

# Stratégie d’exploration en recherche d’images par similarité

Matthieu CORD, Jérôme FOURNIER, Sylvie PHILIPP-FOLIGUET

ETIS (CNRS UMR-8051)ENSEA  
6, av. du Ponceau 95014 Cergy-Pontoise  
cord@ensea.fr, philipp@ensea.fr

**Résumé** – Le problème abordé est celui de la recherche de sous-ensembles d’images (ensembles ayant un thème commun) dans des bases d’images généralistes. Les images recherchées pouvant avoir des contenus très variés, nous proposons de prendre explicitement en compte la complexité de la recherche via la définition d’une requête multiple et d’une fonction de similarité adaptée. De plus, nous introduisons une stratégie d’exploration-exploitation de la base d’images afin de pouvoir traquer des ensembles d’images de grande diversité. La technique proposée repose sur un algorithme stochastique permettant de gérer l’exploration. C’est un outil simple qui présente des performances accrues par rapport à celles des systèmes classiques de recherche par similarité. Des tests comparatifs ont été effectués sur la base généraliste de photographies *Corel* pour valider notre approche.

**Abstract** – This paper concerns the retrieval of image categories in generalist databases. We propose to explicitly take into account the possible multimodality of the categories through the definition of a multiple request and of a similarity function. A strategy of exploration of the database is also proposed which aims at tracking sets of images of high diversity. It is based on a stochastic algorithm able to manage the exploration. It is a simple tool, which improves the performances of classical search-by-similarity systems. Comparative tests on the generalist picture database *Corel* have been carried out in order to validate our approach.

## 1 Introduction

Les techniques traditionnelles de recherche d’images dans de grandes bases de données restent très peu performantes lorsqu’elles opèrent sans interaction avec l’utilisateur. De nombreuses stratégies ont été développées récemment pour améliorer la recherche en impliquant l’utilisateur [6] [4] [7].

Les méthodes de recherche par similarité utilisent principalement la technique dite de bouclage de pertinence pour raffiner la requête et/ou la similarité à partir des données annotées par l’utilisateur [5]. Ces méthodes sont relativement efficaces lorsqu’on cherche une image précise dans la base mais échouent le plus souvent lorsque l’on cherche des ensembles d’images assez diversifiés (catégories d’images ayant une sémantique commune). En effet, dans ce contexte, les images pertinentes ne sont généralement plus regroupées en un seul mode dans les espaces de représentation. Plusieurs modes sont donc nécessaires pour représenter correctement la catégorie recherchée, ce qui n’est pas géré par les méthodes classiques.

Nous proposons dans ce papier une stratégie de recherche par similarité gérant explicitement la multi-modalité des images pertinentes (images de la catégorie recherchée). Toutes les images pertinentes collectées pendant la recherche grâce aux annotations utilisateur (pertinente, non pertinente) forment un ensemble noté  $\mathcal{R}$  qui est directement utilisé comme requête multiple. Il faut modifier en conséquence la fonction de similarité pour calculer non plus une distance entre deux images mais entre une image et l’ensemble  $\mathcal{R}$ . On propose pour cela un schéma de fusion codant au mieux le comportement recherché.

L’autre point crucial concerne l’exploration de la base. En effet, si les images pertinentes peuvent être réparties en plusieurs modes éventuellement éloignés dans les espaces de représentation, il faut pouvoir retrouver ces différents modes. Or, en dé-

but de recherche, la requête se limite généralement à une seule image. Sans stratégie spécifique, les méthodes de recherche par similarité vont retrouver des images proches de l’image initiale, situées dans le même mode, mais plus difficilement identifier d’autres modes.

Nous proposons une stratégie permettant une exploration de la base au cours des premières itérations de la recherche, suivie d’une phase d’exploitation plus importante de l’information accumulée via  $\mathcal{R}$ .

