

# Modèle topologique pour l'interrogation des bases d'images

Mihaela Scuturici\*, Jérémy Clech\*  
Vasile-Marian Scuturici\*, Djamel A. Zighed\*

\*5, avenue Pierre Mendès-France - 69676 Bron cedex - France  
{mscuturi,jclech,vscuturi,dzighed}@eric.univ-lyon2.fr  
<http://eric.univ-lyon2.fr>

**Résumé.** Nous proposons dans cet article un modèle topologique de représentation de bases d'images. Chaque image est représentée à l'aide d'un vecteur de caractéristiques dans  $\mathbb{R}^p$  et figure comme nœud dans un graphe de voisinage. L'exploration du graphe correspond à la navigation dans la base de données, les voisins d'un nœud représentent des images similaires. Afin de pouvoir traiter des requêtes, nous définissons un modèle topologique. L'image requête est représentée par un vecteur de caractéristiques dans  $\mathbb{R}^p$  et insérée dans le graphe en mettant à jour localement les relations de voisinage. Ce travail se positionne dans le domaine de la fouille de données complexes.

## 1 Introduction

La recherche d'information dans les bases d'images reste un défi. Beaucoup de systèmes permettant la recherche d'images par le contenu existent : QBIC, Photobook [Pentland *et al.*, 1996], NeTra [Ma *et al.*, 1999], RETIN [Fournier *et al.*, 2001], etc. Dans ce cadre, l'utilisateur fournit une image exemple ou bien indique des critères de recherche comme les couleurs, les formes ou les régions (sous la forme d'un croquis). Mais il arrive fréquemment que l'utilisateur soit surpris par le résultat retourné par la requête. En effet, une image a un contenu sémantique propre déterminé par son contexte. Or ces éléments sont rarement capturés par les caractéristiques de bas niveau. Dès lors, la similarité en termes de caractéristiques de bas niveau ne correspond pas toujours à la similarité sémantique désirée par l'utilisateur.

Explorer la similarité peut se traduire par la recherche des voisins d'une image dans une base d'images. En effet, à partir des caractéristiques préalablement extraites, chaque image est représentée comme un point dans cet espace multidimensionnel. La plupart des algorithmes de recherche dans des bases d'images proposent de retourner les  $k$  plus proches voisins d'une image (kPPV) [Mitchell, 1997] au sens d'une mesure de similarité [Veltkamp *et al.*, 2000]. Par exemple, le système QBIC [Faloutsos *et al.*, 1994], dans son implémentation pour le musée de l'Hermitage (<http://www.hermitagemuseum.org/>) renvoie les 12 images les plus proches de la requête soumise par l'utilisateur. Dans la section 2, nous présentons un autre modèle de voisinage : le voisinage topologique. La section 3 introduit le modèle topologique de requête et présente ses avantages et ses limites. La section 4 montre les résultats expérimentaux sur une base d'images où les deux méthodes de requête (kPPV et voisinage topologique) sont comparées à travers deux indicateurs : le rappel et la précision. Enfin, les conclusions et perspectives sont présentées dans la section 5.

## 2 Voisinage topologique

Soit un ensemble  $\Omega$  composé de  $n$  images. Chaque image est représentée par un vecteur de caractéristiques de dimension  $p$ , qui est un point dans l'espace  $\mathbb{R}^p$  :

$$X(i) = (X_1(i), \dots, X_j(i), \dots, X_p(i)) \in \mathbb{R}^p, 1 \leq i \leq n.$$

La *topologie* d'un ensemble de données  $\Omega$  précise comment les points (les données) sont reliés entre eux. Dans notre cas, chaque image est représentée comme un nœud dans un graphe de voisinage. L'existence d'une arête entre deux nœuds signifie qu'ils vérifient une relation de similarité particulière. Nous avons choisi un modèle de graphe avec un critère local de construction, ce qui exclu les graphes comme par exemple l'Arbre Minimum Recouvrant. Le graphe des Voisins Relatifs ou le graphe de Gabriel satisfont ce critère. Pour plus de détails sur les modèles de graphe et leurs propriétés voir [Preparata *et al.*, 1985].

