

Accès aux bases d'images par navigation visuelle

G. Lavoué K. Idrissi A. Baskurt

LIRIS, Université Claude Bernard, Lyon 1
Bât. Nautilus, 8 boulevard Niels Bohr, 69622 Villeurbanne cedex – France

{glavoue, kidrissi, abaskurt}@ligim.univ-lyon1.fr

Résumé

Nous abordons dans cet article le thème de l'accès aux bases d'images à travers une approche originale de navigation interactive, de telle sorte que l'utilisateur puisse se déplacer et zoomer dans la base suivant le type d'images qu'il cherche. Le système proposé est basé sur la couleur et intègre la recherche traditionnelle à partir d'images clefs. Pour la navigation, la base d'images est structurée sous la forme d'un arbre hiérarchique en utilisant une méthode de classification floue, l'affichage des images relatives à un nœud se faisant par projection sur le plan le plus discriminant. La recherche, quant à elle, est basée sur la mesure de similarité couleur entre l'image clef et les images de la base.

Mots clefs

Couleur, Navigation, Recherche d'Images par le Contenu.

1 Introduction

Les systèmes classiques de recherche d'images par le contenu (RIPC)[1] retournent les images trouvées sous forme de liste linéaire, ordonnée par ordre de similarité. L'inconvénient de cette recherche directe est qu'elle ne donne ni une vue globale des images retournées, ni une idée des relations qui existent entre elles. Le système que nous proposons permet de coupler la recherche traditionnelle avec un système de navigation visuelle. Ce type de navigation prend tout son sens dans le cas de la recherche d'images par le contenu. En effet, elle permet de donner une vue globale de la base et de sa structure, sachant que l'utilisateur recherche plus souvent un type d'image plutôt qu'une image précise. Les principales difficultés sont de permettre une visualisation de toute la collection et de fournir un mécanisme de navigation efficace.

Pecenovic [2] propose d'effectuer une projection de Sammon sur la base, afin de projeter celle-ci sur le plan le plus discriminant. La base est ensuite classée de manière hiérarchique, afin de permettre à l'utilisateur de zoomer progressivement vers la partie qui l'intéresse. Dans [3], Chen construit à partir de la base une structure hiérarchique en agglomérant les images les plus proches. L'approche proposée dans cet article consiste à classer

hiérarchiquement la base en procédant par la méthode des Nuées Dynamiques (k-means), puis à effectuer une Analyse en Composantes Principales (ACP) sur chaque classe afin que l'affichage fournisse des informations sur la dissemblance entre images d'une même classe. Cette structuration ne nécessite pas d'apprentissage et peut donc s'appliquer sur n'importe quelle base.

2 Indexation

La mise en place du système nécessite une étape préalable d'indexation des images de la base, afin d'extraire des descripteurs qui serviront à classer la base dans le cadre de la navigation et à évaluer la similarité entre images dans le cas de la recherche directe. Dans cette étude, nous avons privilégié la caractéristique couleur pour décrire les images. Les descripteurs choisis sont de 2 types :

2.1 Couleurs dominantes

Ce descripteur est issu de la méthode de quantification couleur présentée en [4]. Chaque classe couleur i est caractérisée par 3 paramètres : La couleur dominante CD_i , le pourcentage p_i que représente la classe par rapport à l'image et la variance σ_i de la couleur dans la classe. Ce descripteur est alors constitué de N triplets (CD_i, p_i, σ_i) , N étant le nombre de couleurs dominantes détectées dans l'image.

2.2 Cohérence spatiale

Ce descripteur explicité en [4] a pour but de combler, en partie, l'absence d'information spatiochromimétrique de l'image dans le descripteur précédent. Une faible valeur de SCR indiquera que la couleur est dispersée dans l'image, tandis que pour une couleur dominante homogène, SCR sera proche de 1.

