

Recherche d'images par composition logique de catégories de régions

Julien Fauqueur et Nozha Boujemaa

INRIA Rocquencourt, BP 105, 78153 Le Chesnay Cedex - France
{Julien.Fauqueur,Nozha.Boujemaa}@inria.fr
[http ://www-rocq.inria.fr/imedia/](http://www-rocq.inria.fr/imedia/)

Résumé Nous présentons un nouveau paradigme de recherche d'images par composition logique de catégories de régions. Il permet d'atteindre les images, quand elles existent dans la base, qui se rapprochent de la représentation mentale de la cible visuelle de l'utilisateur. Ainsi **aucune image ou région exemple n'est nécessaire** au moment de la formulation de la requête.

Ce paradigme repose sur la génération non-supervisée d'un thesaurus photométrique constitué par le regroupement des catégories de régions similaires de la base. Pour formuler sa requête, l'utilisateur interagit avec ce thesaurus en disposant d'opérateurs de composition logique appliqués aux différents types de régions disponibles. Cette approche permet de retrouver des images en spécifiant les types de régions qui les composent et permet également d'effectuer des requêtes par la négative. Les requêtes logiques sur le contenu des images s'apparentent à celles en recherche de texte.

Des tests sur 9995 images de la photothèque Corel sont présentés. L'originalité de ce paradigme ouvre des perspectives riches pour de futurs travaux en recherche d'information visuelle.

Mots clés thesaurus de régions, résumé visuel, requêtes partielles, image mentale, régions d'intérêt, composition logique de régions, interaction utilisateur

1 Introduction

En recherche d'images par le contenu, le paradigme largement répandu est celui de la recherche par l'exemple visuel. A partir d'une image, le système (QBIC [5], par exemple) retrouve les images d'apparences visuelles similaires. Plus proche des besoins de l'utilisateur, la recherche par région exemple (BLOBWORLD[3], NETRA[7], IKONA[4]) permet de désigner explicitement les parties d'images pertinentes pour la requête.

En recherche d'information visuelle, ce paradigme est contraignant puisqu'il suppose l'existence d'un exemple de départ (image(s) ou région(s)). Or il est rare que l'utilisateur dispose d'un exemple pour formuler une requête visuelle. La recherche préalable d'un exemple pour effectuer une recherche s'avère fastidieuse, surtout pour une requête à partir de plusieurs régions. Généralement l'image cible qu'il recherche n'existe que dans son esprit.

Le nouveau paradigme présenté dans cet article diffère radicalement des approches existantes aussi bien du point de vue de la formulation de la requête que du processus de recherche. A l'issue de la segmentation des images de la base (par la méthode proposée dans [4]), les régions extraites sont regroupées en catégories de régions visuellement similaires. Dans l'interface de requête, l'ensemble des régions représentatives de l'ensemble des catégories fournit un résumé visuel des différentes des régions qui composent les images de la base. Ce résumé constitue un thesaurus photométrique de régions. A l'aide de ce thesaurus et d'opérateurs logiques, l'utilisateur peut retrouver les images à partir de requêtes telles que : "retrouver les images composées de régions de ce type et de ce type, mais pas de ce type". Pour plus de détails sur cette approche, nous invitons le lecteur à se référer à [2].

2 Catégorisation de régions pour la recherche de similarité

Après extraction des régions d'images de la base, nous souhaitons définir les catégories de régions comme les regroupements de régions visuellement similaires. Elles définiront les "types" de régions à partir desquels l'utilisateur pourra formuler sa requête de composition.

Le regroupement des régions est obtenu par classification non-supervisée de leurs descripteurs visuels. Notre choix s'est porté sur l'algorithme CA d'agglomération compétitive [6] en raison de son avantage majeur de déterminer *automatiquement* le nombre de classes. Il consiste à minimiser la fonction objectif J suivante :

$$J = \sum_{i=1}^C \sum_{j=1}^N u_{ij}^2 d^2(x_j, \beta_i) - \alpha \sum_{i=1}^C \left[\sum_{j=1}^N u_{ij} \right]^2$$

où $\{x_j, \forall j = 1, \dots, N\}$ désignent les descripteurs de régions à regrouper, $\{\beta_i, \forall i = 1, \dots, C\}$ les prototypes de classes à déterminer, $d(x_j, \beta_i)$ la distance entre un descripteur x_j et un prototype β_i , et u_{ij} le degré d'appartenance floue d'un descripteur x_j à une classe i . En minimisant J , donc les deux termes antagonistes qui la composent, l'algorithme génère une partition optimale des descripteurs et supprime les classes à faible population.