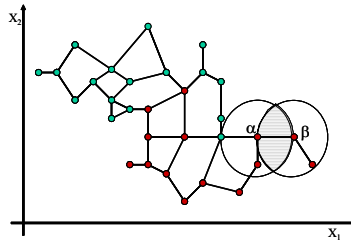


FIG. 1 Exemple de Graphe des Voisins Relatifs dans  $\mathbb{R}^2$

**Le Graphe des Voisins Relatifs** [Toussaint, 1980] que nous avons utilisé dans nos expérimentations est un graphe connexe où deux points  $\alpha$  et  $\beta$  sont voisins s'ils vérifient la condition suivante : « La lunule, correspondant à la surface hachurée dans la Figure 1, doit être vide ». En d'autres termes, deux points  $\alpha$  et  $\beta$  sont reliés par une arête s'ils vérifient l'équation (1) :

$$d(\alpha, \beta) \leq \text{Max}(d(\alpha, \gamma), d(\beta, \gamma)), \forall \gamma \in \Omega \setminus \gamma \neq \alpha, \beta \quad (1)$$

Dans la recherche d'images par le contenu, nous allons nous baser sur la relation de similarité (voisinage) induite par le modèle topologique du Graphe des Voisins Relatifs. Pour chacun des points (image) nous pouvons alors connaître les voisins. Bien sûr, tous ses voisins ne sont pas équidistants, on peut alors prendre cette information pour restreindre ou ordonner les voisins d'un point. Pour  $p \geq 3$  la représentation visuelle peut être projetée dans l'espace  $\mathbb{R}^2$  en utilisant, par exemple, l'analyse en composantes principales ou les arbres phylogénétiques [Hillis *et al.*, 1993]. Cette représentation peut être une alternative aux cartes de Kohonen [Kohonen, 1995].

## 3 Interrogation topologique

En considérant une image requête  $I$  et une base d'images, un utilisateur recherche un ensemble d'images lui étant similaires. En appliquant l'algorithme des kPPV, le système renvoie les  $k$  images les plus proches de  $I$ , en termes de distances.

Fixer la valeur de  $k$  représente un inconvénient de l'approche kPPV. En effet, il n'y a aucune raison de supposer que le nombre de voisins d'un point doit être fixe. Il n'est donc pas naturel de forcer une image à avoir  $k$  voisins si elle en comporte moins, et inversement de la restreindre à  $k$  voisins si elle en comporte davantage.

Nous illustrerons le processus d'interrogation sur un exemple simple. La Figure 2 présente huit images (a) appartenant à trois catégories sémantiques: « plantes », « fontaine » et « montagne ». ainsi que leur matrice des distances (b).

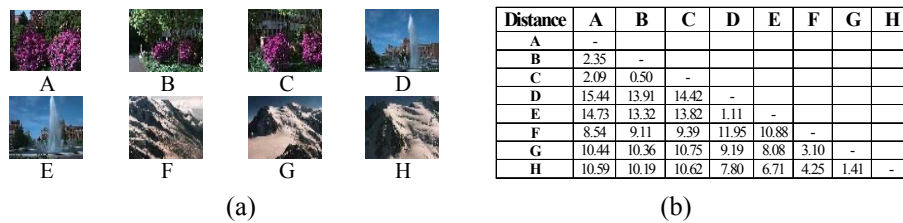


FIG. 2 Images appartenant à trois catégories sémantiques : plantes (A, B et C), fontaine (D et E) et montagne (F, G et H) (a) et la matrice des distances en utilisant 15 caractéristiques de couleur et texture projetées sur les deux premières composantes principales (b)

Si l'utilisateur présente comme image requête une image similaire à A (« plantes »), le système devrait renvoyer A, B et C. Dans le cas de kPPV, si la valeur de  $k$  est fixée à 2 seulement deux résultats seront obtenus, disons A et B. Une autre requête est nécessaire pour obtenir l'image de C, même si elle est très similaire aux deux autres. D'autre part, si  $k$  est fixé à 3 et l'image requête est similaire à D (« fontaine »), le kPPV renverra certes D et E mais également une autre image dont la distance est la plus proche par rapport à l'image requête. La matrice des distances dans notre exemple montre que H est le deuxième voisin le plus proche de l'image D, donc l'utilisateur obtiendra une image de « montagne », visuellement non similaire avec l'image requête représentant une fontaine.

