
Recherche d'images par régions d'intérêt : segmentation grossière rapide et description couleur fine

Julien Fauqueur — Nozha Boujemaa

*Projet Imedia - INRIA
BP 105 - F-78153 Le Chesnay
{Julien.Fauqueur,Nozha.Boujemaa}@inria.fr*

RÉSUMÉ. Les deux problèmes posés par un système de recherche d'images par régions sont la définition et la description automatiques de régions. Dans cet article nous présentons dans un premier temps une technique non supervisée de détection de régions grossières qui améliore leur spécificité visuelle. La segmentation est basée sur la classification de Distributions Locales de Couleurs Quantifiées (ou LDQC pour Local Distribution of Quantized Colors). L'algorithme de classification par agglomération compétitive (CA) est utilisé. Il a l'avantage de déterminer automatiquement le nombre de classes. Puis, en considérant que la description d'une région doit être plus fine pour les régions que pour les images, nous proposons un descripteur de variabilité couleur de région : la Distribution Adaptative de Nuances de Couleurs (ou ADCS pour Adaptive Distribution of Colors Shades). Comparée aux descripteurs couleur existants, la haute résolution couleur de ADCS améliore la similarité perceptuelle des régions retrouvées. En conséquence, le système de recherche d'images par régions d'intérêt est plus intuitif.

ABSTRACT. The two major problems raised by a region-based image retrieval system are the automatic definition and description of regions. In this paper we first present a technique of unsupervised coarse detection of regions which improves their visual specificity. The segmentation scheme is based on the classification of Local Distributions of Quantized Colors (LDQC). The Competitive Agglomeration (CA) classification algorithm is used which has the advantage to automatically determine the number of classes. Then, considering that region description must be finer for regions than for images, we propose a region descriptor of fine color variability: the Adaptive Distribution of Color Shades (ADCS). Compared to existing color descriptors, the high color resolution of ADCS improves the perceptual similarity of retrieved regions. As a consequence, the proposed region-based image retrieval scheme is more efficient.

MOTS-CLÉS : indexation et recherche d'image, segmentation, description couleur fine de région, classification, distribution de couleur.

KEYWORDS: image indexing and retrieval, segmentation, fine region color descriptor, classification, color distribution.

1. Introduction

La première approche de recherche d'images par le contenu est l'approche globale de recherche par l'exemple initialement développée dans les systèmes PhotoBook [PEN 94], QBic [FLI 95] ou Virage [GUP 96] et plus récemment par ImageRover [SCL 97], PicToSeek [GEV 99], Compass [BRU 00], Ikona [BOU 01] : à partir d'une image exemple sélectionnée par l'utilisateur, le système lui retourne les images de la base qui lui sont proches visuellement. La similarité visuelle est alors mesurée sur les images *entières*. Cependant, l'utilisateur souhaite généralement retrouver des objets similaires plutôt que des images dans leur globalité. La recherche globale est alors approximative car elle prend en compte toutes les parties d'images, y compris celles qui ne sont pas pertinentes pour la recherche. Ce biais est d'autant plus gênant dans des bases d'images naturelles composites.

Nous souhaitons permettre à l'utilisateur de spécifier la partie qui l'intéresse dans une image et de retrouver des parties similaires dans d'autres images de la base. Ce mode de recherche permet à l'utilisateur d'exprimer plus précisément ses préférences.

La détection automatique d'objets dans une image quelconque uniquement par extraction de primitives visuelles est une tâche très complexe surtout si l'on considère qu'un même objet peut avoir des aspects visuels différents (un vêtement par exemple). Au cours des dix dernières années, différentes approches de requêtes partielles ont été proposées pour identifier et comparer des parties d'images dans un contexte d'indexation :

- rétroprojection de primitive : les parties similaires dans les images candidates sont identifiées dynamiquement : approche souple mais consommatrice en temps de calcul au moment de la requête (par exemple dans VisualSeek [SMI 96]) ;
- subdivision systématique d'images en carrés : simple et rapide mais imprécis (voir [MOG 99, MAL 99]) ;
- subdivision manuelle : le plus proche de l'attente de l'utilisateur mais non viable pour des grosses bases d'images (voir [DEL 98]) ;
- segmentation non supervisée d'images en régions : les régions sont détectées automatiquement (voir [CAR 99][MA 99]) ;
- points d'intérêt : ils permettent d'effectuer des recherches précises mais au prix d'un temps de calcul très élevé (voir [SCH 98] et [GOU 01]).

Nous adoptons l'approche de segmentation en régions qui fournit une approximation naturelle des objets. Pour satisfaire les besoins d'un système de recherche par régions, nous proposons une approche de segmentation rapide et non supervisée pour détecter des régions grossières et pertinentes pour l'utilisateur.

Etat de l'art en recherche de régions

Concevoir un système de recherche par régions demeure un problème ouvert et non trivial : la détection automatique de régions d'intérêt est une tâche difficile et la

description de régions doit prendre en compte leur spécificité visuelle. Les systèmes existants diffèrent sur ces deux points. Parmi ces quelques systèmes nous pouvons citer Blobworld [CAR 99] et Netra [MA 99].

Dans Blobworld [CAR 02], la segmentation région est effectuée par classification de vecteurs mixtes de couleur et texture avec EM (Expectation/Maximization) qui a l'inconvénient de supposer connu le nombre de classes à déterminer. La segmentation est approximative et beaucoup de petites zones sont supprimées au sein de plus grosses régions cohérentes. Le système Netra [MA 99] utilise l'approche contour « Edgeflow » [MA 97]. Les régions résultantes sont satisfaisantes, mais la technique prend environ plus d'une minute par image (contre moins de deux secondes avec notre technique comme nous le verrons en section 7). Le programme de cet algorithme a été testé à partir du programme disponible sur leur site web¹. Dans tous les cas, les segmentations proposées correspondent à un niveau de granularité *fine* dans la détection des régions. Une technique plus récente impliquant la mise en correspondance de régions pour la recherche d'images est proposée par Wang, Li et Wiederhold dans le système SIMPLicity [WAN 01] et plus tard par Sridhar, Nascimento et Li [SRI 02, STE 02]. La similarité entre deux images est mesurée comme combinaison des similarités entre les régions constituantes de chaque image. Mais le système effectue en fait une *recherche globale* car toutes les primitives de l'image sont prises en compte. De plus, la qualité de la segmentation n'est pas leur objectif.

Concernant la description couleur des régions, les systèmes existants appliquent aux régions les descripteurs utilisés pour la recherche *globale* d'images. Ces descripteurs globaux correspondent à la distribution couleur calculée sur une palette constante de 200 couleurs (en moyenne sur les différents systèmes). Bien qu'efficaces pour décrire l'apparence visuelle globale des images, leur faible résolution couleur les rend moins adaptés pour décrire les régions qui sont, par construction, plus homogènes que les images.

Notre approche diffère par la façon d'extraire et de décrire les régions. Dans chaque image, nous cherchons à détecter les zones saillantes et ignorer les détails trop fins. Les régions détectées doivent intégrer une plus grande variabilité photométrique afin d'être visuellement plus caractéristiques les unes des autres dans la base. Elles doivent aussi correspondre à des zones intuitives pour l'utilisateur afin de constituer des régions requête potentielles. Le principe de segmentation proposé ici est basé sur la classification de distributions locales de couleurs évaluées sur de larges voisinages avec une faible granularité de classification.

Concernant la description, plutôt que de décrire les régions à partir d'une palette constante de 200 couleurs, nous proposons de définir un ensemble adaptatif de couleurs déterminées à une haute résolution qui sont pertinentes pour chaque région. Le descripteur ADCS sera la distribution de ces couleurs dans la région.

1. <http://vision.ece.ucsb.edu/segmentation/edgeflow/software/index.html>

L'idée clé de la détection grossière de région et de la signature fine est la suivante : l'importante variabilité visuelle au sein d'une région est décrite fidèlement par la haute résolution des nuances de couleurs, de sorte que les régions soient vraiment spécifiques les unes par rapport aux autres dans la base. On assure la cohérence entre les étapes de détection et de description des régions car elles sont formées de distributions locales similaires et comparées avec des distributions de nuances de couleur. Une partie de ces travaux a été publiée dans [FAU 02, FAU 03b, FAU 03c].

Nous présenterons également des résultats de la combinaison de ce nouveau descripteur photométrique à des descripteurs géométriques. Bien qu'élémentaires, nous verrons que ces derniers présentent un intérêt évident pour la recherche des régions similaires.

Dans la prochaine section, nous expliquerons l'algorithme de classification par agglomération compétitive (CA), une technique essentielle dans notre travail, car elle est utilisée aussi bien à la phase de segmentation que de description. L'extraction de régions par classification de distributions locales de couleurs sera développée dans la section 3. Le descripteur de variabilité couleur ADCS sera détaillé dans la section 4 ainsi que le processus de recherche de régions. Nous présenterons aussi dans la section 5 l'interface utilisateur de la plate-forme Ikona de recherche de régions. Puis suivront la description des conditions de test et les résultats dans les sections 6 et 7. La performance de recherche du descripteur ADCS sera testée comparativement à l'histogramme couleur classique et à la combinaison de lui-même avec des descripteurs géométriques simples. Nous concluons en section 8 et proposerons des pistes de recherche future.

2. Regroupement de primitives visuelles

Aussi bien pour extraire que pour décrire les régions, notre approche nécessitera d'effectuer des regroupements de primitives visuelles de natures différentes et de façon non supervisée. Notre choix s'est porté sur CA, l'algorithme de classification par agglomération compétitive, initialement présenté dans [FRI 97]. Il s'agit d'une extension évoluée de l'algorithme de classification des C-Moyennes Floues (FCM) [BEZ 81]. CA présente l'avantage majeur de déterminer automatiquement le nombre de classes.

En utilisant les notations de [FRI 97], nous appellerons $\{x_j, \forall j \in \{1, \dots, N\}\}$ l'ensemble des N données que nous souhaitons classifier et C le nombre de classes. $\{\beta_i, \forall i \in \{1, \dots, C\}\}$ représentent les prototypes de classes à déterminer. La distance entre une donnée x_j et un prototype β_i est $d(x_j, \beta_i)$ et doit être choisie efficacement en fonction des données à classifier. La classification CA est obtenue par minimisation de la fonctionnelle J suivante :

$$J = J_1 + \alpha J_2, \quad (1)$$

où

$$J_1 = \sum_{i=1}^C \sum_{j=1}^N u_{ij}^2 d^2(x_j, \beta_i) \quad (2)$$

et

$$J_2 = - \sum_{i=1}^C \left[\sum_{j=1}^N u_{ij} \right]^2 \quad (3)$$

sous la contrainte d'appartenance :

$$\sum_{i=1}^C u_{ij} = 1, \forall j \in \{1, \dots, N\} \quad (4)$$

où u_{ij} représente le degré d'appartenance de la donnée x_j au prototype β_i . Minimiser séparément J_1 revient à effectuer une classification FCM qui détermine C prototypes optimaux et la partition floue U , étant donné x_j et C , en utilisant la distance d . J_2 est un terme de validité de classification (voir [BOU 00]). J est ainsi écrit comme la combinaison de deux termes d'effets opposés J_1 et J_2 . Alors la minimisation de J avec une surestimation du nombre de classes initiales classe les données et, simultanément, détermine automatiquement le nombre de classes C .

