
Approche interactive de la recherche d'images par le contenu

Une stratégie d'indexation et de recherche basée sur le principe de l'exploration

Matthieu Cord — Jérôme Fournier — Sylvie Philipp-Foliguet

Laboratoire ETIS (CNRS - UMR 8051)
Université de Cergy-Pontoise/E.N.S.E.A
6 avenue du Ponceau – BP 44
F-95014 Cergy-Pontoise cedex
{cord,philipp}@ensea.fr

RÉSUMÉ. Cet article traite du problème de l'indexation et de la recherche d'images par le contenu dans les bases généralistes. Le fossé sémantique qui sépare l'information bas niveau extraite de l'image et la requête sémantique de l'utilisateur est la limite majeure rencontrée dans le domaine. L'aspect indexation est abordé sous l'angle de l'optimisation hors-ligne de la taille et de la pertinence des signatures calculées sur l'image. Nous proposons une stratégie de recherche interactive originale permettant l'exploration de la base. Le processus repose sur la construction d'une requête multiple et sur une compétition de modèles pour le raffinement de la mesure de similarité. Une évaluation des performances est réalisée sur deux bases généralistes afin d'illustrer et de valider notre approche.

ABSTRACT. This paper deals with content-based image indexing and retrieval in general databases. The semantic gap which separates the low-level information extracted from images and the semantic user's request is the main problem encountered in this domain. Indexing is presented under the aspect of "off-line" effectiveness optimization and reduction of image signature length. We introduce an original and interactive retrieval strategy designed for database exploration. The process is based on the construction of a compound query and on a competition between models for the similarity measure refinement. A quality assessment is realized on two general databases in order to illustrate and validate our approach.

MOTS-CLÉS : indexation, recherche-par-similarité, fossé sémantique, interaction utilisateur, bouclage de pertinence, compétition de modèles, exploration, exploitation.

KEYWORDS: indexing, search-by-similarity, semantic gap, user interaction, relevance feedback, competition between models, exploration, exploitation.

1. Introduction

Les avancées technologiques et la démocratisation de l'informatique conduisent à une prolifération de documents numériques à laquelle nous devons faire face aussi bien dans notre vie professionnelle que privée. Des outils de compression, archivage et recherche sont nécessaires afin de gérer ces stocks numériques. En particulier, lorsqu'une base de données est à disposition, son exploitation requiert des méthodes d'indexation et de recherche permettant de retourner les documents répondant à l'interrogation d'un utilisateur. Dans cet article, nous nous concentrons sur les bases d'images numériques et posons la question de la représentation de l'image par son contenu.

En imagerie numérique, le problème de l'indexation et de la recherche se pose dans des domaines spécialisés comme le médical aussi bien que dans le domaine grand public. La suite de l'article se focalise sur les bases dites généralistes, c'est-à-dire constituées d'images de divers types, comme c'est le cas sur internet. Les premiers systèmes créés pour la recherche d'images par le contenu sont du type *recherche par similarité*. La requête de l'utilisateur se fait sous la forme d'un exemple (croquis, image, etc.), le système réalise un calcul de similarité entre cet exemple et les images de la base puis sélectionne les meilleures d'entre elles. Ce calcul de similarité se fait à partir de *signatures* pré calculées sur l'image dont l'objectif est de résumer le contenu colorimétrique, textural ou autre de cette dernière.

Le problème majeur rencontré en recherche d'images est le décalage entre la requête subjective de l'utilisateur et les caractéristiques primaires (bas niveau) résumant le contenu de l'image. Ce problème est connu sous le nom de *fossé sémantique*. Il existe deux approches afin de réduire ce fossé, soit travailler à la construction hors ligne d'attributs images plus proches de la sémantique soit encore développer des techniques de recherche interactives permettant l'identification en ligne des caractéristiques pertinentes pour la requête et/ou la production d'un modèle de similarité raffiné. Dans ce dernier cas, on parle de *bouclage de pertinence*.