Quand l'utilisateur soumet une image requête  $I$ , il cherche les images les plus semblables dans la base d'images. Il serait souhaitable de calculer le graphe de voisinage contenant toutes les images dans la base de données plus celle fournie en tant que requête. En raison de la complexité d'ordre  $O(n^3)$  de la construction du graphe, cette opération n'est pas efficace.

Dans la pratique, l'image requête  $I$  est localement insérée dans le graphe de voisinage. Des modifications locales sont exécutées sur les relations de voisinage. Cette approche justifie notre choix du modèle de graphe des voisins relatifs, qui possède un critère local de construction. L'insertion locale se déroule en deux étapes, la première consistant à déterminer l'ensemble des points candidats aux modifications du voisinage locale et la seconde consistant à mettre à jour les relations de voisinage concernées par les points candidats.

Le graphe des voisins relatifs est construit dans la Figure 3 (a) en utilisant des caractéristiques de couleur et de texture. Nous pouvons facilement observer les trois amas qui correspondent aux trois catégories sémantiques. La Figure 3 (b) présente un exemple simple de requête topologique. La nouvelle image est insérée (localement) dans le graphe de voisinage et l'utilisateur peut parcourir tous ses voisins. Une nouvelle image de montagne est utilisée comme image requête (elle n'appartient pas à la base de données, mais elle est visuellement très similaire avec les autres images de montagne dans la base d'images).

## Modèle topologique pour l'interrogation des bases d'images

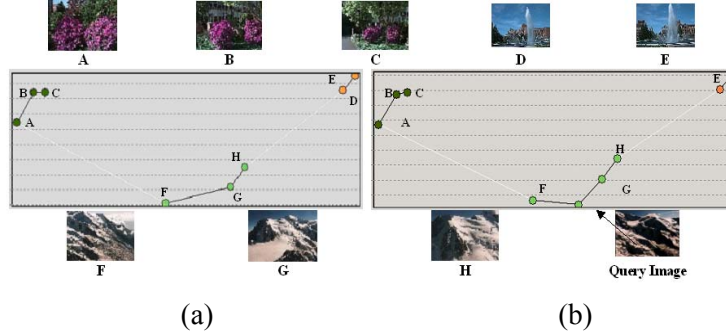


FIG. 3 La projection du graphe des voisins relatifs. La couleur des nœuds représente la catégorie sémantique. A gauche (a), sans l'insertion de l'image requête ; à droite avec (b)

Cette méthode a plusieurs avantages : premièrement, le graphe de voisinage est localement mis à jour au lieu d'être entièrement reconstruit. L'image requête est seulement insérée dans le graphe de voisinage et les relations entre la nouvelle image et ses voisins sont créés ou actualisés. Un autre avantage de la méthode consiste dans le fait que l'utilisateur peut parcourir le graphe de voisinage et analyser tous les voisins de l'image requête, donc le graphe peut être utilisé comme support de navigation entre des images similaires.

## 4 Expérimentation

Nous avons utilisé un jeu de données de 259 images subdivisé en six catégories principales. Ces images ont été extraites de la base de données Ground Truth [Washington, 1999]. Les catégories sont les suivantes : « Arbogreens », « Australia », « Cherries », « SwissMoutains », « Greenlake » et « SpringFlowers ». Nous avons extrait à partir de ces images deux familles de caractéristiques : des caractéristiques de couleurs (L1 et L2 normalisées ; la couleur prédominante) et 14 caractéristiques de textures ([Haralick *et al.*, 1973]).

