 (3), pp 264-323.
- Karoui, L., Aufaure, M.A. and Bennacer N. (2004), Ontology Discovery from Web Pages: application to tourism. Workshop on Knowledge Discovery and Ontologies, collocated with ECML/PKDD, Pisa, Italy.
- Kohonen, T. (1995), Self-Organizing Maps. Springer, Berlin.
- Meharga, M.T. and Monties, S. (2001), An Image Content Data Model for Image Database Interrogation. International Workshop on Content-Based Multimedia Indexing, Brescia.

- Miller, G.A. (1995), WordNet: A Lexical Database for English. Communications of the ACM, 38(11), pp 39-41.
- Nastar, C. et al, SurfImage: a Flexible Content-Based Image Retrieval System. The 6th ACM International Multimedia Conference (MM'98), Bristol, England, 1998.
- Niblack, W. et al. (1998), The QBIC project: Querying images by content using color, texture and shape. In Proc. SPIE Storage and Retrieval for Image and Video Databases.
- Oja, E. and Kaski, S. (2003), Kohonen Maps, Elsevier, 2nd Ed.
- Oria, V., Özsu, T. and Iglinski, P.J. (2001), Querying Images in the DISIMA DBMS. Proc. of the 7th Int. Workshop on Multimedia Information Systems, Capri, Italy, pp 89-98.
- Oria, V., Li, Y. and Dorai, C. (2004), Multimedia Databases: Analysis, Modeling, Querying and Indexing. In Computer Science and Engineering Handbook, 2nd Ed., CRC Press.
- Schmid, C. et al. Comparing and evaluating interest points. In proceedings of the 6th international conference on Computer vision, Bombay, India, 1998.
- Simoff, S.J., Djeraba, C. and Zaïane, O.R. (2002), Multimedia Data Mining between Promises and Problems. MDM/ KDD2002, ACM SIGKDD Explorations, 4(2).
- Smith, J.R. and Chang, S.F. (1996), Tools and Techniques for Color Image Retrieval. Storage & Retrieval for Image and Video databases IV, SPIE Proceedings, 2670.
- Staab, M. and Studer, R. (eds) (2004), Handbook on Ontologies. Springer.
- Stricker, M. et al. (1995), Similarity of color images. Storage and Retrieval for Image and Video databases III, SPIE Proceedings, 2420.
- Swain, M.J. and Ballard, D.H. (1991), Color indexing. Int. journal of computer vision, 7(1).
- Venters, C.C. and Cooper, M.D. (2000), A Review of Content-Based Image Retrieval Systems. JISC Technology Applications Program.
- Zhang, J., Hsu, W. and Lee, M.L. (2001), Image mining: issues, frameworks and techniques. Second International Workshop on Multimedia Data Mining, San Francisco, USA.

Summary

Nowadays, image media is omnipresent for various applications. A considerable volume of data has been produced and we need now to develop powerful tools allowing to efficiently retrieve relevant information. At present, CBIR systems suffer from a lack of expressive power because they do not integrate enough semantics. An interesting way we want to explore to fill the semantic gap between visual features and semantics content is image mining. This research field is recent and still remains immature but seems to be a promising issue. This paper presents some preliminary work and ideas about the way to define new descriptors to integrate image semantics. Clustering and obtained class characterization are used to reduce the research space and to produce a summarized view of the image database. As far as the navigation is concerned, it is based on a visual ontology, a powerful and user-friendly tool to retrieve relevant information.