J est minimisé récursivement. α est le poids de compétition ; il doit permettre un équilibre entre les termes J_1 et J_2 dans l'équation (1). A l'itération k , le poids α s'exprime de la façon suivante :

$$\alpha(k) = \eta_0 \exp\left(\frac{-k}{\tau}\right) \frac{\sum_{i=1}^C \sum_{j=1}^N u_{ij}^2 d^2(x_j, \beta_i)}{\sum_{i=1}^C \left[\sum_{j=1}^N u_{ij} \right]^2} \quad (5)$$

Au fil des itérations, α décroît, donc l'importance est d'abord accordée au processus d'agglomération des classes, puis à l'optimisation de la classification. α est complètement déterminé par les paramètres η_0 et τ .

Pendant l'algorithme, les classes à faible population sont supprimées, *i.e.* celles dont la quantité $\sum_{j=1}^N u_{ij}$ pour une classe i est inférieure à un seuil fixé ϵ . La convergence est établie lorsque les prototypes sont stables. La granularité de la classification est contrôlée par la population minimale ϵ et le facteur α grâce à son amplitude η_0 et son taux de décroissance τ . Etant donné une granularité de classification, CA détermine automatiquement le nombre de classes.

CA sera utilisé à 3 étapes dans notre système avec différents niveaux de granularité, dans l'ordre suivant :

- quantification couleur de l'image (classification de triplets de couleur) ;
- regroupement des primitives LDQC (classification de distributions couleur) ;
- description des régions avec ADCS (classification de triplets de couleur).

Pour plus de détails sur l'agglomération compétitive, nous invitons le lecteur à se référer à l'article original [FRI 97].

3. Extraction des régions d'intérêt par segmentation grossière

La nature composite des images que nous nous proposons de segmenter présente des niveaux de détails variables. Dans le cadre de la recherche d'images par région-exemple, nous allons nous intéresser aux parties saillantes de l'image susceptibles de constituer des clés de requête pertinentes pour l'utilisateur. Nous proposons une approche de segmentation grossière permettant de définir des régions à la fois homogènes photométriquement et englobant une certaine variabilité visuelle. Les saillances géométriques ou photométriques correspondant à un niveau trop fin de détail seront naturellement intégrées dans les régions plus grandes. La variabilité rendra les régions plus caractéristiques les unes des autres dans la base. De plus la technique de segmentation devra être rapide et non supervisée. Le temps d'exécution est un facteur important car nous traitons des bases pouvant contenir plusieurs milliers d'entrées. Mais la rapidité ne doit pas être synonyme de qualité médiocre : les régions obtenues doivent être intuitives pour l'utilisateur.

Notre approche de segmentation est basée sur la classification CA des distributions locales de couleurs de l'image. Ces primitives intègrent naturellement la diversité des couleurs dans les voisinages des pixels. La classification de telles primitives représentatives de larges voisinages donne des régions cohérentes plus naturellement ; en effet, la formation de régions par regroupement de larges voisinages similaires conduit à des régions naturellement plus grossières et nécessite moins de post-traitement spatial. Ces primitives sont cohérentes avec le descripteur couleur ADCS qui sera présenté à la section 4.

Le choix de l'ensemble de couleurs pour calculer ces distributions locales de couleurs est crucial : il doit être compact pour gagner en vitesse de classification et représentatif d'un voisinage de pixels. Les images comportent, en moyenne, 60 000 couleurs originales différentes. Nous définissons cet ensemble de couleurs comme l'ensemble adaptatif obtenu par quantification couleur de l'image pour réduire considérablement le nombre de couleurs à traiter sans perte importante de l'information perceptuelle. Puis, pour chaque voisinage de l'image, la distribution locale est évaluée sur cet ensemble de couleurs réduit et pertinent. Nous les nommerons Distributions Locales de Couleurs Quantifiées (ou *LDQC* pour *Local Distribution of Quantized Colors*). Après classification, chaque distribution LDQC sera remplacée dans l'image par l'étiquette de sa classe. Les étiquettes connexes formeront les régions et les petites régions seront fusionnées ou supprimées à l'aide d'un graphe d'adjacence de régions.

En résumé, les trois phases principales de la segmentation sont les suivantes :

- quantification couleur (produit les *couleurs quantifiées*) ;
- classification des distributions locales de couleurs quantifiées (*LDQC*) ;
- fusion et suppression des petites régions.

3.1. Quantification couleur de l'image

Cette étape de quantification couleur de l'image vise à réduire le nombre de couleurs sur lesquelles les distributions locales (LDQC) vont être évaluées, tout en conservant suffisamment d'information colorimétrique pour distinguer les régions saillantes entre elles. Différentes approches de quantification couleur ont été proposées qui varient en complexité de calculs et en précision (voir [SAN 99]). Elles reposent sur une subdivision plus ou moins adaptative (donc plus ou moins coûteuse et plus ou moins précise) de l'espace couleur en plans ou cellules. De nombreuses approches supposent donné le nombre de couleurs de la palette, ce qui n'est pas notre cas. Nous souhaitons en effet que le nombre de couleurs quantifiées dépende de la complexité photométrique de l'image.

La méthode reconnue de quantification couleur *General Lloyd Algorithm* (aussi appelée *Linde-Buzo-Gray* ou encore *K-Means*) est souvent utilisée dans la littérature. Elle minimise les distances entre les données couleurs et la palette de quantification de façon itérative. Basée sur le même principe, l'approche que nous proposons consiste à effectuer la classification par CA des pixels de l'image. Notre approche présente l'avantage majeur de déterminer automatiquement le nombre de couleurs quantifiées (qui seront les centroïdes des classes). De plus, elle introduit le calcul des appartenances floues qui gère mieux les données ambiguës.

Concernant l'espace de couleur, la classification reposant fortement sur la métrique, il est nécessaire de choisir un espace perceptuellement uniforme (les plus connus étant LUV et LAB)[SAN 99]. Ces espaces ont été conçus de sorte que les différences de couleurs jugées égales par l'œil humain soient aussi égales en distance euclidienne. Par contre, les espaces RGB et HSV ne sont pas perceptuellement uniformes [ROB 88]. Le modèle HSV est intuitif, mais souffre d'irrégularités (la composante H cyclique et H et v n'ont pas de sens pour s petit). RGB présente généralement l'avantage de ne nécessiter aucune transformation, mais sa topologie ne rend pas compte des distances couleurs perçues par l'œil humain.

Le choix de quantification couleur par classification des pixels dans l'espace LUV muni de la distance L^2 semble un choix naturel. L'espace LUV a été préféré à LAB en raison d'un coût de transformation moindre à partir de RGB.

La granularité de la classification (voir section 2) a été choisie de sorte que les grandes zones d'images avec une forte texture soient représentées par plus d'une couleur. A la convergence de la classification, les prototypes des différentes classes obtenues définissent l'ensemble des n couleurs quantifiées. Comme CA détermine automatiquement le nombre de classes, le nombre de couleurs quantifiées n sera représentatif de la diversité couleur des images naturelles.



Figure 1. A gauche : l'image originale (43.217 couleurs). A droite : l'image quantifiée (27 couleurs)



Figure 2. Illustration des LDQC : exemple de calcul sur trois fenêtres

3.2. Regroupement des primitives locales LDQC

3.2.1. Extraction des LDQC

L'image quantifiée ainsi produite fait déjà apparaître visuellement des zones saillantes. Mais ces zones ne peuvent pas être extraites automatiquement à ce stade, car elles comportent différentes couleurs quantifiées dès qu'elles sont texturées. Afin de saisir les caractéristiques visuelles des zones (uniformes ou texturées), il est donc nécessaire d'extraire des primitives sur des voisinages. Les Distributions Locales de Couleurs Quantifiées (LDQC) vont permettre de détecter les variations locales des couleurs dans l'image quantifiée et le regroupement des LDQC de l'image formera la segmentation en régions.

Pour extraire l'ensemble des LDQC de l'image, nous déplaçons une fenêtre sur les pixels de l'image quantifiée et évaluons la distribution locale de couleurs quantifiées au sein de la fenêtre. La figure 2 montre trois exemples de LDQC sur une image en des lieux de couleurs et de variations différentes. Dans le cas d'une image 500x400, la largeur de fenêtre considérée est de 31 pixels et le pas de déplacement, défini comme une demi-largeur de fenêtre, vaut 16 pixels. Le nombre de LDQC extraits de cette image se calcule ainsi : $(500/16) \times (400/16) = 31 \times 25 = 775$.

3.2.2. Distance Couleur Quadratique pour le regroupement des LDQC

Les LDQC extraites dans l'image vont être regroupées par classification avec CA. La classification CA dépend notamment de la définition d'une distance dont la capacité à mesurer les similarités entre les LDQC influera sur la qualité de la segmentation. Les distances couramment utilisées pour comparer des distributions de couleurs, telles les distances L^p , reposent sur une simple comparaison cellule à cellule des distributions (on entend par *cellule* la cellule d'histogramme ou *bin*). Cette approche suppose implicitement que les informations associées à chaque cellule (c'est-à-dire les couleurs ici) sont indépendantes entre elles, ce qui est faux puisque entre toute paire de couleurs on peut mesurer une distance. On pourrait qualifier de « daltoniennes » ces distances de distributions dans la mesure où elles ne prennent pas en compte l'information de couleur associée à chaque cellule. Bien que ces distances donnent des résultats satisfaisants dans le cas, par exemple, de la recherche globale d'images par distributions de couleur, elles s'avèrent trop imprécises pour notre problème. Un exemple de leurs limitations est qu'elles considèrent à distance maximale les distributions dont les histogrammes ne s'intersectent pas. Or, des zones de pixels dont les histogrammes de couleur ne s'intersectent pas peuvent aussi bien paraître visuellement proches que très éloignées. Plus les distributions sont piquées (données homogènes), plus ce phénomène se produit facilement.