### 4.1 Protocole de comparaison

Dans la recherche d'information, les performances sont habituellement mesurées avec les deux notions classiques : le rappel ( $\rho$ ) et la précision ( $\pi$ ) [Swets, 1963]. Pour chaque image requête  $I$  le système retrouve un ensemble d'images que nous notons  $R(I)$ . Il contient des images pertinentes et non-pertinentes. Nous notons  $R^{per}(I)$  l'ensemble d'images pertinentes dans  $R(I)$  et  $\Omega^{per}(I)$  l'ensemble d'images pertinentes dans l'ensemble total de données  $\Omega$ . Le rappel  $\rho(I)$  est défini comme la proportion de documents pertinents retournés par le système comparé à la totalité des documents pertinents existants dans l'ensemble de données. La précision  $\pi(I)$  est définie comme la proportion de documents pertinents parmi les documents retournés par le système.

$$\rho(I) = \frac{|R^{per}(I)|}{|\Omega^{per}(I)|}; \quad \pi(I) = \frac{|R^{per}(I)|}{|R(I)|} \quad (2)$$

Afin de comparer les résultats de notre système de recherche d'information basé sur le voisinage à un système basé sur les  $k$  plus proches voisins, nous avons comparé leurs taux de

rappel et précision dans un contexte de catégorisation : déterminer la catégorie sémantique des images. Lorsque nous soumettons une requête au système, ce dernier retourne un ensemble de  $k$  réponses pour le kPPV et un ensemble de voisins pour le graphe de voisinage. Les réponses fournies par les deux méthodes sont pondérées par l'inverse de leur distance à la requête et constituent un vote à la majorité permettant ainsi de définir la catégorie de la requête. Dans cette étude, nous avons utilisé la distance Cosinus puisqu'elle est invariante aux échelles des axes de l'espace de représentation [Rajman *et al.*, 1998]. Nous avons fait varier  $k$  de 1 à 5 et les résultats de la prédiction ont été évalués après une 10 validation-croisée.

## 4.2 Résultats et discussion

La Figure 4 montre les résultats obtenus lors de nos expérimentations. Nous observons que le Graphe des Voisins Relatifs surpasse les différents modèles de plus proches voisins. Nous avons effectué les expériences jusqu'à  $k=30$ , mais ils étaient toujours inférieurs à ceux du graphe des voisins relatifs. Nous expliquons ces meilleures performances par l'utilisation d'un nombre de voisins dépendant de la disposition des données dans l'espace des caractéristiques qui doit être fortement bénéfique au graphe des voisins relatifs au niveau des points frontières alors que les kPPV utilisent un nombre constant de points quelque soit la disposition des données.

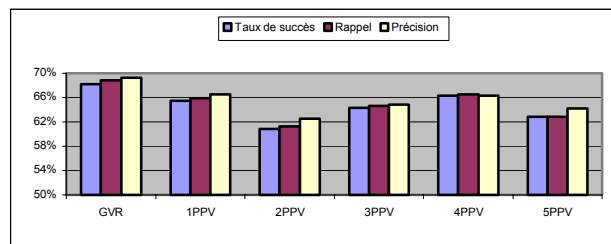


FIG. 4 Comparaison des performances sur la collection d'images extraites de la base Ground Truth entre le Graphe des voisins relatifs (GVR) et le kPPV ( $k$  variant de 1 à 5)