A l’instar de Vendrig et al. [8], l’idée est de définir un espace de recherche suffisamment large au départ (pour pouvoir traquer plusieurs modes), puis de réduire cet espace de recherche au fil des itérations. Pour mener à bien ce schéma d’exploration, notre algorithme est basé sur une optimisation stochastique.

## 2 Requête multiple et similarité

La forme globale de la similarité que nous proposons entre une image cible  $\mathbf{I}$  de la base et la requête multiple  $\mathcal{R} = \{R_i, 1 \leq i \leq L\}$  combine linéairement les similarités calculées dans  $M$  espaces de représentation des images<sup>1</sup> :

$$S(\mathcal{R}, \mathbf{I}) = \sum_{k=1}^M \beta^{(k)} s^{(k)}(\mathcal{R}, \mathbf{I}) \quad (1)$$

où  $s^{(k)}(\mathcal{R}, \mathbf{I})$  est la similarité dans le  $k$ ’ième espace de représentation et  $\beta^{(k)} \in \mathbb{R}^+$ .

Pour prendre correctement en compte la multi-modalité, nous considérons que chaque image  $R_i$  ( $R_i \in \mathcal{R}$ ) doit être utilisée

<sup>1</sup>. Voir [7] pour une discussion sur les espaces de représentation et les index associés.

comme une source d'information propre dans le calcul de la similarité globale. Contrairement aux méthodes de recherche par similarité traditionnelles, le calcul d'un nouveau vecteur requête unique est exclu. Nous lui préférons un schéma de fusion des similarités image par image car le comportement de l'opérateur de fusion peut être précisé. Dans notre contexte, la multimodalité est prise en compte correctement si les contraintes suivantes sont vérifiées :

- Si  $\mathbf{I}$  est proche d'un des exemples  $\mathbf{R}_i$ , la similarité  $s^{(k)}(\mathbf{R}_i, \mathbf{I})$  doit être favorisée par rapport aux autres similarités  $s^{(k)}(\mathbf{R}_{j \neq i}, \mathbf{I})$ .
- Si la cible est à égale distance de toutes les images de  $\mathcal{R}$ , une moyenne est adaptée.

Nous avons choisi en conséquence un opérateur de fusion barycentrique :

$$s^{(k)}(\mathcal{R}, \mathbf{I}) = \frac{\sum_{i=1}^L s^{(k)}(\mathbf{R}_i, \mathbf{I})^2}{\sum_{i=1}^L s^{(k)}(\mathbf{R}_i, \mathbf{I})} \quad (2)$$

Différentes mesures de similarité entre images  $s^{(k)}(\mathbf{R}_i, \mathbf{I})$  ont été testées et une similarité basée sur la distance  $L_1$  entre index normalisés a été adoptée :  $s^{(k)}(\mathbf{R}_i, \mathbf{I}) = 1 - d_{L_1}(\mathbf{R}_i, \mathbf{I})$ .

L'opérateur barycentre est adapté pour gérer automatiquement le type de configurations décrites précédemment. Il s'adapte à la différence de distribution des images pertinentes (multimodalité, dispersion, *etc.*) d'une recherche à l'autre.

De plus, la fonction de similarité est une somme pondérée des similarités requête-cible sur chaque modèle (*cf.* eq. 1). Il est possible d'effectuer un bouclage de pertinence pour affiner les paramètres  $\beta^{(k)}$  de la fonction de similarité. Le réglage des poids peut être réalisé manuellement par l'utilisateur du système [1]. Cependant, une mise à jour semi-automatique, à partir de l'information d'interaction, est également possible. La technique de mise à jour peut reposer sur une *heuristique* [4], ou sur des méthodes *basées-optimisation* [5]. Dans tous les cas, le processus vise à renforcer le poids des attributs les plus discriminants.

Nous avons opté pour la règle d'optimisation LMS car la forme de notre fonction de similarité fournit un calcul simple et explicite des coefficients [2].