Dans cette étape de regroupement, tout type de descripteur de région peut être employé (couleur, texture, forme, géométrie). Dans nos tests, nous avons utilisé comme descripteur de régions la couleur moyenne. Bien que simple, ce dernier permet de former des catégories génériques de régions.

Les **catégories de régions**, notées $C_1, \dots, C_P$, sont les classes obtenues à l'issue de la classification par CA. Elles constituent la base de la similarité entre les régions au moment de la recherche d'images. L'existence de groupements naturels de régions dans l'espace de description ne pouvant pas être exhibée dans le cas général, la classification est effectuée avec une granularité *fine* qui permet d'obtenir des catégories cohérentes de régions. Au moment de la recherche, la similarité de régions étant définie par rapport à leurs appartenances aux catégories, la cohérence des catégories pourra garantir de retrouver les régions très similaires. Afin de permettre la recherche de régions à des degrés *variables* de similarité, nous définirons aussi comme similaires les régions se trouvant dans des catégories proches dans l'espace de description, appelées "catégories voisines". La **catégorie voisine** d'une catégorie C_q de prototype p_q est définie comme la catégorie C_j dont le prototype p_j satisfait $\|p_q - p_j\|_{L^2} \leq \gamma$, pour un rayon de recherche donné γ . Le rayon γ sera fixé au moment de la requête. Nous notons $V^\gamma(C_q)$ l'ensemble des catégories voisines d'une catégorie C_q . La figure 1 illustre l'influence du choix du rayon pour les range-queries sur les catégories voisines.

L'idée-clé est de permettre d'effectuer des "range-queries" sur les catégories cohérentes dans l'espace de description des régions grâce à la prise en compte des catégories voisines.

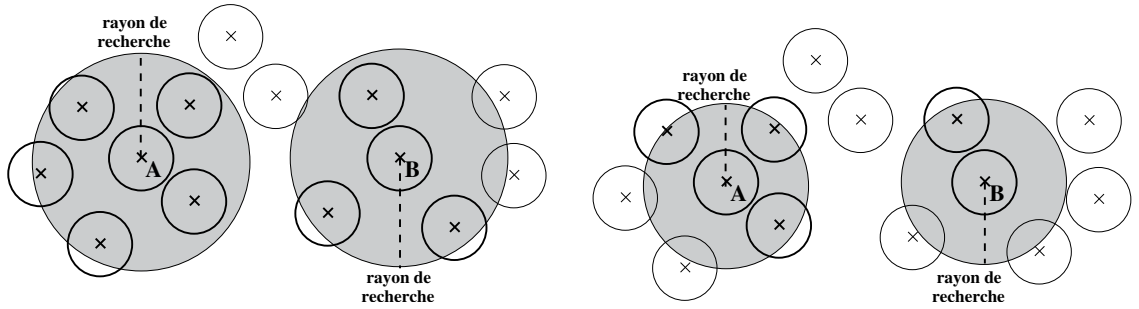


Fig. 1. Rayon de recherche et sélection des catégories voisines. A et B désignent deux catégories requête. Le choix du rayon de recherche détermine l'ensemble des catégories voisines de A et B. Un rayon faible (à gauche) et un rayon élevé (à droite) vont couvrir plus ou moins de catégories voisines pour définir les types de régions recherchées. Les disques grisés couvrent les catégories voisines. Les catégories voisines ont des contours plus épais que les autres catégories. Les prototypes sont repérés par des croix.

3 Recherche d'images par composition logique de catégories de régions

Les catégories de régions que nous venons de définir permettent à l'utilisateur de désigner les types de régions qui composent l'image mentale cible. Dans l'interface de requête, l'utilisateur sélectionne les catégories, appelées Catégories Requête Positives (**CRP**), qui définissent les types de régions qui doivent composer les images qu'il recherche. Il peut aussi indiquer les types de régions qui *ne doivent pas* les composer en sélectionnant des catégories, appelées Catégories Requête Négatives (**CRN**). On note $\{C_{pq_1}, \dots, C_{pq_M}\}$ et $\{C_{nq_1}, \dots, C_{nq_R}\}$ les ensembles de CRP et de CRN sélectionnées par l'utilisateur pour une requête donnée. Dans sa formulation la plus complexe, la requête par composition s'exprime ainsi : "trouver les images composées de régions de ces différents types (les CRP), mais pas de ces différents types (les CRN)".