Notre système a été conçu pour la *recherche de catégorie*, c'est-à-dire pour répondre à une requête générale du type : retrouver les images d'automobiles. Les deux aspects fondamentaux de notre problème sont l'indexation et la recherche. L'objet de la phase d'indexation est la structuration de l'information portée par le contenu de l'image sous la forme d'une ou plusieurs signatures. Dans la suite de l'article, nous présentons notre technique d'indexation basée sur une quantification adaptative des espaces d'attributs. La phase de recherche doit employer au mieux les index pré-calculés afin de fournir des résultats satisfaisant la requête de l'utilisateur. Dans l'objectif de réduire le fossé sémantique, nous proposons une nouvelle stratégie interactive basée sur le principe de l'exploration de la base. Les deux idées sont d'apprendre et d'exploiter un prototype de la catégorie d'images recherchée ainsi que de raffiner la mesure de similarité par un mécanisme de compétition de modèles.

La section 2 est consacrée à la bibliographie sur les signatures images et à la présentation de notre méthode d'indexation intégrant une signature spatiale originale. La section 3 se focalise sur un état de l'art synthétique des techniques de recherche

interactive. Les sections 4, 5 et 6 présentent en détail notre stratégie de recherche interactive. La section 7 propose une synthèse de notre algorithme de recherche et discute des parallèles existant avec quelques méthodes de référence. Pour finir, la section 8 est dédiée aux expérimentations et à l'évaluation des techniques proposées.

2. Indexation

Le problème de l'indexation est de déterminer pour chaque image une représentation qui soit, d'une part compacte, pour être rapidement accessible et facilement comparable, et d'autre part suffisamment complète pour bien caractériser l'image. Les signatures sont le plus souvent basées sur des indices de couleur, texture et forme [BIM 99]. Des critères topologiques (information spatiale) peuvent également être pris en compte.

Schettini *et al.* ont rassemblé dans [SCH 00] les signatures couleur et les similarités entre ces signatures employées dans les principaux systèmes d'indexation. Tous les espaces colorimétriques, de RVB aux espaces CIELab ou CIELuv en passant par HSV sont employés par l'un ou l'autre des systèmes. La discrétisation est faite en général par partition de l'espace colorimétrique. Les auteurs indiquent que la couleur ne suffit pas à indexer des grandes bases d'images, toutefois peu de méthodes intègrent l'information spatiale. A titre d'illustration, citons les corrélogrammes couleur [HUA 97], les histogrammes spatiaux-colorimétriques [CIN 99] et les matrices de transition de couleur [SMI 99].

En ce qui concerne la texture, outre les moments, les attributs les plus courants sont basés sur les matrices de cooccurrence, les filtres de Gabor et les décompositions en ondelettes. Les méthodes à base de bancs de filtres type Gabor ou ondelettes ont été comparées dans [MA 95], les fonctions de Gabor se sont montrées les plus efficaces. Les critères de forme sont très nombreux, mais ils nécessitent d'avoir extrait des régions de l'image [BIM 98] [PHI 02]. La topologie est codée par des graphes d'adjacence, des arbres, des chaînes, etc [BIM 99].

L'ensemble d'attributs que l'on peut calculer est donc très vaste, en particulier si les espaces de représentation sont continus. De plus, ces espaces sont très vides par rapport au nombre d'images présentes dans une base. Donc, le problème est de déterminer un ensemble d'attributs pertinent puis de discrétiser au mieux les espaces.

2.1. Quantification des espaces d'attributs

Pour chacun des espaces d'attributs, il est nécessaire de trouver un nombre raisonnable de classes ou de bins afin d'obtenir une signature compacte. La réduction de la taille des signatures se fait par un découpage de l'espace de représentation qui est soit statique, soit dynamique.

Le découpage statique de l'espace des attributs est la méthode la plus simple de réduction du nombre de classes, elle est indépendante des données. Il existe de nombreuses méthodes pour le découpage régulier ou irrégulier de l'espace colorimétrique. L'espace HSV peut être, par exemple, quantifié sur 3 valeurs par axe colorimétrique, ou encore découpé irrégulièrement en 166 bins, comme dans Visual Seek [SMI 96]. Une telle quantification ne tient pas compte des données. Le découpage se fait de la même façon, pour des couleurs peu représentées, que pour des couleurs dont le nombre de nuances aurait réclamé un découpage plus fin.