### 3 Stratégie d'exploration

Supposons que  $\mathcal{R}$  ne contienne qu'une image  $\mathbf{R}_1$  au début de la recherche. Les dissimilarités calculées sur la base peuvent être classées pour retrouver les images proches de  $\mathcal{R} = \{\mathbf{R}_1\}$ . Cependant, dès lors que la catégorie recherchée contient aussi des images éloignées de  $\mathbf{R}_1$  et du mode qui la contient, cette approche par similarité à une requête simple n'est plus adaptée.

C'est pourquoi nous avons modifié la stratégie de recherche par similarité classique de la façon suivante : nous avons associé à chaque image de la base une probabilité d'être pertinente pour la recherche de catégories en cours. Cette loi de probabilité est paramétrée par un coefficient  $T$  (paramètre de température) gérant explicitement l'exploration de la base. Il permet donc de sélectionner des images relativement éloignées de  $\mathcal{R}$  en début

de recherche, puis de restreindre au fil des itérations l'espace de recherche autour de  $\mathcal{R}$ . Dans le même temps,  $\mathcal{R}$  s'enrichit de toutes les images annotées positivement.

#### 3.1 Loi de probabilité pour l'exploration

La probabilité d'une image  $\mathbf{I}$  d'être pertinente par rapport à la catégorie recherchée est une variable aléatoire qui s'exprime en fonction de la dissimilarité  $D(\mathcal{R}, \mathbf{I}) = 1 - S(\mathcal{R}, \mathbf{I})$  selon :

$$P(\mathbf{I}) = \frac{1}{Z_T} \exp\left(\frac{-D(\mathcal{R}, \mathbf{I})}{T}\right) \quad (3)$$

où  $Z_T$  est la somme des exponentielles de toutes les images cibles et  $T$  le paramètre de température.

A chaque itération, le système tire au sort et affiche des images selon cette loi de probabilité. Toutes les images que l'utilisateur annoté pertinentes viennent compléter l'ensemble  $\mathcal{R}$ . Quand le paramètre  $T$  est grand, l'influence de la dissimilarité à  $\mathcal{R}$  est réduite, et donc le voisinage exploré autour de  $\mathcal{R}$  est étendu ; l'exploration est favorisée. Quand  $T$  décroît, l'influence de la dissimilarité à  $\mathcal{R}$  dans le calcul de la probabilité augmente. L'espace de recherche se resserre autour de  $\mathcal{R}$ . Le contenu de  $\mathcal{R}$  accumulé durant la recherche est pleinement exploité.

#### 3.2 Paramétrage du système

Le nombre d'itérations de la phase d'exploration doit être faible afin de ne pas décourager l'utilisateur. C'est pourquoi nous avons adopté une loi de décroissance exponentielle

$$T_j = C^j \cdot T_0 \quad (4)$$

Les paramètres  $C$  et  $T_0$  déterminent la distribution de probabilité sur la base, ces derniers sont réglés expérimentalement pour que l'exploration dure entre 5 et 10 itérations.

## 4 Expérimentations

Les tests ont été réalisés sur une base généraliste extraite de la galerie commerciale *Corel*. Elle comprend 6000 images catégorisées par le fabricant en 29 classes distinctes (portraits, espace, iceberg, *etc.*).

Les critères de qualité retenus pour l'évaluation sont le *rapport* (rapport du nombre d'images pertinentes retrouvées et du nombre total d'images pertinentes de la base) et la *précision* (rapport du nombre d'images pertinentes retrouvées et du nombre d'images retrouvées).

Toute recherche est initiée par une image requête dont nous précisons, à chaque test, la catégorie d'origine. La connaissance de la vérité-terrain permet de réaliser une évaluation automatique (le système peut déterminer si une image est pertinente ou non) et de fournir des performances fiables, moyennées sur plusieurs requêtes. Le nombre total d'images pertinentes de la base variant suivant la catégorie, les performances ne sont jamais moyennées sur des requêtes appartenant à des classes différentes : l'évaluation se rapporte à une seule catégorie d'images.