3 Mise en place de la structure de navigation

Notre point de départ étant les descripteurs couleur des images de la base, nous proposons d'utiliser l'histogramme pour évaluer la similarité couleur entre images. En supposant que l'histogramme couleur peut être

modélisé par un mélange de distributions gaussiennes, un nouvel histogramme est créé à partir des N couleurs dominantes fournies par le descripteur de l'image, et intégrant les 3 paramètres constituant ce descripteur (couleur dominante, pourcentage et variance), en combinant N distributions gaussiennes. La contribution de la i ème couleur dominante CD_i est donnée par :

$$H_i(x) = \frac{p_i}{\sigma_i \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x-CD_i)^2}{2\sigma_i^2}}$$

où p_i représente son pourcentage et σ_i sa variance.

Ce modèle de distribution nous servira aussi bien pour la navigation que pour la recherche.

Le mécanisme de navigation nécessite 3 étapes de traitements. Celles-ci ne sont pas exécutées en ligne, mais une seule fois lors de l'indexation de la base.

3.1 Création d'un vecteur représentatif pour chaque image

A partir des descripteurs (couleur, cohérence spatiale) issus de l'indexation, on crée pour chaque image, un vecteur descriptif V , afin de pouvoir représenter la base comme un nuage de points en dimension réduite $n = n_c + 1$.

$$V_i = \omega_c \times C_j \quad \text{pour } 0 \leq i < n_c \text{ et } 0 \leq j < n_c$$

$$V_{n-1} = \omega_{cs} \times SCR$$

Les C_j représentent les coefficients couleur, issus de la discrétisation des histogrammes couleur L, a et b, en $n_c/3$ intervalles. ω_c et ω_{cs} sont les poids respectifs que l'on veut donner à la couleur et à la cohérence spatiale.

3.2 Structuration de la base

La base d'images est organisée en une structure d'arbre hiérarchique afin de permettre à l'utilisateur de zoomer successivement sur des régions particulières de la base. Pour créer cette structure, les images, considérées comme des points dans l'espace à n dimensions, sont regroupées récursivement en régions contenant des images similaires. Pour chaque région, l'image la plus proche du barycentre est choisie comme image représentative. Cette structure permet une navigation efficace et rapide.

A chaque niveau, l'étape de regroupement est réalisée par un algorithme simple et rapide de classification non supervisée, type nuées dynamiques, suivi d'une régularisation des groupes (élimination des groupes trop petits, fusion des groupes trop proches).

Les images étant regroupées sur des critères de ressemblance liés à leurs descripteurs et non sur des critères sémantiques, il apparaît donc évident que certaines images ne vont pas se retrouver dans le groupe que l'utilisateur aurait choisi par rapport à ses critères.

Afin de minimiser ce problème, la classification par nuées dynamiques est suivie d'un algorithme de classification floue (Fuzzy k-means)[5]. A chaque image P_i est associé, non pas un groupe, mais un pourcentage d'appartenance μ_{ik} à chaque groupe G_k :

- En premier lieu, les μ_{ik} sont initialisés à partir du résultat des nuées dynamiques :

$$\mu_{ik} = 1 \quad \text{si } P_i \in G_k$$

$$\mu_{ik} = 0 \quad \text{si } P_i \notin G_k$$

Puis les centres de classe floue G_k^0 sont calculés :

$$G_k^0 = \frac{\sum_{i=1}^N (\mu_{ik})^m P_i}{\sum_{i=1}^N (\mu_{ik})^m} \quad k = 1 \dots K \quad (1)$$

avec K le nombre de groupes. m est un facteur de poids tel que $m \in]1 \dots +\infty[$. Plus m est grand, plus la quantification est floue, et plus m est proche de 1, plus la quantification est stricte.

- Ensuite, les pourcentages d'appartenance μ_{ik} sont mis à jour :

$$\mu_{ik} = \frac{1}{\sum_{j=1}^K \left(\frac{d_{ik}}{d_{ij}}\right)^{\frac{2}{m-1}}} \quad k = 1 \dots K \text{ et } i = 1 \dots NI \quad (2)$$

avec NI le nombre d'images de la base et d_{ik} la distance euclidienne entre la i ème image et G_k^0 .