La distance couleur quadratique [HAF 95] développée dans le cadre du système QBic, constitue une solution élégante à ce problème en intégrant la similarité couleur intercellules. Si l'on définit X et Y deux distributions de couleurs sur les n couleurs quantifiées, on peut les écrire comme paires de couleur/pourcentage :

$$X = \{(c_1, p_1^X), \dots, (c_n, p_n^X)\} \text{ et } Y = \{(c_1, p_1^Y), \dots, (c_n, p_n^Y)\}.$$

Alors la distance quadratique est définie ainsi :

$$\begin{aligned} d_q(X, Y)^2 &= (X - Y)^T A (X - Y) \\ &= \sum_{i=1}^n \sum_{j=1}^n (p_i^X - p_i^Y)(p_j^X - p_j^Y) a_{ij} \end{aligned} \quad (6)$$

où $A = [a_{ij}]$ est la matrice de similarités couleurs a_{ij} entre les couleurs c_i et c_j :

$$a_{ij} = 1 - d_{ij}/d_{max} \quad (7)$$

où d_{ij} est la distance euclidienne dans l'espace LUV et d_{max} le maximum de cette distance dans l'espace couleur. Notons que si l'on remplace A par la matrice identité, nous retrouvons la distance euclidienne elle-même, *i.e.* $d_q(X, Y) = \|X - Y\|_{L^2}$.

Les figures 3 et 4 illustrent l'amélioration apportée par la distance quadratique par rapport à la distance L^1 . Nous avons choisi 3 imagettes « ciel », « brique » et « osier ». Du fait de leur homogénéité, elles peuvent aussi bien correspondre à un voisinage de pixels qu'à une région. Chacune a été transformée selon 6 facteurs d'intensité. Nous obtenons un ensemble de trois familles d'imagettes d'homogénéités et d'intensités

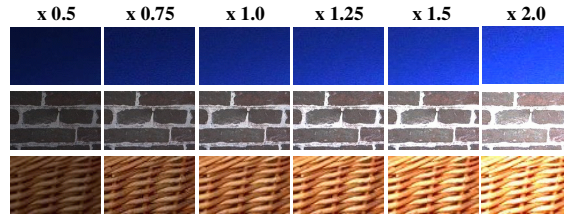


Figure 3. 3 imagettes extraites de régions d'images ont été transformées selon 6 facteurs d'intensité. Elles diffèrent en couleurs et textures

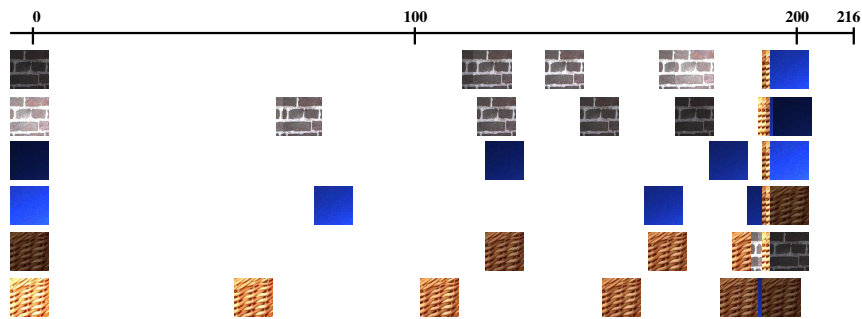


Figure 4. 6 comparaisons avec L^1 . Sur chaque ligne, les imagettes sont positionnées selon la valeur de leur distance L^1 en distribution à la première imagette. La graduation indique les valeurs de ces distances à l'exemple. Les imagettes très similaires sont correctement jugées les plus proches avec L^1 , mais les autres sont agglomérées à proximité de la distance maximum (autour de 200) et ne sont plus discernables les unes des autres : en moyenne, 13 imagettes sur 18 se trouvent à une distance dans l'ensemble $\{197, 198, 199, 200, 201\}$. Ceci illustre l'influence de l'homogénéité des données sur l'imprécision des distances de types cellule à cellule

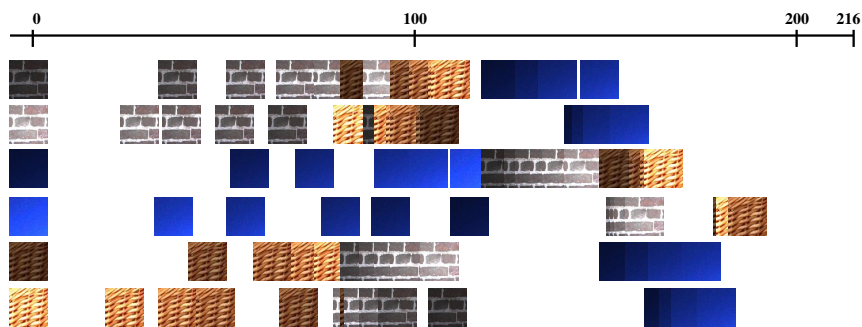


Figure 5. 6 comparaisons avec d_{quad} . Même illustration qu'à la figure 3 mais avec la distance quadratique. Les distances sont plus étalées qu'avec L^1 et fournissent des résultats nettement plus satisfaisants. Notre perception du continuum visuel entre les images est mesurée plus fidèlement avec la distance quadratique

différentes, soit 18 images. Les distributions de couleur des 3 images les plus claires (correspondant au facteur 2) et celles des 3 les plus sombres (facteur 0.5) ont été comparées à celles des 17 autres imagerie avec la distance L^1 (fig. 3) et avec la distance quadratique (fig. 4).

On observe la capacité de la distance quadratique à mesurer le continuum visuel entre les différentes intensités d'une même imagerie tout en discernant les familles d'imagerie entre elles (ciel, brique, osier), bien que les distributions soient piquées donc d'intersections parfois nulles. Il est important de noter que ces propriétés de continuité de la distance quadratique par rapport aux changements d'intensité est aussi vérifiable pour les changements de couleurs et de manière générale à tout changement qui entraîne une translation de la distribution sur des cellules voisines. Alors que ces changements pourront se traduire par des valeurs maximales avec la distance L^1 , ils correspondront à des variations progressives de la distance quadratique. Il en résulte une meilleure fidélité dans la mesure de similarité visuelle. Cette propriété de la distance quadratique nous intéressera pour la classification des LDQC (distributions locales de couleurs) et la comparaison des ADCS (distributions de couleurs sur des régions).

3.2.3. Regroupement des LDQC

L'algorithme de classification CA tel qu'il est présenté à la section 2 est utilisé pour classer les LDQC extraites de l'image avec la distance quadratique.

A la convergence de la classification, on obtient un ensemble de classes de LDQC ainsi que leurs prototypes. L'image segmentée est obtenue en attribuant à chaque pixel l'étiquette de la classe à laquelle la LDQC autour de ce pixel appartient. Un filtre de vote est ensuite appliqué pour supprimer les étiquettes isolées dans l'espace image.

3.3. Information d'adjacence

A ce point, nous avons obtenu une partition complète de l'image en régions adjacentes. Quelques régions peuvent s'avérer trop petites pour constituer des régions d'intérêt ; elles augmentent alors inutilement le nombre total de régions dans la base. De plus, dans des scènes complexes, elles sont souvent situées à la frontière entre deux grandes régions ou à l'intérieur d'une grande région. Elles doivent être fusionnées pour améliorer la topologie des régions d'intérêt.

Nous désirons que chaque région couvre au minimum 1,5 % de la surface de l'image. En dessous de ce seuil, une région est fusionnée à sa région voisine la plus proche visuellement si elle en a une. Deux régions sont dites « proches visuellement » si la moyenne des LDQC qu'elles contiennent sont proches, c'est-à-dire si la distance quadratique entre les LDQC moyennes est inférieure à un seuil fixé. La stratégie itérative de fusion des petites régions est la suivante : tant qu'il existe une région de surface inférieure à 1,5 % et possédant une région voisine « proche visuellement », on fusionne la petite région dans sa voisine. A l'issue des itérations, s'il reste des petites

régions (*i.e.* qui n'ont pas été jugées suffisamment proches de leurs voisines pour être fusionnées), elles sont supprimées et ne sont pas indexées.

La fusion de régions s'opère à l'aide d'un graphe d'adjacence de régions [PAV 77] (ou RAG pour *Region Adjacency Graph*). Les attributs des régions (surface, distribution de couleur, contours, barycentre) sont stockés dans les nœuds et l'information d'adjacence de régions dans les arêtes du graphe (adjacence, longueur des contours communs). Notons que l'étape de fusion de régions à l'aide du graphe est un traitement marginal dans notre approche. La formation des régions repose principalement sur le regroupement des primitives LDQC et l'étape de fusion constitue un post-traitement de la segmentation.

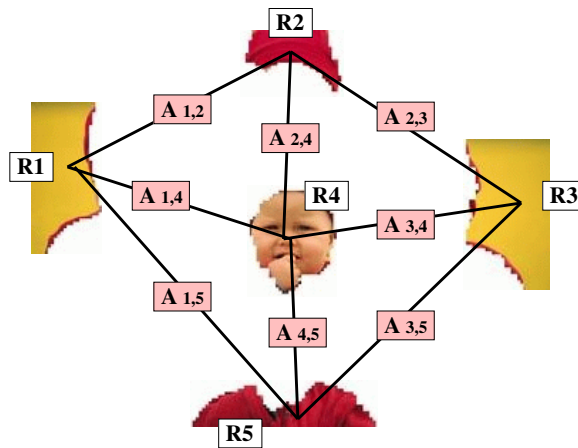


Figure 6. Structure RAG de l'image partitionnée. Le RAG représente les attributs des régions R_i dans les nœuds et les informations d'adjacence entre toutes les paires (i, j) de régions adjacentes dans les arêtes du graphe. La segmentation exploite ces informations pour la fusion des régions



Figure 7. L'image segmentée finale est représentée avec la couleur moyenne originale de chaque région

4. Indexation et recherche de régions

4.1. ADCS, une signature fine de variabilité couleur pour les régions

Une fois les régions détectées de façon grossière, nous souhaitons décrire finement leur apparence.

La plupart des systèmes de recherche par régions dérivent les descripteurs couleur globaux pour les régions. Ils consistent en des distributions de couleurs calculées sur le sous-échantillonnage prédéfini d'un espace de couleur. Le nombre de ces couleurs est de 166 dans VisualSeek [SMI 96] et 218 dans Blobworld [CAR 99] (en fait réduit à 5 par décomposition en valeurs propres). Dans Netra [DEN 99] une moyenne de 3,5 couleurs dominantes est choisie dans une table de 256 couleurs qui est construite pour chaque base. Le point commun entre les descripteurs couleur existants est l'utilisation d'un ensemble fixé de couleurs comportant en moyenne 200 couleurs supposées couvrir l'intégralité de l'espace couleur qui peut en contenir des millions (qu'il s'agisse de RGB, LAB, LUV ou HSV). Choisir systématiquement les mêmes 200 couleurs pour chaque région parmi les millions de couleurs possibles dégrade fortement la résolution couleur.