L'étape d'indexation est basée sur l'information couleur et texture, ainsi que sur les cooccurrences de couleurs et de textures (voir [3] pour plus de détails).

Nous présentons figure 1 un résultat de recherche pour des requêtes appartenant à la classe "tigre". Une partie des images annotées comme pertinentes durant les 6 premières itérations est affichée.



FIG. 1: Exemples d'images pertinentes accumulées pendant la recherche (image requête initiale en haut).

Les figures 2, 3 et 4 présentent une comparaison entre deux méthodes :

- Notre méthode de recherche, appelée stratégie A. Elle part d'une image requête et comporte l'exploration et la requête multiple, ainsi que le bouclage de pertinence.
- Une stratégie de compétition entre espace d'attributs, appelée stratégie B. Elle part d'une image requête, contient le même bouclage de pertinence que la stratégie A mais ne comporte pas la stratégie d'exploration de la base. A chaque itération, les images sont triées et les images ayant le meilleur score sont affichées.

Sur la fig. 4, nous présentons un résultat sans bouclage (en haut), un résultat avec la stratégie B ( au milieu) et un résultat avec la stratégie A (en bas, images présentées sans ordre). L'utilisateur recherche ici des châteaux et la requête initiale est l'image en haut à gauche. Les bouclages ont été réalisés avec le même nombre d'itérations dans les deux cas. La stratégie B est bien entendu capable de trouver des images pertinentes, mais la stratégie A retrouve des images de châteaux beaucoup plus variés.

Dans la fig. 2, l'évaluation de la qualité est réalisée sur 20 requêtes différentes de la catégorie "fleur". Nous comparons le rappel (en fonction du nombre d'itérations) pour les stratégies A et B et également pour un tirage aléatoire. Avec la stratégie A, le rappel est assez faible durant les premières itérations, mais il croît rapidement après 4 ou 5 bouclages de pertinence. Ceci est dû au fait que pendant les premières itérations, le système explore la base et construit la requête multiple  $\mathcal{R}$ , ce qui accroît sa faculté à retrouver des images de la catégorie, même

éparpillées dans la base, par rapport à une simple compétition entre attributs.

L'efficacité de la stratégie exploration-exploitation est confirmée par les courbes de précision-rappel (cf. fig. 3), tracées pour des paramètres de pondération ( $\beta$ ) obtenues après 10 itérations.

La stratégie A permet de construire, à partir des exemples annotés par l'utilisateur, une requête multiple qui est en quelque sorte un prototype de la catégorie cherchée. Au vu des résultats, cette stratégie se révèle plus efficace après quelques itérations pour retrouver des images d'une catégorie sémantique donnée que la classique recherche-par-similarité.

## 5 Conclusion

Nous avons proposé une nouvelle méthode pour la recherche d'ensembles variés d'images dans des bases généralistes.

La phase de recherche est basée sur un algorithme interactif visant à explorer la base. Notre technique accumule des ensembles d'images les plus représentatifs possibles des catégories recherchées, contenant de nombreuses images de la catégorie avec une grande variabilité sur la base. L'algorithme, de type stochastique, gère l'exploration de la base par diminution progressive de l'espace de recherche des images. Indispensable à ce type de recherche, une requête multiple et une similarité adaptée ont également été introduites.

Les tests réalisés sur la base généraliste *Corel* ont permis de comparer et de valider notre stratégie de recherche d'images. L'ensemble des résultats met en évidence l'importance de la requête multiple et la nécessité d'une stratégie d'exploration dès lors que les recherches sont complexes.