- Enfin, on compare les matrices d'appartenance U avant et après l'itération :

$$\|U^p - U^{p-1}\| < \varepsilon \quad (3)$$

où p est le numéro de l'itération courante, ε le seuil d'arrêt et U telle que : $U[i][j] = \mu_{ij}$.

Les deux étapes (1) et (2) sont répétées jusqu'à ce que la condition (3) soit vérifiée. Une image peut alors, suivant son pourcentage finale d'appartenance, être affectée à plusieurs groupes, réduisant ainsi les problèmes d'images sémantiquement mal placées.

3.3 Mécanisme de visualisation

Le but de cette étape est d'apporter un maximum d'information à l'utilisateur lors de l'affichage. Les images étant vues comme un nuage de points en dimension n , cela revient à appliquer une projection multidimensionnelle de données d'un espace à n dimensions dans l'espace à 2 dimensions de l'écran. Il

existe plusieurs méthodes pour effectuer ce type d'opération, notamment l'analyse en composante principale (ACP)[6] utilisé par Moghaddam[7], la projection de Sammon, le « multidimensional scaling »(MDS)[8] utilisé par Rubner[9]. L'ACP est l'approche que nous avons choisie, c'est une méthode linéaire très rapide.

Pour chaque groupe de l'arbre hiérarchique, l'ACP est effectuée sur les images représentatives de ses sous groupes afin de les projeter de l'espace de dimension n , dans le sous espace de dimension 2 tel que l'inertie du nuage de points-images projetés soit maximale. Un tel affichage apporte beaucoup d'information à l'utilisateur, dans la figure 2, on peut nettement voir deux groupes se détacher, les images de teinte jaune (en haut), et celle de teinte verte (en bas), de gauche à droite, on a aussi une distinction entre images foncées et images plus claires. L'expérience montre que la quantité d'information (le pourcentage d'inertie) portée par les deux axes principaux est en moyenne de 80%. Cette forte valeur est due au fait que peu d'images sont affichées en même temps (20 au maximum).

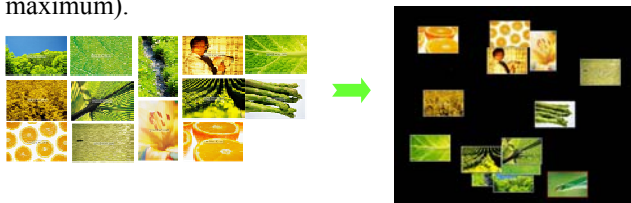


Figure 2 – Affichage selon les deux axes principaux, calculés par ACP sur les vecteurs représentatifs, d'un groupe d'images quelconques.

4 Fonctionnement de la navigation

Grâce aux pré-traitements hors ligne de structuration de la base, le système présente celle-ci sous forme d'un espace navigable en temps réel. En se plaçant à un nœud de l'arbre hiérarchique (A dans la Fig. 3), l'utilisateur peut voir les images représentatives des nœuds fils (B1, B2, B3 et B4), ou bien directement les images du nœud si celui-ci est terminal.

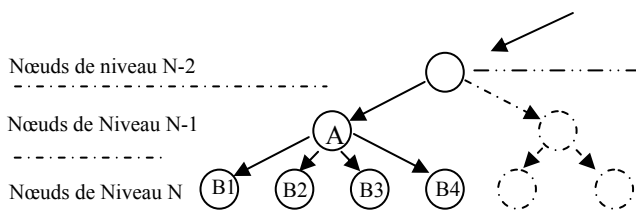


Figure 3 - Structure hiérarchique de la base. Chaque nœud situé au niveau N correspond à une classe d'images, elle-même constituée de sous-classes (et de nœuds fils) de niveau $N+1$.

La figure 4 illustre le principe de navigation. En cliquant sur une image représentative située au niveau (N),

l'utilisateur pénètre alors dans le sous groupe correspondant situé au niveau (N+1). L'utilisateur a aussi la possibilité de sélectionner une image et de se placer automatiquement dans tous les groupes terminaux où cette image est présente, ce qui permet une navigation transversale.