Une telle description force la distance minimum entre deux couleurs à être élevée car le sous-échantillonnage est fixé et car on ne considère que peu de couleurs parmi plusieurs millions. Cette perte de résolution couleur peut ne pas affecter la description globale d'images naturelles car celles-ci comportent une vaste dynamique de couleurs mais les régions sont, par construction, plus homogènes que les images. Comparée à la base d'index d'un système classique de recherche globale d'images, la spécificité d'une base d'index pour la recherche de régions est qu'elle contient plus de descripteurs (car une image définit plusieurs régions) et qu'ils décrivent des zones visuelles plus homogènes. Le besoin d'une résolution plus fine, donc plus fidèle au contenu visuel, est double.

Nous proposons de décrire chaque région avec la distribution adaptative de nuances de couleurs (ou ADCS pour *Adaptive Distribution of Colors Shades*). Pour chaque région de l'image originale, ses pixels couleur sont regroupés avec l'algorithme CA. La classification est effectuée dans l'espace couleur LUV avec la distance euclidienne. Les prototypes couleur obtenus fournissent l'ensemble de nuances de couleurs. Il s'agit ici d'effectuer une *quantification adaptative à chaque région*. Bien que la méthode de quantification employée ici soit celle présentée pour le prétraitement de la segmentation en section 3.1, il serait erroné d'utiliser les mêmes couleurs quantifiées pour décrire les régions².

2. En effet, dans la quantification globale, la participation des pixels d'une région dans la classification CA est proportionnelle à la taille de la région, ce qui signifierait que les couleurs quantifiées seraient plus représentatives des grandes régions. Donc, afin d'assurer une finesse de description égale pour chaque région, il est nécessaire de classifier séparément les pixels de chaque région.

Nous employons les termes *nuances de couleurs* par opposition aux *couleurs dominantes* pour exprimer la présence de couleurs mineures. L'index est constitué de la distribution de ces nuances de couleurs : il représente la variabilité couleur de la région et donne une approximation de l'apparence de texture des régions. Plus la diversité couleur d'une région est grande, plus CA détectera de nuances de couleur. La nature et le nombre de ces nuances de couleur ne sont pas fixés mais spécifiques à chaque région, contrairement aux descripteurs existants. De plus, les nuances sont choisies parmi l'espace couleur LUV *entier* qui contient 5,6 millions de couleurs potentielles, tandis que les descripteurs couleur existants ne peuvent pas en distinguer plus de 200 en moyenne.

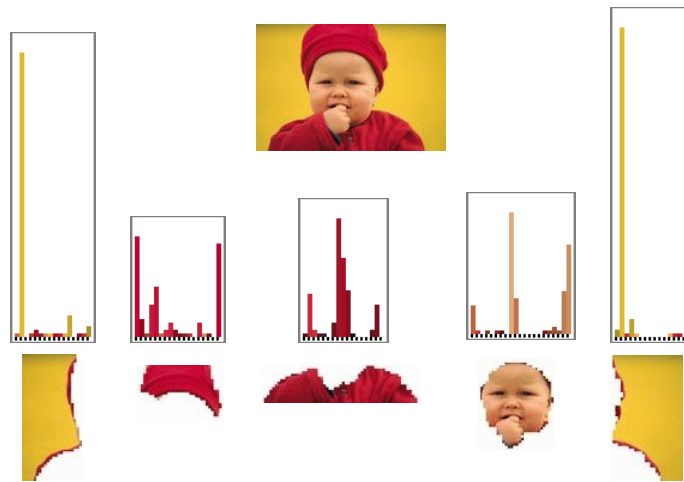


Figure 8. *l'image originale, ses régions détectées et leurs index ADCS respectifs. Notons que l'ordre des couleurs dans la représentation des ADCS est sans importance. Nous remarquons que les zones texturées du bonnet et du manteau sont représentées par des nuances différentes de rouges et que le fond jaune quasiment uniforme est décrit par un pic majeur de jaune et par quelques autres couleurs mineures. L'ensemble des couleurs sélectionnées est pertinent pour chaque région associée*

4.2. Descripteurs géométriques

Pour tester la combinaison de ADCS avec d'autres types de descripteurs nous avons implémenté deux descripteurs géométriques simples de région : la surface et le barycentre. La surface est représentée comme le rapport de surface de la région sur celle de l'image et le barycentre comme le couple de coordonnées (x, y) normalisées par rapport à la largeur et la hauteur de l'image respectivement. La distance L^1 est utilisée pour comparer les surfaces et L^2 pour les barycentres.

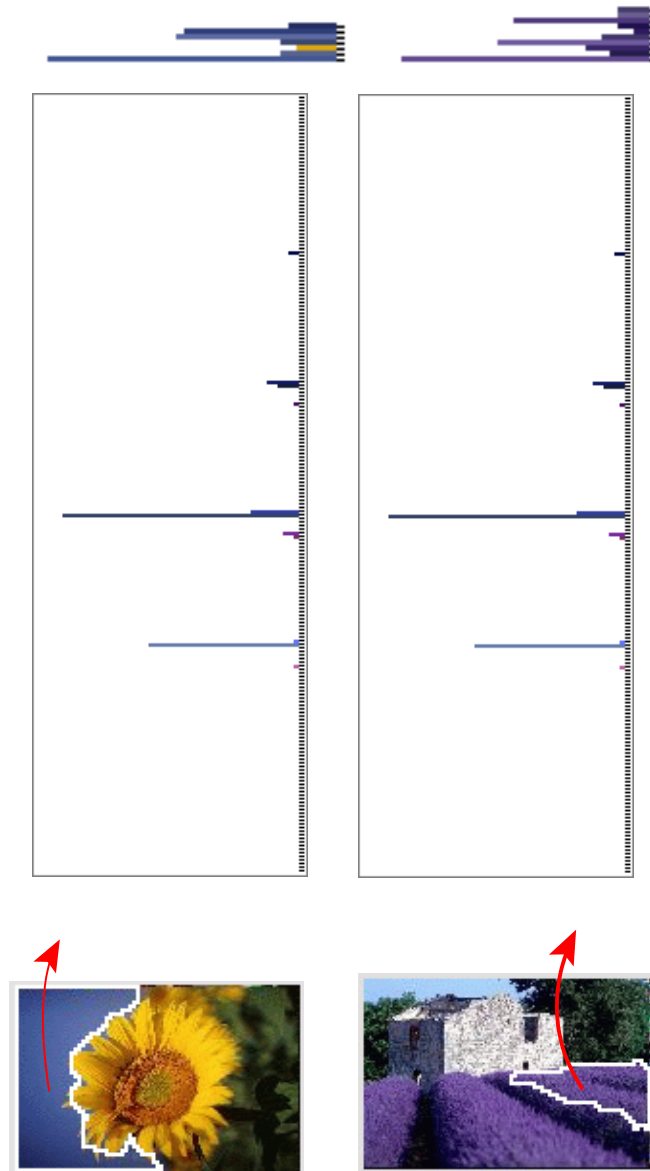


Figure 9. Limitation de l'histogramme couleur classique : deux régions perceptuellement différentes peuvent avoir la même description (au milieu). Sa faible résolution couleur amène des couleurs perceptuellement différentes dans des mêmes cellules, tandis que les nuances de couleur du ADCS offrent une description plus fidèle (en haut)

4.3. Mise en correspondance de régions

Deux index ADCS ne sont *a priori* pas exprimés sur le même ensemble de nuances de couleurs et nous devons utiliser une distance adéquate pour les comparer.

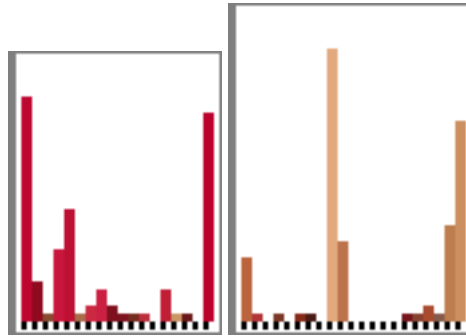


Figure 10. Exemple de deux index ADCS à comparer : ils sont exprimés sur deux ensembles différents de couleurs

La distance quadratique est ici aussi le choix pertinent pour comparer les distributions ADCS, pour deux raisons : la première concerne ses propriétés de continuité relativement aux changements photométriques telles qu'elles ont été illustrées en 3.2.2, qui en font une distance pertinente pour comparer les apparences visuelles des régions de la base. La seconde raison est que sa précision n'est pas affectée par les intersections nulles d'histogrammes, contrairement aux distances L^p . En effet, la résolution couleur des ADCS étant fine, la comparaison de deux histogrammes ADCS quelconques a de grandes chances statistiquement de donner une intersection nulle.

Pour une région requête donnée de distribution ADCS X , les régions similaires sont telles que leur distribution Y minimise $d_{quad}(X, Y)$. Ecrivons X et Y comme paires de couleur/pourcentage :

$$X = \{(c_1^X, p_1^X), \dots, (c_{n_X}^X, p_{n_X}^X)\} \text{ et}$$

$$Y = \{(c_1^Y, p_1^Y), \dots, (c_{n_Y}^Y, p_{n_Y}^Y)\}$$

La formule (6) donne la distance quadratique entre deux distributions de couleurs x et y évaluées sur le *même* ensemble de couleurs, ce qui n'est *a priori* pas le cas pour deux distributions pour ADCS.

A présent, nous introduisons la forme généralisée de la distance quadratique qui va permettre de comparer des distributions exprimées sur des ensembles de couleurs *différents*. Nous reformulons l'expression (6) de la distance quadratique pour supprimer les termes impliquant des différences de cellule. Considérons x comme l'extension de la distribution X sur la *totalité* de l'espace couleur et y comme l'extension de Y .