Nous allons exposer le principe de résolution de cette requête par étapes en considérant une catégorie requête CRP, puis ses voisines, puis plusieurs CRP et finalement les CRN. Dans un premier temps, nous recherchons les images composées d'une région appartenant à une seule catégorie CRP que nous notons C_{pq} . Pour une catégorie C_{pq} donnée, nous définissons $IC(C_{pq})$ comme l'ensemble des images composées d'au moins une région appartenant à C_{pq} . Pour étendre la recherche par range-query, nous prenons aussi en compte les catégories voisines de C_{pq} . Ainsi nous définissons comme pertinentes, les images composées d'une région appartenant à une catégorie voisine de C_{pq} :

$$S(C_{pq}) = \bigcup_{C \in N^\gamma(C_{pq})} IC(C) \quad (1)$$

Ensuite, pour l'extension de la recherche aux M CRP $C_{pq_1}, \dots, C_{pq_M}$, nous devons déterminer les images composées d'une région dans C_{pq_1} ou ses voisines et ... et d'une région dans C_{pq_M} ou ses voisines. L'ensemble S_Q d'images satisfaisant cette requête multiple s'écrit à présent :

$$S_Q = \bigcap_{i=1}^M S(C_{pq_i}) \quad (2)$$

Finalement, afin de satisfaire aussi la requête par la négative, nous devons déterminer les images composées d'une région dans les R CRN : $C_{nq_1}, \dots, C_{nq_R}$. Comme précédemment, les catégories voisines sont également prises en compte. L'ensemble S_{NQ} des images composées des CRN s'écrit ainsi :

$$S_{NQ} = \bigcap_{i=1}^R S(C_{nq_i}) \quad (3)$$

L'ensemble S_{res} des images pertinentes qui ont des régions dans les différentes CRP et pas dans les CRN s'exprime comme la soustraction ensembliste de S_Q avec S_{NQ} :

$$S_{res} = S_Q \setminus S_{NQ} \quad (4)$$

L'ensemble S_{res} constitue l'ensemble final des images pertinentes. Les unions, intersections et soustractions dans l'expression de S_{res} sont directement équivalentes à la formulation de la requête par des opérateurs logiques comme l'illustre la figure 3 : OU entre les voisines (expression (1)), ET entre les différentes catégories requête (expression (2)), ET NON pour les catégories requête négatives (expression (4)).

Afin d'évaluer l'expression (4) de S_{res} , l'approche "naïve" consisterait à tester, pour chaque image de la base, si elle est composée d'une région appartenant à une CRP (ou ses voisines) mais d'aucune région dans les CRN (ni leurs voisines). Notre solution pour réduire considérablement ce nombre de tests de façon simple est d'exploiter le fait que S_{res} s'exprime comme des intersections et soustractions ensemblistes. L'idée est d'initialiser S_{res} avec l'un des ensembles d'images puis de supprimer les images qui n'appartiennent pas aux autres ensembles d'images. Cette initialisation évite de tester individuellement chaque image de la base et permet de démarrer directement avec un ensemble d'images potentiellement pertinentes. S_{res} est graduellement réduit de la façon suivante :

1. initialiser S_{res} comme l'ensemble $S(C_{pq_1})$.
2. supprimer les images de S_{res} qui n'appartiennent à aucun des ensembles $S(C_{pq_i}), i = 2, \dots, M$ afin d'obtenir l'expression (2) de S_Q . A ce point, nous avons $S_{res} = S_Q$.
3. pour soustraire S_{NQ} de S_{res} , supprimer dans S_{res} les images qui appartiennent à aux ensembles $S(C_{nq_i}), i = 2, \dots, R$. Nous avons finalement : $S_{res} = S_Q \setminus S_{NQ}$ (expression (4)).

Grâce à cette approche, nous verrons dans la section suivante que nous évitons l'accès à une fraction importante de la base.

La mise en oeuvre effective de cette approche repose sur l'utilisation de trois tables qui fournissent les associations entre les catégories, les catégories voisines et les images. Le processus de recherche est très rapide car il n'implique que des opérations élémentaires sur les entiers, contrairement aux approches usuelles qui nécessitent des calculs coûteux de distances entre des vecteurs de primitives multidimensionnels.

4 Résultats et interface utilisateur

La base test est constituée de 9.995 images de la photothèque Corel dont ont été extraites automatiquement 50.220 régions. Le regroupement par CA des 50.220 vecteurs de couleurs moyennes a automatiquement généré 91 catégories. Au sein des différentes catégories, on observe que les régions présentent des couleurs moyennes très similaires. La variabilité photométrique intra-catégorie est principalement due aux régions ayant des textures différentes.

L'interface de requête présente à l'utilisateur les 91 régions représentatives¹ des catégories qui produisent un résumé visuel des types de régions disponibles dans la base (figure ??). Ce résumé visuel constitue le **thesaurus photométrique des régions** à partir duquel l'utilisateur peut sélectionner chaque catégorie comme CRP ou CRN pour formuler sa requête. L'étendue de similarité des types de régions recherchées est contrôlée par l'utilisateur en fixant, dans l'interface, la valeur du rayon de recherche γ .