La quantification dynamique de l'espace des attributs s'adapte aux données, soit globalement par une prise en compte de toutes les images de la base, soit individuellement pour chaque image. Dans le premier cas, on utilise les méthodes traditionnelles de la classification, le découpage de l'espace se fait selon différents critères, tels que la séparabilité des classes. Citons les méthodes de clustering de type nuées dynamiques ou les cartes topologiques de Kohonen. Dans le deuxième cas, les classes diffèrent d'une image à l'autre. Par exemple, Rubner utilise une palette de couleur adaptée à chaque image [RUB 98].

2.2. Notre approche : signature adaptée à la base

Notre système emploie des attributs de couleur, de texture ainsi que des informations spatiales. Pour la couleur, en l'absence d'un espace colorimétrique prédominant, le choix est laissé à l'utilisateur parmi les 4 espaces RVB, HSV, CIELab et CIELuv. Dans la pratique, ce choix a peu d'influence sur les résultats. Par la suite, nous nous concentrons sur l'espace HSV. Pour la texture, nous utilisons un banc de filtres de Gabor sélectif à trois fréquences et quatre orientations [FOU 01].

La quantification se fait séparément dans chacun des deux espaces (de dimension 12 pour la texture et 3 pour la couleur) par classification pixellaire de type k-means. Seul le nombre de classes est fixé par l'utilisateur, nous étudierons l'influence de ce paramètre sur les résultats, dans la section 8.1.

La composition spatiale est étudiée comme dans la méthode de Smith et Li [SMI 99] en analysant les transitions colorimétriques ou texturales verticales. Nous nous limitons aux transitions verticales car celles-ci sont beaucoup plus informatives que les transitions horizontales pour les images généralistes (une symétrie verticale ne nuit pas à la reconnaissance d'un paysage ou d'une personne, alors qu'une symétrie horizontale modifie complètement l'image). La méthode est fondée sur des cooccurrences de couleurs ou de textures dans un balayage vertical de l'image, de haut en bas. Pour compacter l'information, nous travaillons sur l'image après quantification en couleur ou en texture. Puis une partition spatiale est effectuée par division de l'image en blocs rectangulaires. Le choix du nombre de blocs dépend de la résolution voulue, pour nos tests nous avons divisé l'image en 15×15 blocs. Chaque bloc est affecté à la classe la plus fréquente le composant. On obtient ainsi une image-bloc qui est un résumé basse résolution de l'image originale (cf. fig. 1).

Si la quantification est réalisée sur N classes, une matrice de composition de l'image de dimension $N \times N$ est construite à partir de l'image-bloc. Pour chaque colonne de cette dernière, on comptabilise le nombre de transitions d'un bloc de la classe C_m à un bloc de la classe C_n , que ceux-ci soient adjacents ou non. Pour les transitions d'une classe vers elle-même, seuls sont pris en compte les blocs séparés par au moins un bloc d'une autre classe. Ceci conduit à une invariance par rapport à la taille des régions homogènes de l'image.

En notant $I_{b_{i,j}}$ la classe du bloc $b_{i,j}$ (i et j sont les indices de ligne et de colonne dans l'image-bloc, l'origine du repère est fixée en haut à gauche de celle-ci), la matrice de composition M est définie par :

$$\begin{cases} M_{m,n} = \text{card} \left\{ (b_{i,j}, b_{i',j}), I_{b_{i,j}} = C_m, I_{b_{i',j}} = C_n, i' > i \right\} & \text{pour } m \neq n \\ M_{m,m} = \text{card} \left\{ (b_{i,j}, b_{i',j}), I_{b_{i,j}} = I_{b_{i',j}} = C_m, \exists b_{i'',j}, i < i'' < i', I_{b_{i'',j}} \neq C_m \right\} \end{cases} \quad (1)$$