L'extension consiste à mettre à zéro les valeurs de cellules pour lesquelles les couleurs ne sont pas des nuances de couleur. Donc nous avons $d_{quad}(x, y) = d_{quad}(X, Y)$.

$$\begin{aligned} d_{quad}(X, Y)^2 &= d_{quad}(x, y)^2 \\ &= (x - y)^T A (x - y) \\ &= x^T A x - x^T A y - y^T A x + y^T A y \end{aligned} \quad (8)$$

Comme la matrice A de similarité couleur est symétrique, nous en déduisons l'égalité suivante :

$$d_{quad}(x, y)^2 = x^T A x + y^T A y - 2x^T A y \quad (9)$$

Par construction de x et y , nous avons :

$$\begin{aligned} x^T A x &= X^T A^X X \\ y^T A y &= Y^T A^Y Y \\ x^T A y &= X^T A^{XY} Y \end{aligned} \quad (10)$$

où les matrices A^X , A^Y et A^{XY} sont les restrictions de la matrice A et donnent les similarités de couleur entre, respectivement, les nuances de X avec elles-mêmes (de dimension $n_X.n_X$), celles de Y avec elles-mêmes (de dimension $n_Y.n_Y$) et celles de X avec celles de Y (de dimension $n_X.n_Y$). D'où :

$$d_{quad}(x, y)^2 = X^T A^X X + Y^T A^Y Y - 2X^T A^{XY} Y \quad (11)$$

Nous obtenons finalement l'expression suivante *sans différence de cellules*. Elle constitue la **forme généralisée de la distance quadratique** qui permet de comparer, de façon plus générale, deux distributions évaluées sur des *ensembles de couleurs différents* :

$$\begin{aligned} d_{quad}(X, Y)^2 &= \sum_{i,j=1}^{n_X} p_i^X p_j^X a_{ij}^X + \sum_{i,j=1}^{n_Y} p_i^Y p_j^Y a_{ij}^Y - 2 \sum_{i=1}^{n_X} \sum_{j=1}^{n_Y} p_i^X p_j^Y a_{ij}^{XY} \\ &= F(X, X) + F(Y, Y) - 2.F(X, Y) \end{aligned} \quad (12)$$

où l'application F est définie ainsi :

$$F(X, Y) = \sum_{i=1}^{n_X} \sum_{j=1}^{n_Y} p_i^X p_j^Y a_{ij}^{XY} \quad (13)$$

Comme les termes $F(X, X)$ et $F(Y, Y)$ dépendent individuellement d'un seul index, le précalcul de $F(X, X)$ pour chaque région d'index X à la phase d'indexation accélère les calculs de la distance quadratique à la phase de requête. En effet, parmi les trois termes dans (12) seul le terme $F(X, Y)$ sera calculé au moment de la requête.

Notons que dans Netra, la mesure de comparaison de leurs couleurs dominantes est une approximation de la distance quadratique : le terme intercellules n'existe pas

étant donné leur hypothèse selon laquelle les couleurs dominantes d'une même région sont à distance maximale dans l'espace couleur. Nous ne faisons de telle hypothèse car nos nuances de couleurs peuvent être proches entre elles à des degrés différents.

L^1 est la distance choisie pour comparer les surfaces, L^2 pour les barycentres et d_{quad} pour les ADCS. La distance finale entre une région requête R_r et une région candidate R_c est :

$$\begin{aligned} d_{finale}(R_r, R_c) = & \alpha_A \cdot d_{quad}^A(R_r, R_c) + \\ & \alpha_S \cdot d_{L^1}^S(R_r, R_c) + \\ & \alpha_B \cdot d_{L^2}^B(R_r, R_c) \end{aligned} \quad (14)$$

Les poids α_A , α_S et α_B sont initialisés avec des valeurs par défaut qui accordent une importance égale à chaque critère. L'utilisateur a néanmoins la possibilité de les modifier dynamiquement dans notre interface Ikona (voir image 5). Selon la nature de la recherche de régions, plus ou moins d'importance peut être accordée aux poids relatifs entre la similarité de diversité couleur, la similarité de position et la similarité de surface. Puisque $d_{L^1}^S(R_r, R_c)$ et $d_{L^2}^B(R_r, R_c)$ sont des quantités simples et rapides à déterminer, elles vont être utilisées pour rejeter rapidement les régions qui sont d'improbables candidates : pour une région requête donnée R_r , si les quantités $d_{L^1}^S(R_r, R_c)$ ou $d_{L^2}^B(R_r, R_c)$ sont suffisamment grandes, alors $d_{finale}(R_r, R_c)$ sera nécessairement trop grande et il sera alors inutile de calculer la quantité $d_{quad}^A(R_r, R_c)$ plus complexe à évaluer.

Si les poids α_S et α_B sont nuls, cette technique d'optimisation n'est pas mise en œuvre car $d_{finale}(R_r, R_c)$ est réduite à $d_{quad}^A(R_r, R_c)$. La stratégie de rejet se déroule ainsi :

```

initialisation :  $d_{finale}(R_r, R_c) = 0$ 
si (  $\alpha_S > 0$  et  $d_{L^1}^S(R_r, R_c) > \text{SEUIL\_SURFACE}$  )
    alors  $d_{finale}(R_r, R_c) = \text{VALEUR\_PSEUDO\_INFINIE}$ 
sinon
     $d_{finale}(R_r, R_c) += d_{L^1}^S(R_r, R_c)$ 
    si (  $\alpha_B > 0$  et  $d_{L^2}^B(R_r, R_c) > \text{SEUIL\_BARYCENTRE}$  )
        alors  $d_{finale}(R_r, R_c) = \text{VALEUR\_PSEUDO\_INFINIE}$ 
    sinon
         $d_{finale}(R_r, R_c) += d_{L^2}^B(R_r, R_c)$ 
         $d_{finale}(R_r, R_c) += d_{quad}^A(R_r, R_c)$ 

```

La quantité `VALEUR_PSEUDO_INFINIE` est une valeur de distance arbitrairement grande. Nous verrons dans la section 7 la conséquence de cette stratégie de rejet sur le processus de recherche.

5. Interface utilisateur

Le système de recherche de régions est intégré à notre plateforme Ikona, développée selon une architecture client-serveur (serveur écrit en C++ et interface utilisateur client en Java). Un aperçu des fonctionnalités de recherche et de navigation d'Ikona est présenté dans [BOU 01].

Pour le scénario de recherche de régions, l'utilisateur commence à naviguer aléatoirement dans la base grâce à l'interface client (voir capture d'écran 5). Chaque région de chaque vignette image peut être cliquée pour désigner la région requête. Dans une deuxième fenêtre, l'utilisateur peut ajuster l'importance relative des différents descripteurs (ADCS, surface et barycentre) selon que la taille et la position de la région dans l'image sont pertinentes ou non pour le type de régions (ou objets) recherchés.

Le serveur retourne les images comportant une région similaire à la région requête affichées en ordre croissant de distance finale d_{finale} . Dans les images retournées, les contours de la région sont tracés en blanc.

6. Protocole de test

Notre système a été testé avec un PC standard à 2.0 GHz (512 Mo de mémoire vive). La base de test comporte au total 11 479 images (couleur majoritairement et noir et blanc) réparties de la façon suivante : 792 images de textures de la base Vistex, 552 de la banque d'images *Images Du Sud* et 10 135 la base *Corel*. Les deux dernières bases sont des images génériques de fleurs, dessins, portraits, paysages, architecture, fractales, gens, fruits, jardins, voitures, cuisines...

L'évaluation des performances d'un système de recherche d'images par le contenu est une tâche ardue car elle dépend étroitement de la perception visuelle de l'humain, du domaine d'application (générique ici) et du contenu de la base elle-même. Pour un système de recherche de régions, la difficulté est accrue car elle requiert l'identification préalable de régions spécifiques pour la création d'une base vérité-terrain de régions.

Nous avons construit notre base vérité-terrain de régions en partie avec la base Vistex (disponible sur le site web du MIT³) qui est construite à partir de parties d'images de scènes réelles (88 classes de textures différentes où chaque classe se compose de 9 images). L'autre partie de notre base vérité-terrain de régions est constituée de régions détectées par notre système. Elles ont été étiquetées manuellement pour définir leur appartenance à l'une des trois classes suivantes : personne (régions de peau), lavande, piscine. Au sein de chaque classe, les régions se réfèrent au même *objet*, mais sont aussi perceptuellement proches.

3. <http://www-white.media.mit.edu/vismod/imagery/VisionTexture/vistex.html>

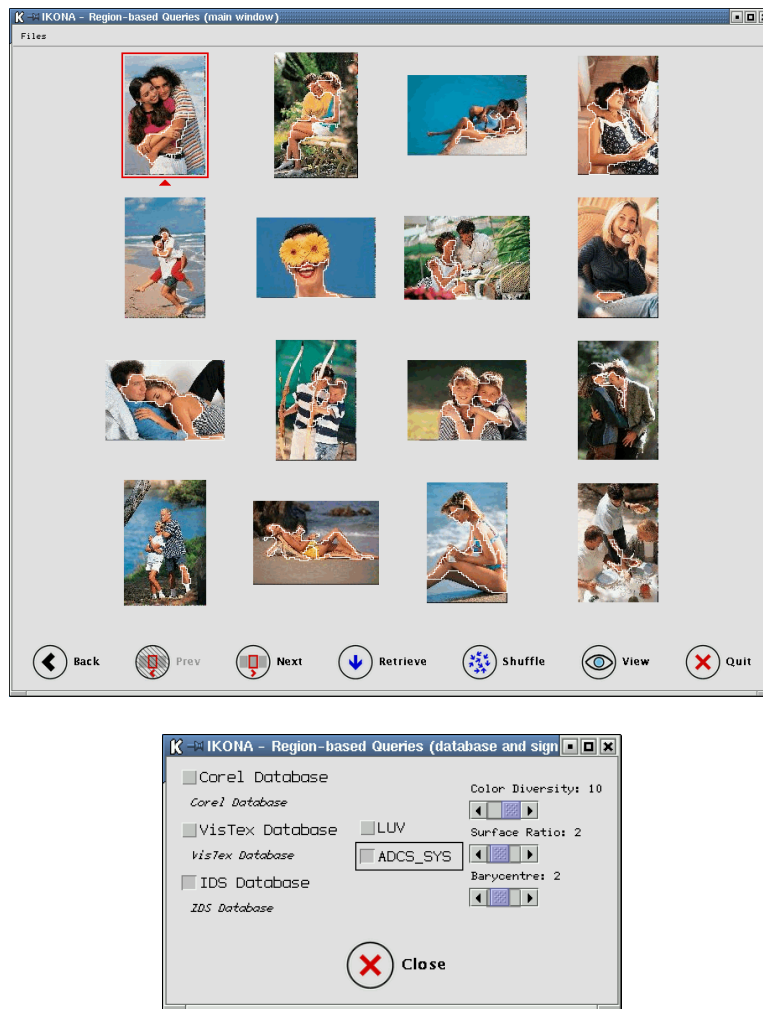


Figure 11. L'interface utilisateur Ikona pour la recherche de régions. Chaque région dans chaque vignette est cliquable dans la fenêtre principale. Les régions retrouvées dans les images sont identifiables par leurs contours blancs. La seconde fenêtre de paramétrage permet l'ajustement dynamique de l'importance relative de la diversité couleur, de la position et du barycentre

7. Résultats

7.1. Détection de régions

Même dans les scènes complexes, les régions extraites présentent une cohérence visuelle et sont généralement intuitives pour l'utilisateur. La segmentation grossière prouve sa capacité à intégrer au sein des régions des zones formées de différentes nuances d'une même teinte, de fortes textures, de détails spatiaux isolés. Une telle variabilité perceptuelle rend chaque région plus spécifique vis-à-vis des autres régions de la base. Les régions supprimées (petites et grisées dans les exemples de la figure 12) représentent une très faible proportion des surfaces d'images. Les images difficiles à segmenter sont celles qui comportent majoritairement des détails fins et saillants colorimétriquement. Pour de telles images, il pourra être intéressant d'appliquer une description partielle par points d'intérêt.