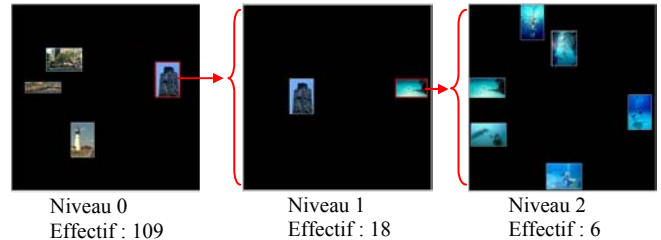


Figure 4 - Illustration de la navigation sur une base de 109 images issues de la base GoodShoot©.

5 Le système IMALBUM

Le système IMALBUM intègre la recherche directe et la navigation. La recherche par image-clef est une approche complémentaire à la navigation. Elle intervient lorsque l'utilisateur dispose d'une image similaire à ce qu'il cherche, surtout si celle-ci ne fait pas partie de la base. Lorsque l'utilisateur lance une recherche avec une image-clef, les descripteurs sont extraits de celle-ci. Une distance est alors calculée entre les descripteurs de l'image requête et ceux des images de la base. Les images les plus similaires sont ensuite retournées à l'utilisateur.

La distance couleur utilisée est issue de la distance de Kullback [10] dans sa forme symétrique donnée par :

$$d_c(R, I) = \sum_{n=1}^H \sum_{m=1}^3 (r_{n,m} - i_{n,m}) \log_2 \left(\frac{r_{n,m} + 1}{i_{n,m} + 1} \right)$$

où H est le nombre d'éléments de l'histogramme de chaque composante couleur (classiquement $H=256$), et $r_{m,n}$ et $i_{m,n}$ sont les pourcentages de la $m^{\text{ième}}$ composante de la $n^{\text{ième}}$ couleur respectivement de l'image-clef et d'une image de la base. La distance Euclidienne est utilisée pour la cohérence spatiale. La distance totale est alors donnée par la somme pondérée des 2 distances précédentes.

La figure 5 donne une vue de l'interface utilisateur du système IMALBUM. Celle-ci comporte 6 fenêtres notées de (a) à (f) de gauche à droite, et de haut en bas. La fenêtre (a) est la fenêtre de navigation sur laquelle figurent les images représentatives des sous-groupes d'un nœud de l'arbre de la base, les axes du plan d'affichage étant les axes principaux des images affichées.

Dans la fenêtre (b), le système fournit les images résultats de la requête utilisateur, affichées dans un plan dont les axes sont la distance couleur en abscisse et la distance de cohérence spatiale en ordonnée. Il est possible de considérer d'autres paramètres pour les axes de projection, comme la texture ou la forme, si ceux-ci sont utilisés comme descripteurs, faisant ainsi ressortir les paramètres de dissemblance entre des images similaires. Dans le cas de l'exemple de la figure 5, les images retournées par la recherche (fenêtre en haut à droite) forment 3 groupes, celles qui sont similaires à l'image-clé aussi bien au niveau de la couleur que de la cohérence spatiale, celles s'en éloignant selon les 2 critères et enfin celles qui ont la même cohérence spatiale mais pas la même couleur. L'intérêt de cet affichage paramétrique est d'apporter une aide à l'utilisateur en lui montrant comment se répartissent les images résultats de sa requête en fonction des critères qu'il considère comme importants.

La fenêtre (c) représente toutes les images appartenant aux nœuds représentés dans la fenêtre (a), sous forme de points, donnant ainsi à l'utilisateur une idée de la taille des sous-classes correspondant à chaque nœud. L'affichage se fait dans le plan des axes principaux calculés pour la fenêtre (a).

Les fenêtres (d) et (e) contiennent respectivement l'image sélectionnée et l'image-clé servant de modèle pour la recherche, tandis que la fenêtre (f) donne le résultat de la recherche sous forme linéaire.

