

Systemes de representation d'images haut-niveau pour l'indexation

J. Dombre

N. Richard

C. Fernandez-Maloigne

Laboratoire IRCOM-SIC, UMR n°6615 CNRS
Université de Poitiers, UFR SFA
Bât. SP2MI, Boulevard 3, Téléport 2 BP 179
86960 FUTUROSCOPE CEDEX
{dombre,richard,fernandez}@sic.sp2mi.univ-poitiers.fr

Résumé

En indexation d'images, la plupart des descripteurs développés sont basés sur des statistiques globales aux images. Aujourd'hui, des méthodes essayant d'aller plus loin dans la représentation apparaissent. Sans aller jusqu'à une description sémantique des images, elles représentent les images plus finement en utilisant des statistiques localisées ainsi que les interactions entre elles. Pour la plupart, des structures de graphes sont utilisées. Cet article présente ces diverses approches ainsi que celle que nous avons développée en construisant un graphe pyramidal. Les premiers résultats obtenus avec cette méthode sont aussi présentés.

Mots Clef

Indexation, représentations de haut-niveaux, graphe.

1 Introduction

Après de nombreux développements sur des bases très variées, l'indexation doit relever le défi de bases spécialisées à large spectre de variations. Dans les approches classiques des descripteurs plus ou moins globaux à l'image apportent des réponses de qualité variable, qui montrent leurs insuffisances pour ces bases. Quelques méthodes ont cherché à prendre en compte une forme d'information spatiale plus générique, mais peu se sont hasardées sur une description de haut-niveau.

Nous introduisons le terme de *traitement orienté objets* pour définir que notre description cherche à se rapprocher d'une description sémantique du contenu de l'image. Celle-ci n'est plus considérée comme un assemblage de zones homogènes connexes mais comme une composition d'objets complexes. Le problème de l'indexation se résume alors dans la recherche d'images de même composition ou comportant une partie des objets inclus (cas des requêtes partielles).

Cet article présente un état de l'art des diverses struc-

tures de données existantes utilisées pour décrire les images dans un but d'indexation, leurs intérêts et leurs limites. Nous proposons aussi notre structure qui apporte une description multi-résolution orientée régions et/ou frontières des images. Quelques résultats seront présentés à la suite pour illustrer notre propos.

2 Structure de données pour l'indexation

2.1 Statistiques globales ou structure à une région

La majeure partie des travaux en indexation utilisent diverses statistiques globales à l'image[10, 12]. Swain et Ballard[15] ont été les premiers à utiliser les histogrammes qu'ils intersectèrent pour calculer une distance entre images. Depuis, différentes approches statistiques ont été utilisées pour décrire plus ou moins précisément l'ensemble des régions composant l'image (couleur, texture, matrice de cooccurrence, coefficients d'ondelettes ou de Gabor, ...)[4, 14, 3, 9].

S'il existe quelques variantes de ces méthodes adaptées au problème des requêtes partielles[3, 6], elles restent très attachées à la conservation de statistiques globales. De plus, cette hypothèse n'est pas toujours vérifiée lors de changements d'échelle importants ou d'altérations (occlusions, changement d'illuminant, ...)[2].

Le point commun de ces méthodes est finalement de considérer l'image comme une structure unique et non comme une ensemble d'objets complexes. Les statistiques calculées décrivent alors l'image dans son ensemble.

2.2 Structures spatiales

Pour tenir compte de l'arrangement spatial des différentes statistiques de l'image et des relations entre les diverses régions de l'image, Matas[7] et Park[8] ont proposé des descripteurs spatiaux à partir

de graphes de régions (RAG, CAG, MCAG).

Pour contourner l'énorme complexité des calculs lors de la comparaison des RAGs (*Region Adjacency Graphs*), Matas[7] propose d'utiliser un graphe des couleurs, le CAG (*Colour Adjacency Graph*). Chaque nœud représente alors une couleur (représentant un amas de l'histogramme) et non plus une région. Les arêtes symbolisent alors les adjacences des couleurs au sein de l'image et contiennent le coefficient de réflexion mutuelle. Les différences entre représentation sont obtenues à partir d'un formalisme de type *graph-matching* d'où une complexité importante.