Des exemples d'images segmentées sont présentés dans la figure 12. Afin de juger perceptuellement du comportement de la technique de segmentation sur de nombreux exemples de la base, le lecteur est invité à se rendre à l'adresse suivante : <http://www-rocq.inria.fr/~fauqueur/ADCS/>

Le processus de segmentation est rapide (une moyenne de 1,9 seconde par image) ce qui est convenable pour de grandes bases d'images. 56 374 régions ont été automatiquement extraites des 11 479 images. On obtient donc une moyenne de 5,2 régions par image (calculée sur les bases Corel et IDS mais pas Vistex dont chaque image est réduite à une région).

nombre d'images	11 479
nombre de régions	56 374
nombre de régions par image (hors vistex)	5,2
temps de segmentation par image	1,9 s

7.2. Description de régions

Dans la figure 12, la troisième image de chaque exemple est l'image créée à partir des nuances de couleur utilisées pour indexer chaque région. La forte similarité visuelle entre ces images avec leur image originale montre la précision de la description de variabilité couleur ADCS.

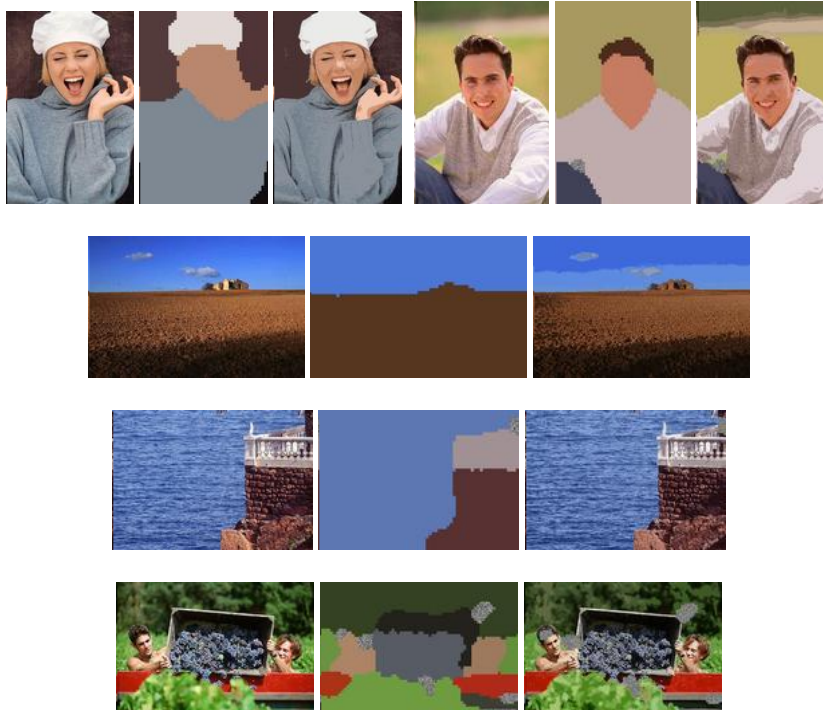
Un total de 963 215 couleurs dans l'espace couleur LUV a été automatiquement déterminé pour indexer les 56 374 régions ce qui donne une moyenne de 17 nuances de couleurs par région. Parmi ces couleurs, 690 419 sont uniques, nombre à comparer à la palette de 200 couleurs fixées dans les descriptions couleur usuelles.

L'extraction des index ADCS pour les régions d'une image prend environ 0,8 s. La taille de stockage d'un index est de 69 octets (un scalaire est stocké comme un octet).

Celui-ci s'avère trois fois plus compact qu'un histogramme couleur classique.

nombre d'images	11 479
nombre de régions	56 374
nombre total de nuances de couleurs	963 215
nombre total de nuances uniques	690 419
nombre de couleurs par région	17
temps d'indexation par image	0,8 s

Les descripteurs couleur proposés dans la littérature reposent sur une représentation fixe et grossière des couleurs commune à toute la base, qu'il s'agisse de description de région ou d'image. Il en est de même pour les variantes de l'histogramme couleur, proposées pour la description *globale* d'images. Leur amélioration porte principalement sur l'intégration d'information de structure locale (tels les « geometric histograms » [RAO 00] ou les « weighted histograms » [VER 00]). Ces représentations de la couleur étant équivalentes à celle de l'histogramme couleur classique, nous avons comparé ADCS à ce dernier. Nous avons utilisé un histogramme couleur reposant sur une quantification systématique de l'espace LUV en 6 cellules par composante, donnant une palette de 216 couleurs. Pour la recherche de similarité, la comparaison d'histogramme est effectuée avec la distance L^1 .



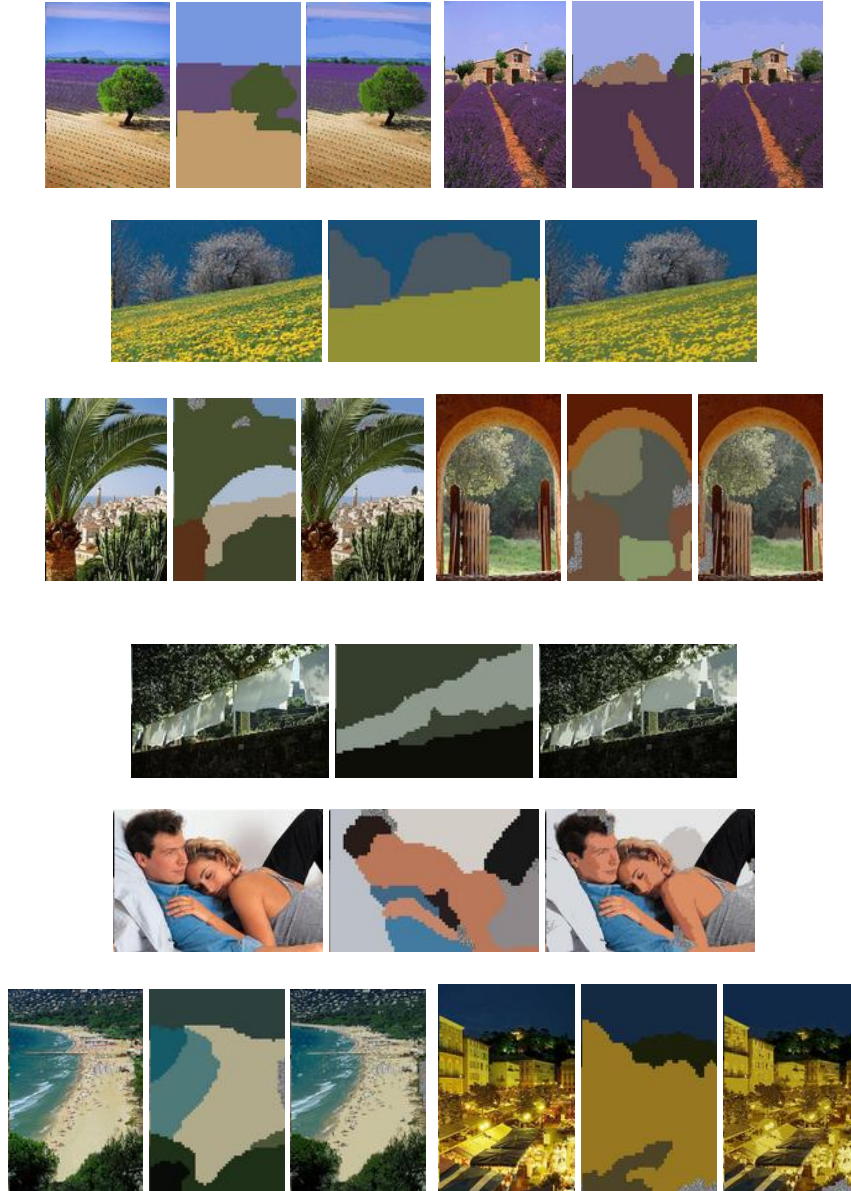


Figure 12. Illustration de la segmentation grossière et de la description fine. Chaque triplet d'images comporte l'image originale, l'image des régions détectées représentées par leur couleur moyenne et l'image des régions avec leurs nuances couleur utilisées pour leur indexation. Les petites régions supprimées sont grisées. La forte similarité visuelle entre chaque image originale et l'image des nuances de couleur montre la précision du descripteur ADCS

7.3. Recherche

7.3.1. Evaluation qualitative

A l'aide de notre système, des centaines de requêtes ont toujours retourné des régions présentant des distributions de couleurs perceptuellement proches pour différents types de région-requête : uniformes ou texturées ou avec différentes nuances de la même teinte. Les régions retournées donnent une impression de continuum visuel au fil des rangs. La prise en compte de la position et de la surface des régions (c'est-à-dire en attribuant des valeurs non nulles aux poids dans la fenêtre de paramètres d'Ikona) améliore presque systématiquement la pertinence des résultats tout en accélérant les temps de requête grâce à la stratégie de rejet de régions décrite en section 4.3.

Lors de la comparaison des résultats obtenus avec l'histogramme couleur classique et avec le descripteur ADCS, l'amélioration de la similarité perceptuelle est remarquable parmi les régions retournées. La meilleure fidélité de description des régions participe à la réduction du « fossé sémantique ». Selon l'apparence visuelle de la région requête, les résultats peuvent être plus ou moins pertinents. Par exemple, une requête sur une petite région noire n'a pas de sens dans l'absolu puisqu'il peut s'agir d'une partie d'objets très différents mais aussi d'ombre. Inversement, des « concepts » peuvent présenter des apparences très différentes ; par exemple, les concepts « chien », « vêtement », « voiture » peuvent avoir des teintes et des textures opposées. Ci-dessous nous avons relevé des exemples d'objets pour lesquels les requêtes régions ont montré une forte corrélation avec leur apparence visuelle :

taille	position dans l'image	teinte	variabilité couleur	« objet » probable
grand	en bas	blanc	faible	neige
grand	en bas	violet	forte	champ de lavande
grand	en haut	bleu	dégradé	ciel
<i>non discriminant</i>	<i>non discriminant</i>	cyan	faible	piscine
petit à moyen	centre	rose clair	dégradé	peau

Il est fréquent que l'apparence visuelle d'une région, aussi cohérente visuellement soit-elle, ne soit pas caractéristique d'une unique classe d'objets dans une base hétérogène. La « valeur sémantique » des régions détectées dépend donc souvent du scénario de recherche et de l'intérêt de l'utilisateur.

Ces considérations sur la « valeur sémantique » des régions détectées nous ont aidés à établir la base vérité-terrain de régions, afin de produire une évaluation quantitative de la précision de notre système.