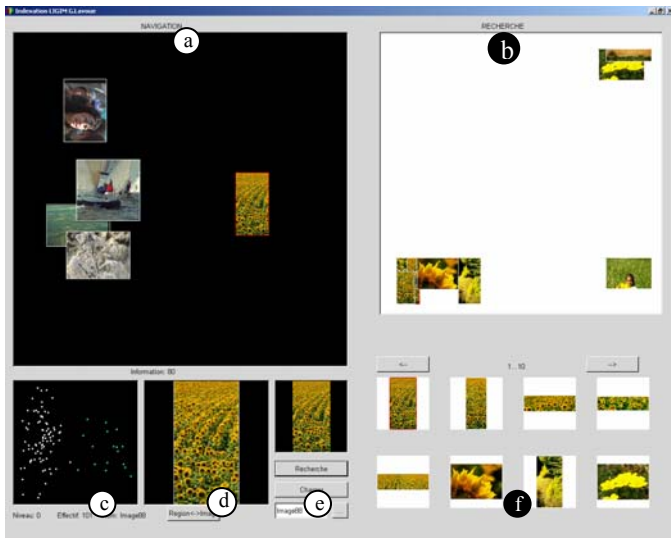


Figure 5 - Interface utilisateur d'IMALBUM.

La navigation et la recherche directe sont des outils complémentaires. Dans notre système ils sont étroitement mêlés. A n'importe quel niveau de la navigation, l'utilisateur peut sélectionner une image dans la fenêtre de navigation et l'utiliser comme image-clé pour lancer une recherche. Réciproquement, à la suite d'une requête de recherche, l'utilisateur peut sélectionner une des images retournées et visualiser dans la fenêtre de navigation les groupes terminaux contenant l'image sélectionnée.

6 Conclusion

Nous avons présenté IMALBUM, un système d'accès aux bases d'images intégrant la navigation à travers la base et la recherche par image-clé. Ces deux modes d'accès sont complémentaires et offrent à l'utilisateur plus d'efficacité et de flexibilité. La structuration de la base d'image en arbre hiérarchique par une approche floue permet d'avoir une structure souple, autorisant plusieurs chemins pour accéder à une même image ou classe d'images. Les possibilités d'extension de ce système sont grandes, outre l'intégration d'autres paramètres que la couleur (texture, forme...), on peut réfléchir sur une manière de pré-classifier sémantiquement la base avant la navigation, afin de présenter plusieurs sous-bases à l'utilisateur. Vailaya [11], par exemple, propose des descripteurs spécifiques permettant de distinguer « Villes » et « Paysages naturels ».

Références

- [1] Y. Rui, T. S. Huang, S. F. Chang. Image retrieval : Past, Present and Future. Int. Symposium on Multimedia Information Processing. Taipei, Taiwan, pp. 27-43, Dec 1997.
- [2] Z. Pecenovic, M.N. Do, M. Vetterly and P. Pu. Integrated Browsing and Searching of Large Image Collections, Proc of 4th Conf. On Visual Info. Systems, 2000.
- [3] J. Chen, C.A. Bouman and J.C. Dalton. Similarity Pyramids for browsing and organization of Large Image Databases, Proc. of SPIE/IS&T Conf. On Human Vision and Electronic Imaging III, volume 3299, pp. 563-575, 1998.
- [4] K. Idrissi, J. Ricard and A. Baskurt. An objective performance evaluation tool for color based image retrieval systems. ICIP-2002, Rochester, USA, September 2002.
- [5] J. C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms. Plenum Press, New York, 1981.
- [6] Jolliffe, I.T., Principal Component Analysis, Springer-Verlag, New-York, 1986.
- [7] B. Moghaddam, Q. Tian, Thomas S. Huang, Spatial Visualization for Content-Based Image Retrieval, ICME2001, August 22-25, Waseda University, Tokyo, Japan, 2001.
- [8] W. S. Torgeson, Theory and methods of scaling, John Wiley and Sons, New York, NY, 1958.
- [9] Y. Rubner, *Perceptual metrics for image database navigation*, Ph.D. dissertation, Stanford University, 1999.
- [10] S. Kullback. *Information theory and statistics*. New York: John Winley & Sons, 1959.
- [11] A. Vailaya, A. Jain, H. Z. Zhang. On Image Classification : City Images vs. Landscapes. Pattern Recognition, vol. 31, pp 1921-1936, December, 1998.