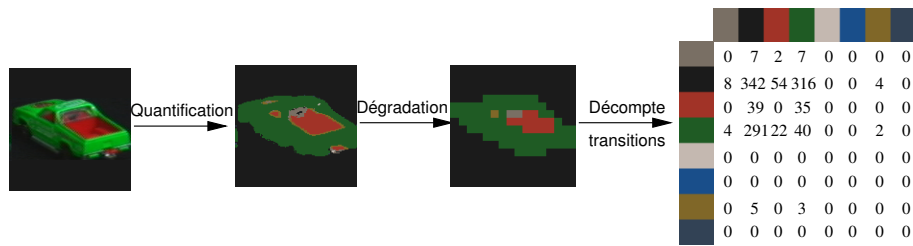


Figure 1. Illustration du processus de calcul de la matrice de composition image. L'espace couleur est quantifié sur 8 classes (cf. échelle de couleur)

Cette méthode diffère de celle de Smith et Li par le fait qu'elle ne nécessite pas de segmentation. De plus, elle s'adapte à tous types d'espaces d'attributs. Les matrices de composition sont donc calculées sur les attributs colorimétriques et sur les attributs texturaux.

Les matrices de composition sont de grande taille ($N \times N$), et souvent très creuses. Nous réduisons leur dimension en effectuant une analyse en composantes principales sur toutes les matrices de la base (après dépliement des matrices sous la forme d'un vecteur de dimension N^2) et en réduisant l'espace de représentation aux N premiers axes.

En résumé, nous obtenons finalement une signature constituée de quatre vecteurs : la distribution couleur (N classes), la distribution texture (N classes), le vecteur de composition colorimétrique (N composantes), le vecteur de composition texturale (N composantes).

3. Recherche interactive d'images : bibliographie

Les techniques classiques de recherche d'images par l'exemple sont limitées par le fossé sémantique. La bibliographie sur le sujet regroupe l'approche indexation et l'approche recherche. Dans le premier cas, les auteurs se focalisent sur les descripteurs (*cf.* section 2) et la fonction de similarité afin de produire une évaluation de la similarité entre images plus proche du niveau interprété. Cox et al. [COX 00] utilisent des expériences réalisées "hors-ligne" pour construire un "modèle utilisateur" intégré dans leur fonction de similarité Bayésienne. Mojsilovic et Rogowitz [MOJ 01] proposent d'identifier les descripteurs bas-niveaux et la mesure de similarité décrivant au mieux des "catégories sémantiques" d'images construites lors d'expériences psychovisuelles. Dans le second cas, il s'agit d'algorithmes de recherche interactive élaborés dans le but de limiter, "en-ligne", le fossé sémantique. L'idée d'introduire l'utilisateur dans la boucle de la recherche (bouclage de pertinence) a été proposée en indexation textuelle [ROC 71]. De nombreux travaux ont, depuis quelques années, montré l'apport incontestable de ce type d'approches pour la recherche d'images.

Comme proposé par Vasconcelos et Kunt [VAS 01], les techniques de recherche interactive peuvent être regroupées en deux catégories : statistique et géométrique.

Sous la dénomination statistique sont regroupées, par exemple, les approches proposées par Geman [GEM 00], Cox [COX 00] ou Meilhac [MEI 99], dont l'objectif est de mettre à jour une fonction de probabilité sur la base d'images, en fonction de l'information obtenue de l'utilisateur durant toute la phase de recherche. Les méthodes d'analyse discriminante [HUA 01] cherchant à construire un classifieur optimal pour la séparation des exemples et des contre-exemples, appartiennent également à ce type d'approches.

La catégorie des méthodes géométriques se rapporte aux systèmes de recherche-par-similarité ou recherche-par-l'exemple. L'utilisateur fait une requête sous la forme d'un exemple (image, croquis, *etc.*) et la recherche se base sur un calcul de similarité entre la signature de cette requête et les signatures des images cibles (c'est-à-dire les images de la base). Le bouclage de pertinence intervient alors pour raffiner la requête et/ou la fonction de similarité. Il s'agit de l'approche la plus répandue, une méthode optimale de ce type est proposée par Rui et Huang [RUI 00].

Les deux catégories de méthodes statistique et géométrique reposent sur des concepts distincts avec leurs avantages et leurs inconvénients. En particulier, l'approche statistique intègre la notion d'incertitude et d'exploration, base de la remise en cause du modèle de similarité bas-niveau, ce qui n'est pas le cas de l'approche géométrique. En effet, cette dernière reste attachée à une mesure de similarité et une requête dont le choix, non trivial et souvent négligé dans la littérature, conditionne la pertinence des résultats. D'un autre côté, les méthodes statistiques sont souvent basées sur une interface plus complexe [CAE 00], des modèles simplificateurs et requièrent un long apprentissage donc plus d'itérations au cours de l'interaction.