7.3.2. Evaluation quantitative

L'évaluation quantitative de la précision de recherche de notre système a été testée en considérant chaque région de la base vérité-terrain comme région requête parmi les classes suivantes :

- 88 classes de Vistex (792 images et 792 régions vérité-terrain) ;
- classe « lavande » (108 images et 134 régions vérité-terrain) ;
- classe « personnes » (371 images et 634 régions vérité-terrain) ;
- classe « piscine » (26 images et 29 régions vérité-terrain).

Chacune de ces régions a été utilisée comme région requête. Parmi les premières régions retournées (analysées jusqu'au rang 50) la précision au rang k est mesurée comme le quotient par k du nombre de bonnes régions retournées jusqu'au rang k . La figure 13 montre les courbes de précision obtenues par requêtes automatiques de chacune de ces régions avec les descripteurs ADCS, histogramme classique et la combinaison de ADCS avec les descripteurs géométriques proposés dans la section 4.2. Par souci de clarté, la précision sur la base Vistex est présentée sur la moyenne des 88 classes.

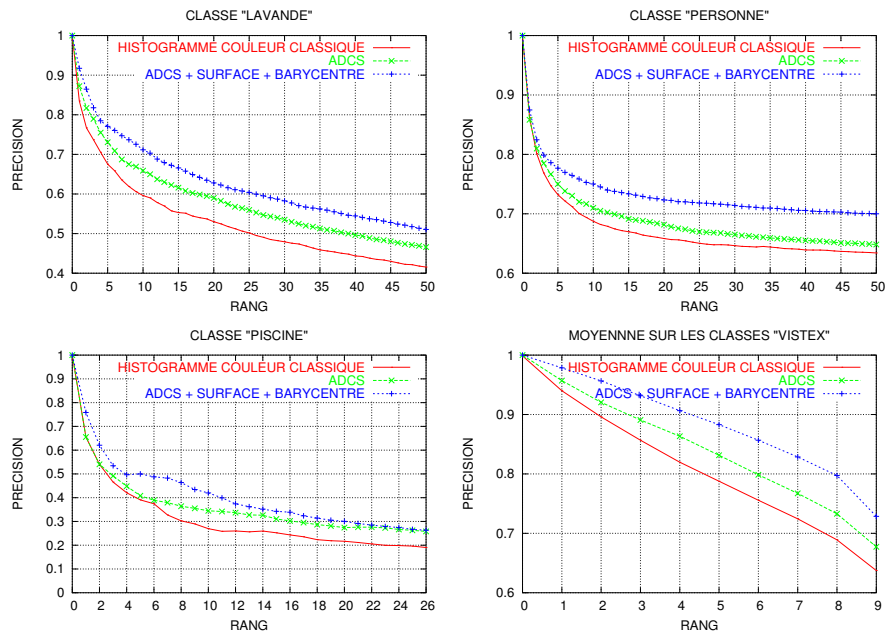


Figure 13. Courbes de précision sur les classes « lavande », « personne », « piscine » et « Vistex » en utilisant les 3 modes de recherche : histogramme classique contre ADCS et contre ADCS combiné avec la surface et le barycentre

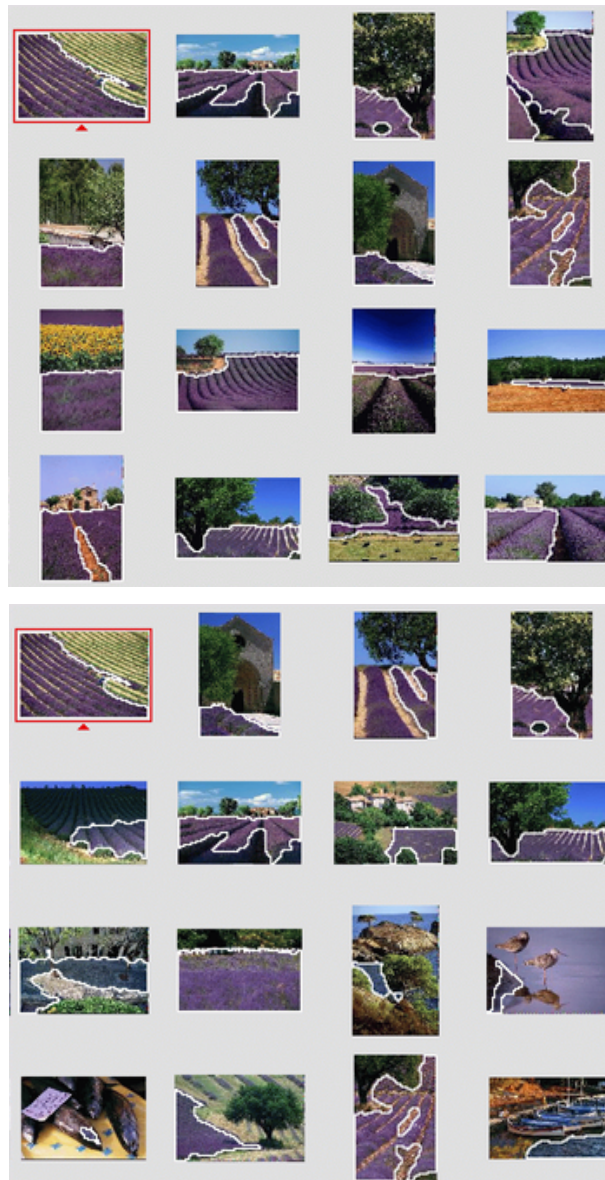


Figure 14. Résultats de recherche à partir de la région de lavande en haut à gauche : avec ADCS (écran supérieur) et avec l'histogramme couleur classique à 216 cellules (écran inférieur). L'histogramme classique ne peut pas distinguer des nuances de mauve de nuances de bleu (ils se retrouvent dans la même cellule). Aucun descripteur géométrique n'est utilisé ici

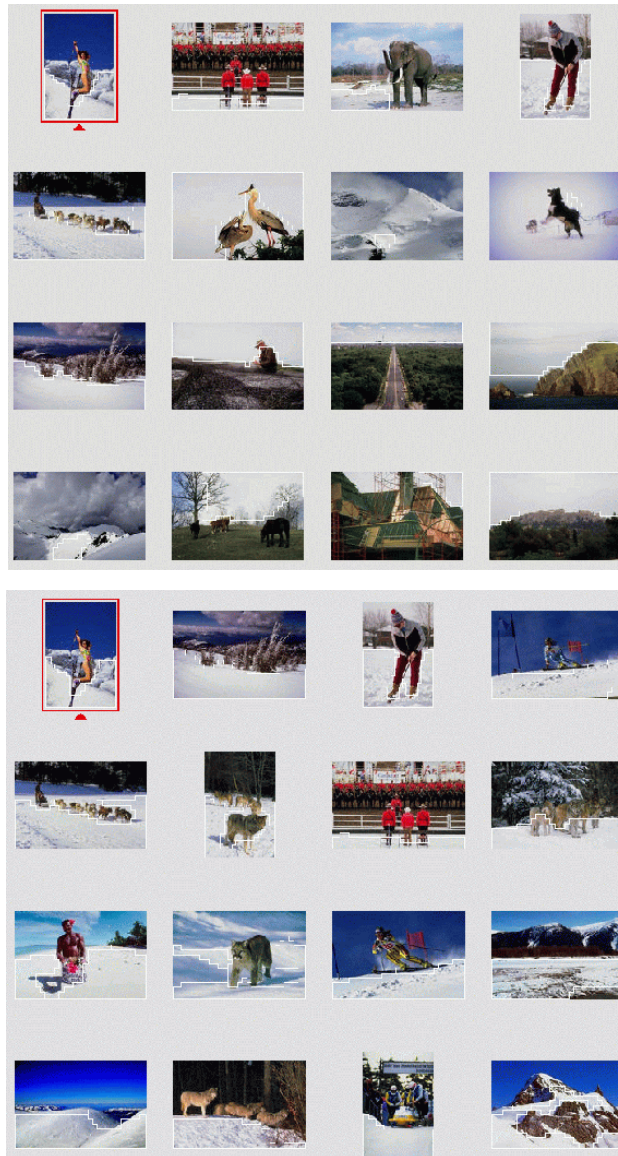


Figure 15. Résultats de recherche à partir de la région de neige en haut à gauche : avec ADCS (écran supérieur) et la combinaison de ADCS avec la surface et le barycentre (écran inférieur). Bien que les régions retournées avec le descripteur ADCS soient pertinentes en termes de photométrie, la taille et la position améliorent nettement la qualité des résultats. Les régions de neige sont un exemple de région requête pour laquelle les descripteurs géométriques sont discriminants

Pour toutes les classes, ADCS améliore la précision avec un gain positif par rapport à l'histogramme classique. Ce gain est variable selon les classes considérées. Les régions décrites par plusieurs nuances de couleurs ont été mises en correspondance avec des régions comportant plusieurs nuances de couleurs et *idem* pour les régions comportant une (ou très peu de) couleur. Nous avons observé que le nombre de nuances de couleur est aussi une information exploitée dans les descripteurs ADCS. Les faux positifs d'images retournées avec l'histogramme classique étaient dus à une imprécision dans la similarité lorsque celui-ci ne permettait pas de distinguer deux régions perceptuellement différentes correspondant à des objets différents (voir capture d'écran 14). Nous en concluons que le gain en précision des régions retournées par rapport aux descripteurs couleur existants est cohérent avec le gain en fidélité de description avec le nouveau processus de description et de mise en correspondance.

La combinaison de ADCS avec des descripteurs géométriques simples a conduit à une seconde amélioration significative, à l'exception de la classe piscine pour laquelle le gain est quasiment nul dans les derniers rangs. Dans notre base vérité-terrain, les régions associées à une piscine étaient soit de petites parties du fond de l'image, soit le fond entier, soit une tâche en haut ou en bas de l'image. Cette remarque est cohérente avec le fait que les descripteurs géométriques ne sont pas discriminants pour cette classe dans notre base. Cependant, pour des régions telles que *lavande*, *ciel*, *peau* ou *neige* (voir les exemples de requêtes à la figure 15), ils sont très discriminants.

Les index de régions sont comparés de manière exhaustive avec la région requête. Le temps moyen de recherche parmi les 56 374 régions est de 0,8 s avec leur descripteur ADCS seul et de 0,5 s avec ADCS et la surface et le barycentre (nous rappelons que la stratégie de rejet accélère les comparaisons).

7.3.3. Passage à l'échelle et optimisation

Les recherches de régions s'opèrent par comparaison exhaustive avec toutes les régions de la base. La base n'est pas prestructurée. Nous avons vu que jusqu'à 11 479 (donnant 56 374 régions) le temps de réponse est faible : 0,5 seconde. En effet, dans le domaine des bases de données on estime que jusqu'à 2 secondes, l'utilisateur n'a pas le sentiment d'attendre le résultat de sa requête. Dans ces conditions, nous estimons que le temps de recherche devient critique à partir de 40 000 images.