## Références

- [1] C. Carson, M. Thomas, S. Belongie, J. Hellerstein, and J. Malik. Blobworld: A system for region-based image indexing and retrieval. In *Third Int. Conf. on Visual Information Systems*, June 1999.
- [2] J. Fournier, M. Cord, and S. Philipp-Foliguet. Back-propagation algorithm for relevance feedback in image retrieval. In *International Conference in Image Processing (ICIP'01)*, volume 1, pages 686–689, Thessaloniki, Greece, October 2001.
- [3] J. Fournier, M. Cord, and S. Philipp-Foliguet. Retin: A content-based image indexing and retrieval system. *Pattern Analysis and Applications Journal, Special issue on image indexation*, 4(2/3):153–173, 2001.
- [4] A. Heinrichs, D. Koubaroulis, B. Levenaise-Obadia, P. Rovidia, and J. Jolion. Image indexing and content based search using pre-attentive similarities. In *RIA02000*, volume 2, pages 1616–1631, Paris, France, April 2000.
- [5] Y. Rui and T. Huang. Optimizing learning in image retrieval. In *Conf on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 236–243, Hilton Head, SC, June 2000.
- [6] S. Santini, A. Gupta, and R. Jain. Emergent semantics through interaction in image databases. *IEEE Transactions on Knowledge and Data Engineering*, 13(3):337–351, 2001.

- [7] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12):1349–1380, December 2000.
- [8] J. Vendrig, M. Worring, and A. W. M. Smeulders. Filter image browsing, exploiting interaction in retrieval. In D. P. Huijsmans and A. W. M. Smeulders, editors, *Visual Information and Information Systems*, volume 1614 of *Lecture Notes in Computer Science*, pages 147–154, Berlin, 1999. Springer-Verlag.

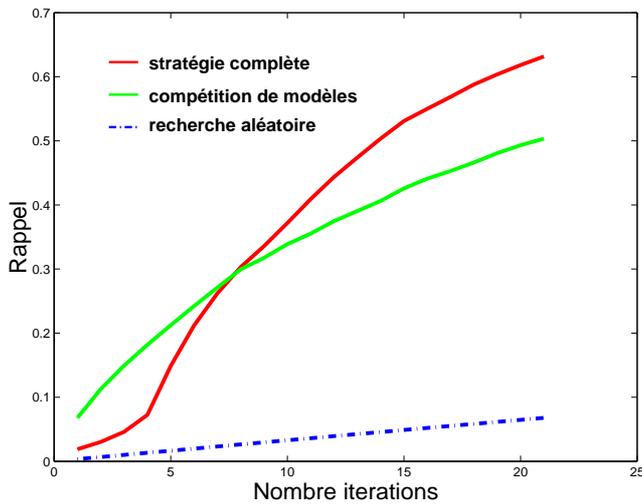


FIG. 2: Performances (base Corel, catégorie "fleur"): rappel en fonction du nombre d'itérations.

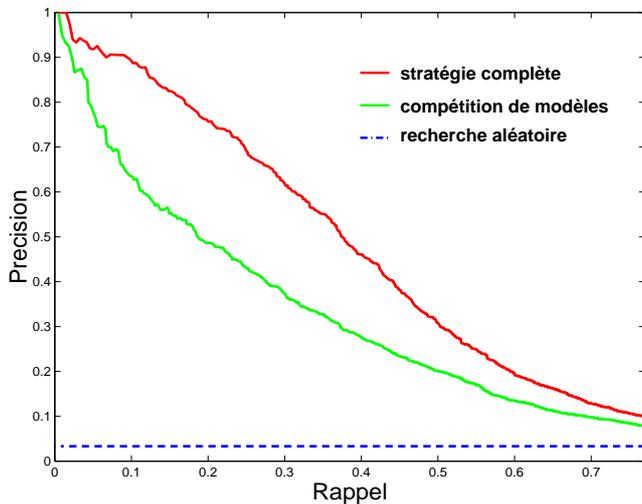


FIG. 3: Performances (base Corel, catégorie "fleur"): courbes de précision-rappel.



FIG. 4: Comparaison des stratégies pour une recherche de châteaux: résultat de la recherche par similarité sans feedback (en haut), résultat de la stratégie B (au milieu), résultat de la stratégie A par exploration-exploitation (en bas).