## 5 Conclusions et perspectives

Dans cet article nous avons proposé un modèle topologique pour la représentation de base de d'images et pour les requêtes. Notre idée est de construire un graphe de voisinage sur les caractéristiques de bas niveau automatiquement extraites à partir d'images. La navigation dans une base d'images consiste à explorer les nœuds du graphe de voisinage dans l'espace  $\mathbb{R}^p$ . Nous avons aussi proposé un modèle d'interrogation topologique. La requête est représentée comme un vecteur dans  $\mathbb{R}^p$ . Ensuite le graphe de voisinage est localement mis à jour afin d'établir les relations de voisinage pour l'image requête et ses voisins. Ces voisins sont retournés comme résultat de la requête. Le modèle de recherche par le contenu proposé ne se distingue pas par son mode d'indexation (vectorisation des images à partir de caractéristiques bas-niveau), mais par son mode d'appariement et son mode de navigation. D'un autre côté, la norme MPEG 7 est graduellement mise en place et des modèles de requête pourraient être établis sur des descripteurs MPEG 7 des données multimédias. L'extraction automatique des descripteurs sémantiques de contenu audiovisuel reste toujours un problème, aussi bien que

leur exploitation convenable. Le modèle topologique proposé dans cet article offre la possibilité d'une analyse exploratoire des descripteurs MPEG 7 et permet leur utilisation dans un processus d'interrogation. Nous travaillons actuellement sur les mesures de qualité de l'espace de représentation. Cette mesure permettra d'apprécier la cohérence de l'annotation.

## Références

- [Faloutsos, C. et al., 1994] Faloutsos, C. et al. Efficient and Effective Querying by Image Content, *Journal of Intelligent Information Systems* 3(3):231-162, 1994.
- [Fournier, J. et al., 2001] Fournier, J. et al. Retin: A content based image indexing and retrieval system, *Pattern Analysis and Applications Journal* 4(2/3):153-173, 2001.
- [Haralick, R. M. et al., 1973] Haralick, R. M. et al. Texture features for image classification, *IEEE Transactions Systems, Man and Cybernetics* 3:610-621, 1973.
- [Hillis, D. M. et al., 1993] Hillis, D. M. et al. Analysis of DNA sequence data: phylogenetic inference, *Methods Enzymol* 224:456-487, 1993.
- [Kohonen, T., 1995] Kohonen, T. *Self-Organizing Maps*. Berlin/Heidelberg, Springer, 1995.
- [Ma, W.-Y. et Manjunath, B. S., 1999] Ma, W.-Y. et Manjunath, B. S. NeTra: A toolbox for navigating large image databases, *ACM Multimedia Systems* 7(3):184-198, 1999.
- [Mitchell, T. M., 1997] Mitchell, T. M. *Machine Learning*. New York, McGraw-Hill, 1997.
- [Pentland, A. et al., 1996] Pentland, A. et al. Photobook: Content-based manipulation of image databases, *Journal of Computer Vision* 18(3):233-254, 1996.
- [Preparata, F. et Shamos, M. I., 1985] Preparata, F. et Shamos, M. I. *Computational Geometry. An introduction*. New York, Springer-Verlag, 1985.
- [Rajman, M. et Lebart, L., 1998] Rajman, M. et Lebart, L. (1998). Similarités pour données textuelles. *4th International Conference on Statistical Analysis of Textual Data (JADT'98)*, pp. 545-555, Nice, France, 1998.
- [Swets, J. A., 1963] Swets, J. A. Information retrieval systems, *Science* 141(3577):245-250, 1963.
- [Toussaint, G. T., 1980] Toussaint, G. T. The Relative Neighbourhood Graph of a Finite Planar Set, *Pattern Recognition* 12(4):261-268, 1980.
- [Veltkamp, R. C. et Tanase, M., 2000] Veltkamp, R. C. et Tanase, M. *Content-based image retrieval systems : A survey.*, Utrecht University, 2000.
- [Washington, 1999] Washington. *Ground Truth Database by the University of Washington*. 2003, 1999.

## Summary

In this paper we propose a topological model for image database query using neighborhood graphs. A related neighborhood graph is built from automatically extracted low-level features, which represent images as points of  $\mathbb{R}^p$  space. Graph exploration corresponds to database browsing, the neighbors of a node represent similar images. In order to perform query by example, we define a topological query model. The query image is inserted in the graph by locally updating the neighborhood graph. The topology of an image database is more informative than a similarity measure usually applied in content based image retrieval, as proved by our experiments.