L'aspect exploration de la base nous apparaît comme déterminant en recherche interactive. Des approches sans requête initiale, du type de celles proposées par Cox et

al. [COX 00] ou encore Geman et Moquet [GEM 00], gèrent cet aspect en initialisant uniformément sur la base les probabilités des images d'être pertinentes (au sens de la recherche) puis en raffinant itérativement la densité de probabilité grâce à l'information apprise lors de l'interaction avec l'utilisateur. Une technique de recherche-par-l'exemple peut également autoriser l'exploration de la base. Ainsi, le système proposé par Vendrig et al. [VEN 99] sélectionne, à chaque itération, autour de la requête, un large ensemble d'images appelé "ensemble actif" parmi lesquelles une nouvelle sélection détermine les images présentées à l'utilisateur. Ces dernières constituent une "vue d'ensemble" de l'ensemble actif. En choisissant un ensemble actif initialement grand (un quart de la base d'images) puis en réduisant sa taille au fil des itérations, la méthode permet d'explorer au delà des plus proches voisins de la requête.

Comme nous le verrons par la suite, cette idée d'exploration est la base de notre stratégie de recherche. Nous situons notre travail parmi les méthodes géométriques ; toutefois, l'algorithme de recherche est reformulé sous forme stochastique.

4. Stratégie de recherche interactive

Nous nous focalisons sur la recherche de catégories : l'utilisateur souhaite retrouver un ensemble d'images auxquelles il attribue des caractéristiques sémantiques communes. Par exemple, le but peut être de retrouver les images de voiture, de chien ou les portraits, dans une base généraliste. Lorsque l'utilisateur expertise une image, il doit simplement préciser si elle appartient ou non à la catégorie qu'il recherche. Notre système fonctionne sur le principe de la recherche-par-l'exemple, cette dernière étant initiée par une image requête. Le calcul de similarité entre les images cibles et cette requête est à la base de la technique. Toutefois, un de nos objectifs prioritaires est la remise en cause de la requête par exploration de la base.

Notre stratégie de recherche interactive apprend et exploite un *prototype* de la catégorie recherchée, à partir de l'information issue de l'annotation (comme pertinentes ou non-pertinentes) par l'utilisateur des images présentées par le système. Ce prototype (noté \mathcal{P}) est constitué d'un ensemble d'images exemples ou images "types" de la catégorie en question. Nous l'appelons indifféremment *requête composée* ou *requête multiple*.

La recherche est itérative et interactive puisque basée sur le principe du bouclage de pertinence. Nous distinguons deux phases de traitement :

- une phase d'apprentissage du prototype de catégorie, appelée *phase d'exploration*, pour la collecte d'images types pouvant être dispersées dans les espaces de recherche,
- une phase d'utilisation du prototype de catégorie, appelée *phase d'exploitation*, pour l'accumulation des images résultats. Celle-ci intègre une étape de raffinement de la mesure de similarité utilisée pour la comparaison des images de la base à la requête composée et a pour objectif de satisfaire la requête de l'utilisateur en fournissant le plus grand nombre d'images pertinentes.

4.1. Pourquoi construire un prototype de catégorie ?

Intéressons-nous à la répartition des images appartenant à une catégorie donnée (ou classe sémantique) parmi l'ensemble des images d'une base généraliste de 1200 éléments. Nous avons réalisé une analyse en composantes principales sur les signatures de la base entière et projeté l'index de chaque image dans le plan formé par les deux directions de plus grande variance. La figure 2 présente la répartition des images d'une classe "avion" parmi le reste de la base, au sein de deux espaces d'attributs distincts (un espace couleur et un espace texture).