Park[8] propose alors le MCAG (Modified Colour Adjacency Graph) pour s'affranchir de ce problème. Après filtrage et quantification de l'image, chaque nœud contient la fréquence d'apparition d'une couleur et les arêtes codent la longueur de la frontière entre couleurs. Ces simples informations peuvent alors être stockées dans une matrice de faible taille (égale au nombre de couleurs utilisées). La comparaison de deux représentations est alors grandement facilitée. Pour prendre en compte la géométrie de la répartition des diverses couleurs, Park met en place un autre descripteur, le SVG (Spatial Variance Graph). Les nœuds de ce nouveau graphe contiennent la variance spatiale de chaque classe colorimétrique et les arêtes les variances relationnelles entre classes. La combinaison de ces deux graphes permet donc de prendre en compte la distribution des différentes couleurs au sein de l'image, leurs relations ainsi que leur répartition géométrique propre.

Si l'utilisation des graphes résout un grand nombre de problèmes, elle représente aussi un coût de traitement assez important. De plus, une structure de graphes à *plat* n'apporte que peu d'intérêt vis-a-vis des requêtes partielles.

Parallèlement à ces approches formelles, Smith[13] et Santini[11] proposent des critères spatiaux de décision.

Le modèle FFC[11] (Fuzzy Feature Contrast) permet de définir une mesure de similarité entre deux arrangements spatiaux. Par exemple, pour comparer deux visages, divers critères tels que la position de la bouche et des yeux sont utilisés.

En revanche, le système SaFe[13] développé par Smith combine une approche globale (histogramme couleur) avec des descripteurs spatiaux. La définition de régions colorées et de leur arrangement spatial permet alors d'affiner la recherche. Pour cela chaque zone requête fournit une liste de régions similaires. La discrimination se fait alors grâce à leur disposition au sein des images. On peut ainsi facilement obtenir

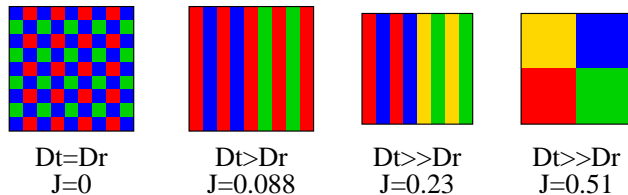


FIG. 1 – Exemple de calcul du coefficient J.

les images possédant du bleu en haut (du ciel par exemple) et du vert en dessous (de l'herbe).

Ces techniques, plus intuitives dans leur formalisme, sont très performantes mais peu évolutives. Bien qu'intéressantes dans des cadres applicatifs précis, elles ne règlent pas non plus le problème des requêtes partielles.

3 Structure multi-résolution par graphe pyramidal

Une structure efficace doit pouvoir osciller entre des descripteurs statistiques garant du contenu des régions et des descripteurs topologiques garant de l'organisation des objets au sein de l'image. Sans aller jusqu'à une description topologique totale de l'image segmentée comme les G-cartes[5], notre proposition s'appuie sur une structure de description haut-niveau utilisant une approche multi-résolution orientée région et/ou contours.

3.1 Segmentation de l'image

La structure multirésolution est portée par une segmentation région multiéchelle proposée par Deng[1]. La prise en compte de l'information colorimétrique est effectuée lors d'une quantification préalable de l'image, chaque couleur étant ensuite considérée comme une classe. La segmentation est ensuite obtenue par l'intermédiaire d'un coefficient, J, de rupture de modèle textural.

L'hypothèse initiale du calcul de J est de considérer la texture couleur uniforme sur la région étudiée, les centres de gravité des différentes couleurs sont alors confondus et leur dispersion est identique. L'indice J est alors calculé à partir des localisations des centres des classes et de leur répartition spatiale. La figure 1 montre 4 exemples simples de calcul de J dans un voisinage.

L'aspect multirésolution est pris en compte au travers de la taille des fenêtres, l'utilisation successive de diverses tailles de masque permettant de détecter des contours plus ou moins important. La génération des régions s'effectue simplement par une technique de *watershed* de ce coefficient. L'aspect multi-échelle évoqué



FIG. 2 – Exemple d’images segmentées issues de la base Columbia (sans critère de regroupement).

précédemment permet d’affiner de façon significative cette étape. Cette approche permet d’obtenir des segmentations cohérentes entre elles et donc un graphe pyramidal rationnel (cf. figure 2).