La manipulation de plus grandes bases nécessite l'accélération du processus de recherche. La première solution serait d'optimiser la calcul de la distance quadratique en utilisant la distance minorante rapide à calculer proposée dans [HAF 95]. La seconde solution consiste à réduire l'espace de recherche en prestructurant la base à l'aide, par exemple, de catégories de régions. L'approche de catégorisation de régions proposée dans [FAU 03a] (dont le but est d'effectuer des recherches d'images par composition de régions) pourrait être exploitée.

8. Conclusions et perspectives

Grâce au paradigme de recherche d'images par régions d'intérêt, l'utilisateur peut explicitement désigner les zones visuelles pertinentes pour sa requête. Nous avons présenté une nouvelle approche de *segmentation grossière et description fine de régions*.

La nouvelle méthode de segmentation est rapide et repose sur la classification d'une unique primitive locale riche (LDQC) avec l'algorithme CA dont l'avantage majeur est de déterminer automatiquement le nombre de classes. Les régions ainsi détectées constituent des clés de requête visuellement pertinentes : d'une part, car elles englobent une variabilité photométrique qui les rend caractéristiques les unes des autres dans la base et, d'autre part, car elles sont intuitives pour l'utilisateur.

Nous avons ensuite introduit un nouveau descripteur de variabilité couleur des régions (ADCS). Il offre une caractérisation fine, compacte et adaptative de l'apparence visuelle des régions. Combinée à une mesure de similarité adéquate et performante (la forme généralisée de la distance quadratique), nous avons montré que la fidélité de description améliorerait la précision des recherches de régions par rapport aux descripteurs existants. De plus, une seconde amélioration significative des résultats a été observée en combinant ADCS avec des descripteurs géométriques élémentaires. De cette approche résulte un système de recherche d'images plus intuitif.

Différents axes futurs de recherche sont à envisager. L'approche de caractérisation fine de la couleur dans ADCS peut être appliquée à des descripteurs qui intègrent, en plus de la couleur, l'information de répartition spatiale de la couleur. Par ailleurs, nous avons proposé des pistes pour accélérer le processus de recherche dans le but de traiter de plus grandes bases (40 000 images et plus). Finalement, nous envisageons de permettre à l'utilisateur de formuler des requêtes qui tiennent compte des relations spatiales à la fois directionnelles et topologiques entre les régions.

9. Bibliographie

- [BEZ 81] BEZDEK J. C., *Pattern Recognition with Fuzzy Objective Functions*, Plenum, New York NY, 1981.
- [BOU 00] BOUJEMAA N., « On Competitive Unsupervised Clustering », *International Conference on Pattern Recognition (ICPR'00)*, 2000.
- [BOU 01] BOUJEMAA N., FAUQUEUR J., FERECATU M., FLEURET F., GOUET V., SAUX B. L., SAHBI H., « IKONA : Interactive Generic and Specific Image Retrieval », *International workshop on Multimedia Content-Based Indexing and Retrieval (MMCBIR'2001)*, Rocquencourt, France, 2001.
- [BRU 00] BRUNELLI R., MICH O., « COMPASS : an Image Retrieval System for Distributed Databases », *IEEE International Conference on Multimedia and Expo (ICME'00)*, New York City, 2000.
- [CAR 99] CARSON C. *et al.*, « Blobworld : A system for region-based image indexing and retrieval », *Proc. of International Conference on Visual Information System*, 1999.

- [CAR 02] CARSON C., BELONGIE S., GREENSPAN H., MALIK J., « Blobworld : Image segmentation using Expectation-Maximization and its application to image querying », *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)* 24(8) : 1026-1038, 2002.
- [DEL 98] DEL BIMBO A., VICARIO E., « Using Weighted Spatial Relationships in Retrieval by Visual Contents », *IEEE workshop on Image and Video Libraries*, 1998.
- [DEN 99] DENG Y., MANJUNATH B. S., « An efficient low-dimensional color indexing scheme for region based image retrieval », *Proc. IEEE Intl. Conference on Acoustics, Speech and Signal Processing (ICASSP'99)*, Phoenix, Arizona, 1999.
- [FAU 02] FAUQUEUR J., BOUJEMAA N., « Region-based Retrieval : Coarse Segmentation with Fine Signature », *IEEE International Conference on Image Processing (ICIP)*, 2002.
- [FAU 03a] FAUQUEUR J., BOUJEMAA N., « New image retrieval paradigm : logical composition of region categories », *IEEE International Conference on Image Processing (ICIP)*, 2003.
- [FAU 03b] FAUQUEUR J., BOUJEMAA N., « Region-Based Image Retrieval : Fast Coarse Segmentation and Fine Color Description », *Journal of Visual Languages and Computing (JVLC), special issue on Visual Information Systems*, 2003 à paraître.
- [FAU 03c] FAUQUEUR J., Contributions pour la Recherche d'Images par Composantes Visuelles, Thèse de doctorat, INRIA, 2003.
- [FLI 95] FLICKNER M. *et al.*, « Query by image and video content : the QBic system », *IEEE Computer*, vol. 28, n° 9, 1995, p. 23-32.
- [FRI 97] FRIGUI H., KRISHNAPURAM R., « Clustering by competitive agglomeration », *Pattern Recognition*, vol. 30, n° 7, 1997, p. 1109-1119.
- [GEV 99] GEVERS T., SMEULDERS A., « The PicToSeek WWW Image Search System », *ICMCS*, Vol. 1, 1999, p. 264-269.
- [GOU 01] GOUET V., BOUJEMAA N., « Object-based queries using color points of interest », *IEEE Workshop on Content-Based Access of Image and Video Libraries (CBAIVL'01)*, 2001.
- [GUP 96] GUPTA A. *et al.*, « The Virage image search engine : an open framework for image management », *SPIE Storage and Retrieval for Image and Video Databases*, vol. 2670, 1996.
- [HAF 95] HAFNER J., SAWHNEY H. *et al.*, « Efficient Color Histogram Indexing for Quadratic Form Distance Functions », *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 17, n° 7, 1995, p. 729-736.
- [MA 97] MA W. Y., MANJUNATH B., « EdgeFlow : A framework of boundary detection and image segmentation », *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'97)*, 1997, p. 744-749.
- [MA 99] MA W. Y., MANJUNATH B. S., « NeTra : A Toolbox for Navigating Large Image Databases », *Multimedia Systems*, vol. 7, n° 3, 1999, p. 184-198.
- [MAL 99] MALKI J., BOUJEMAA N., NASTAR C., WINTER A., « Region Queries without Segmentation for Image Retrieval by Content », *Proc. of International Conference on Visual Information System (VIS'99)*, 1999, p. 115-122.
- [MAN 02] MANJUNATH B., SALEMBIER P., SIKORA T., *Introduction to MPEG-7 : Multimedia Content Description Interface*, Wiley, 2002.

- [MOG 99] MOGHADDAM B., BIERMANN H., MARGARITIS D., « Defining image content with multiple regions of interest », *IEEE Workshop on Content-Based Access of Image and Video Libraries (CBAIVL'99)*, 1999.
- [PAV 77] PAVLIDIS T., *Structural Pattern Recognition*, Springer-Verlag, Berlin, 1977.
- [PEN 94] PENTLAND A., PICARD R., SCLAROFF S., « Photobook : Content-based Manipulation of Image Databases », *SPIE Storage and Retrieval for Image and Video Databases*, vol. II, n° 2185, 1994.
- [RAO 00] RAO A., SRIHARI R., ZHANG Z., « Geometric histogram : A distribution of geometric configurations of color subsets », *SPIE : Internet Imaging*, 3964, 2000.
- [ROB 88] ROBERTSON P. K., « Visualizing Color Gamuts : A User Interface for the Effective Use of Perceptual Color Spaces in Data Displays », *IEEE Computer Graphics and Applications*, 1988, p. 50-64.
- [SAN 99] SANGWINE S., HORNE R., *The colour image processing handbook*, Chapman and Hall, 1999.
- [SCH 98] SCHMID C., MOHR R., BAUCKHAGE C., « Comparing and Evaluating Interest Points », *International Conference on Computer Vision (ICCV)*, 1998.
- [SCL 97] SCLAROFF S., TAYCHER L., CASCIA M. L., « ImageRover : A Content-Based Image Browser for the World Wide Web », *IEEE Workshop on Content-based Access of Image and Video Libraries (CBAIVL'97)*, 1997.
- [SMI 96] SMITH J. R., CHANG S. F., « VisualSEEK : A Fully Automated Content-Based Image Query System », *ACM Multimedia Conference, Boston, MA, USA*, 1996.
- [SRI 02] SRIDHAR V., NASCIMENTO M. A., LI X., « Region-based Image Retrieval Using Multiple-Features », *Proc. of International Conference on Visual Information System (VIS'02)*, Hsin-Chu, Taiwan, 2002.
- [STE 02] STEHLING R. O., NASCIMENTO M. A., FALCAO A. X., « MiCRoM : A Metric Distance to Compare Segmented Images », *Proc. of International Conference on Visual Information System (VIS'02)*, Hsin-Chu, Taiwan, 2002.
- [VER 00] VERTAN C., BOUJEMAA N., « Upgrading Color Distributions for Image Retrieval : can we do better? », *Proc. of International Conference on Visual Information System (VIS'00)*, Lyon, France, 2000.
- [WAN 01] WANG J. Z., LI J., WIEDERHOLD G., « SIMPLcity : Semantics-sensitive Integrated Matching for Picture Libraries », *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2001.

Julien Fauqueur a reçu le diplôme d'ingénieur de l'Ecole Nationale Supérieure des Télécommunications de Paris et le diplôme du DEA d'Intelligence Artificielle, Reconnaissance des Formes et Applications en 1999. Au sein de l'équipe IMEDIA à l'INRIA (Rocquencourt), il a obtenu en 2003 son doctorat en informatique sur la recherche d'images par composantes visuelles. Ses thèmes de recherche portent sur la reconnaissance des formes, la recherche d'information par le contenu visuel, l'interaction utilisateur.

Nozha Boujemaa est directeur de recherche à l'INRIA (Rocquencourt), responsable scientifique du projet IMEDIA. Ses domaines d'intérêt sont l'analyse d'image, la reconnaissance des formes, la recherche d'information dans les archives multimedia.

Elle a obtenu son doctorat en informatique en 1993 et son habilitation à diriger des recherches en informatique en 2000. Auteur de plus de cinquante publications, elle a été membre de plusieurs comités de programmes de conférences internationales (WWW Multimedia, ACM Multimedia, ICPR, IEEE Fuzzy systems, ICME...) dans le domaine de la recherche d'information visuelle et la reconnaissance des formes. Elle a effectué des missions d'expert scientifique auprès de nombreuses institutions dont la National Science Foundation et l'Union européenne.

Article reçu le 13 juin 2002

Version révisée le 26 mars 2003

Rédacteur responsable : Frédéric Jurie