Fig. 2. Le thesaurus des régions constitue l'interface de requête. L'utilisateur peut sélectionner l'une des 91 catégories comme CRP (case verte) ou CRN (case rouge). **Fig. 4.** Images retournées par la requête par composition

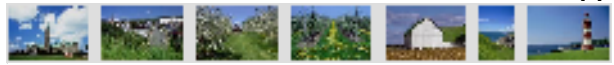
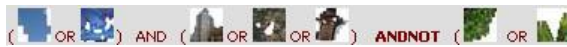


Fig. 3. Le système traduit la requête de l'utilisateur (ici, deux CRP et une CRN) comme expression logique de à cause de la présence d'une zone verte. **Fig. 5.** Images rejetées par la requête de paysages urbains

composition de catégories de régions. Les images retournées satisfont toujours la composition elle-même dont la mise en correspondance est exacte. En effet, la présence ou l'absence de types de régions est représentée par des tables d'association. Dans les images retournées, les régions appartenant aux catégories requête positives sont visuellement saillantes. En observant les images résultat, l'utilisateur peut modifier la formulation de sa requête pour en améliorer la pertinence : en ajoutant ou en supprimant des catégories requête ou en modifiant le rayon de recherche.

Les figures 3, 4 et 5 illustrent une requête par composition logique destinée à retrouver des paysages urbains. En pensant à des scènes composées d'immeubles, de ciel et sans végétation

¹ La région représentative d'une catégorie est définie comme la région la plus proche de son prototype.

par exemple, l'utilisateur formule la requête par composition suivante : "images composées d'une région grise et d'une région bleue mais pas d'une région verte". Dans les images retournées (figure 4) sur la base Corel, les régions grises correspondent à des immeubles, des monuments ou de la roche et les régions bleues principalement à du ciel. Les images retournées par cette approche vérifient effectivement les contraintes de composition de régions. Le système montre aussi les images rejetées à cause de la présence d'une région verte (fig. 5). Il est intéressant de remarquer que, pour cette requête, les images rejetées correspondent à des paysages de nature (fig. 5).

Avec l'approche présentée, les temps de recherche sont très rapides : au plus 0.03 seconde pour les requêtes par composition les plus complexes sur un PC à 498 MHz. Le processus de recherche ne considère pas les 50.220 régions, mais uniquement leurs 91 catégories. Par ailleurs, il n'y a pas de recherche séquentielle sur les images : en moyenne, on accède seulement à une fraction de 12% des 9.995 images.

5 Conclusions et Perspectives

Nous avons introduit le nouveau paradigme de recherche d'images par composition logique de catégories de régions. Il permet de formuler des requêtes telles que : "retrouver des images composées de régions de ces types mais pas de ces types". Il ne nécessite aucune image ou région exemple. Avec l'usage d'opérateurs logiques de requête, l'interrogation du thesaurus photométrique de régions permet de formuler des requêtes directement sur la composition des images cible de façon semblable à la recherche de texte. Les temps de recherche s'avèrent très rapides même pour des compositions complexes. De plus, nous avons pu observer l'émergence d'une forme de sémantique visuelle dans la contrainte de composition de types de régions.

Simple, ouverte et totalement non-supervisée, cette approche se prête à différents travaux futurs tels que l'intégration de différents descripteurs de région, la prise en compte des relations spatiales dans la composition, la catégorisation hiérarchique des régions pour de très grandes bases. De plus cette approche peut aisément se transposer à des types d'attributs d'images autres que les descripteurs de régions et aussi être appliquées à des documents ayant d'une description numérique ses composantes (par exemple une video et ses scènes).

Références

1. <http://www-rocq.inria.fr/fauqueur/COMPOCAT/>
2. J. Fauqueur. Contributions pour la Recherche d'Images par Composantes Visuelles. *Thèse de doctorat, INRIA*, 2003.
3. C. Carson and al. Blobworld : A system for region-based image indexing and retrieval. *Proc. of International Conference on Visual Information System*, 1999.
4. J. Fauqueur and N. Boujemaa. Region-based retrieval : Coarse segmentation with fine signature. *IEEE International Conference on Image Processing (ICIP)*, 2002.
5. M. Flickner and al. Query by image and video content : the QBic system. *IEEE Computer*, 28(9),1995.
6. H. Frigui and R. Krishnapuram. Clustering by competitive agglomeration. *Pattern Recognition*, 30(7), 1997.
7. W. Y. Ma and B. S. Manjunath. Netra : A toolbox for navigating large image databases. *Multimedia Systems*, 7(3) :184–198, 1999.