3.2 Graphe pyramidal

Cette nouvelle représentation, basée sur des graphes, mélange les approches régions et frontières. Les régions amènent à une description intrinsèque des divers objets (ou parties d’objets) de la scène et les frontières à une description des diverses interactions entre objets.

Cet aspect multi-résolution permet d’appréhender les différents niveaux de perception d’une image : du premier regard à la vision en détail, il est porté par le graphe pyramidal qui représente à chaque niveau une description plus ou moins fine de l’image. Cette structure multi-résolution permet aussi d’obtenir des critères de rejet très efficaces dans le cadre de l’indexation.

La figure 3 présente le schéma d’un graphe pyramidal issu d’une segmentation manuelle de l’image. On part de l’image entière considérée comme une seule région ; c’est le sommet. La segmentation la plus

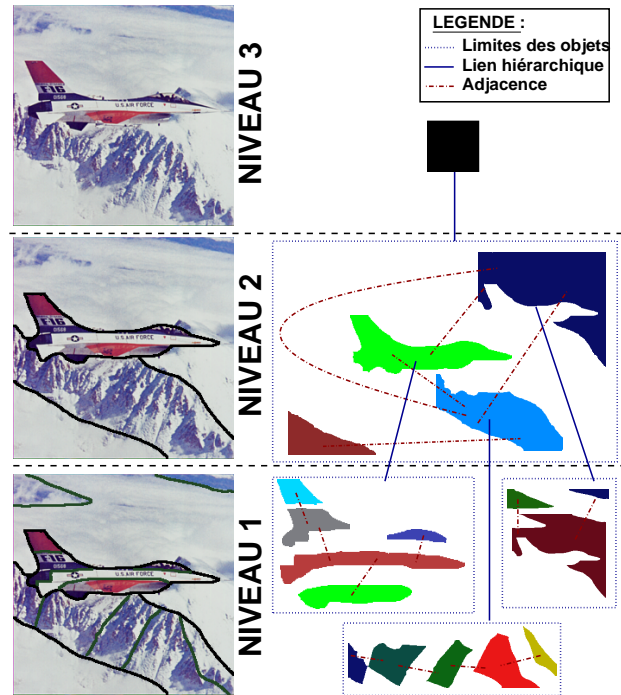


FIG. 3 – Exemple de graphe pyramidal sur une segmentation manuelle.

grossière nous fournit alors le niveau suivant. L’image se décompose alors en plusieurs objets (régions). Ils sont alors représentés par un nœud du graphe et leurs adjacences par les arêtes. Ils possèdent tous la même région mère : l’image. Ainsi de suite, les divers segmentations fournissent alors un nouveau niveau. Cette structure peut se comparer à un arbre d’inclusion au sein duquel le graphe plat de chaque niveau aurait été incorporé.

Les diverses statistiques des régions sont alors codées dans les nœuds du graphe. Quant aux caractéristiques des frontières elles se retrouvent dans les arêtes. Le côté pyramidal de la représentation, nous permet d’effectuer les calculs de ses statistiques de façon récursive et donc de manière très rapide.

4 Premiers résultats

Dans le cadre de l’indexation, cette structure est validée à partir de critères de comparaison simples. Les approches de type *graph-matching* ne sont pas utilisées. Les résultats obtenus pour les requêtes globales sont bons et très prometteurs dans le cas des requêtes partielles.

Nous avons ainsi pu montrer l’intérêt de notre structure. Pour cela, nous avons travaillé sur la base d’images *Columbia* pour valider notre approche. Nous avons alors utilisé un simple algorithme d’intersection d’histogrammes couleurs entre régions au sein des

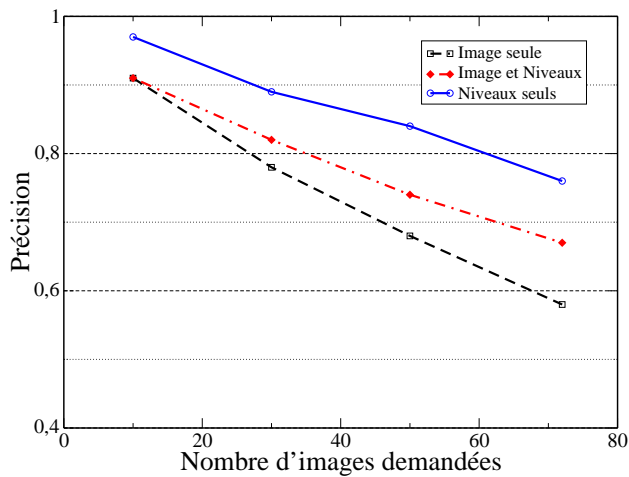


FIG. 4 – Mise en évidence de l'amélioration apportée.

divers niveaux des pyramides comparées. Les résultats obtenus (cf. figure 4) montrent une augmentation moyenne de 15% de la précision lors de l'utilisation des niveaux inférieurs (i.e. sans le niveau supérieur qui correspond à l'image entière) par rapport à l'utilisation de la même statistique sur l'image entière.

Ceci montre bien que l'ensemble des statistiques localisées porte plus d'information que la même statistique au niveau globale.

5 Applications et conclusion

Nous présentons ici un schéma d'indexation de type *coarse-to-fine*. L'architecture proposée s'organise autour de la notion de **traitement orienté objets**. Dans une telle vue, nous cherchons à décrire la projection d'un assemblage d'objets (l'image) par une structure d'information la plus complète possible. Par l'utilisation de l'aspect multi-résolution nous dotons notre système de moyens efficaces pour l'indexation, tout en répondant également au délicat problème de la requête partielle.

Ce travail est effectué dans le cadre d'un partenariat avec le Centre d'Etude Supérieur des Civilisations Médiévales (CESCM). L'enjeu de cette application est de proposer un outil capable d'aider l'historien dans sa recherche de représentation d'images d'archives médiévales. Les diversités des alterations (détériorations, moisissures, effacements partiels ...) imposent ici de disposer d'un schéma de représentation extrêmement robuste et précis dans la description.

Références

[1] Y. Deng, B. Manjunath, and H. Shin. Color image segmentation. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 1999.

[2] B. Funt and G. Finlayson. Color constant color indexing. In *IEEE Transactions on PAMI*, volume 17, pages 522–529, 1995.

[3] J. Huang, S. R. Kumar, M. Mitra, W. Zhu, and R. Zabih. Image indexing using color correlograms. In *IEEE Computer Vision and pattern Recognition conference*, June 1997.

[4] F. Idris and S. Panchanathan. Image indexing using wavelet vector quantization. In *SPIE Vol. 2606–Digital Image Storage and Archiving Systems, Philadelphia, PA, USA*, pages 269–275, Oct. 1995.

[5] P. Lienhardt. Topological models for boundary representation : A comparison with n-dimensional generalized maps. In *Computer Aided Design*, 23(1), pages 59–82, Jan. 1991.

[6] H. Lu, B. Ooi, and K. Tan. Efficient image retrieval by color contents. In *Proceedings of the 1994 International Conference on Applications Database*, 1994.

[7] J. Matas, R. Marik, and J. Kittler. The color adjacency graph representation of multicolored objects, 1995.

[8] K. Park, I. Yun, and S. Lee. Color image retrieval using a hybrid graph representation. *Journal of Image and Vision Computing*, (17) :465–474, 1999.

[9] G. Pass, R. Zabih, and J. Miller. Comparing images using colour coherence vectors. In *ACM conference on Multimedia*, 1996.

[10] Y. Rui, T. Huang, and S. Chang. Image retrieval : current techniques, promising directions, and open issues. *Visual Communication and Image Representation*.

[11] S. Santini and R. Jain. Similarity measures. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 21, 1999.

[12] R. Schettini, G. Ciocca, and S. Zuff. Color in databases : Indexation and similarity. In *CGIP 2000*, pages 244–249, Oct. 2000.

[13] J. Smith and S. Chang. Integrated spatial and feature image query. In *Multimedia Systems 7 (2)*, pages 129–140, 1999.

[14] M. Stricker and M. Orengo. Similarity of color images. In *Storage and Retrieval for Image and Video Databases SPIE*, volume 2420, pages 381–392, Feb. 1995.

[15] M. J. Swain and D. H. Ballard. Color indexing. *IJCV*, 7(1) :11–32, 1991.