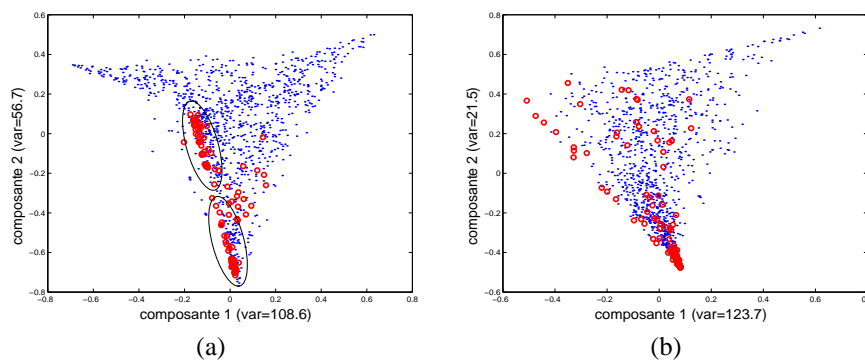


Figure 2. Répartition des images de la classe "avion" (ronds) parmi le reste de la base (points) pour l'espace de caractéristiques couleur HSV (a) et pour un espace texture (b).

La figure 2(a) montre une distribution plutôt bimodale des images pertinentes dans l'espace couleur. En effet, elles se dispersent autour de deux centres d'intérêt (voir les zones entourées) alors que dans l'espace texture (figure 2(b)) on note plutôt une distribution unimodale accompagnée d'une plus forte dispersion.

Les images pertinentes ne sont pas regroupées en un mode unique dans les espaces de recherche. La recherche-par-similarité "classique" (une seule image requête) ne permettant de retrouver que les images qui sont proches de l'image requête (en termes de caractéristiques bas-niveaux), risque de "tomber" dans un maximum local de similarité. En outre, ce maximum est plus ou moins intéressant du point de vue sémantique suivant l'image choisie comme requête. Ces observations illustrent le fossé sémantique.

L'objectif de notre stratégie est de sortir des minima locaux de similarité par l'utilisation d'une technique d'optimisation globale et par la construction d'un prototype multimodal de la catégorie, reflétant la distribution réelle des images pertinentes dans les espaces d'attributs.

4.2. Prototype de catégorie et calcul de similarité

Avec la construction du prototype de catégorie, la requête n'est plus composée d'une image, comme dans le cas classique de la recherche-par-similarité, mais par un ensemble de L images aux contenus plus ou moins variés. En notant R les images de ce prototype, on écrit $\mathcal{P} = \{R_l, 1 \leq l \leq L\}$.

Pour une image I_i de la base (indexée par I_i)¹ et pour une recherche basée sur M attributs (l'espace d'attributs est indicé par k), la mesure de similarité à la requête multiple est calculée comme suit [FOU 01] :

$$S(\mathcal{P}, I_i) = \sum_{k=1}^M \beta^{(k)} s^{(k)}(\mathcal{P}, I_i) \quad (2)$$

où $s^{(k)}(\mathcal{P}, I_i)$ est la similarité requête-cible pour le k -ième espace d'attributs.

Il n'y a pas de refonte des images exemples sous la forme d'une requête unique, contrairement à l'approche classique rencontrée en bouclage de pertinence (cf. section 3). Chacune des images exemples est utilisée : la similarité globale entre la requête et la cible est obtenue en fusionnant les similarités de la cible à chaque image du prototype. Dans le cas où l'image cible est proche d'un des exemples, la similarité à cette image requête doit être favorisée. Dans le cas où la cible est éloignée de toutes les images du prototype, une similarité de type moyenne, est plus adaptée.

Les similarités données par chaque image du prototype sont fusionnées par l'opérateur *barycentre* défini par Huet et Philipp [HUE 98] qui possède les propriétés recherchées :

$$s^{(k)}(\mathcal{P}, I_i) = \frac{\sum_{l=1}^L s^{(k)}(R_l, I_i)^2}{\sum_{l=1}^L s^{(k)}(R_l, I_i)} \quad (3)$$

Différentes mesures de similarité entre images $s^{(k)}(R_l, I_i)$ ont été testées et une similarité basée sur la distance L_1 a été adoptée : $s^{(k)}(R_l, I_i) = 1 - d_{L_1}(R_l, I_i)$.

L'équation (3) se réécrit sous la forme $s^{(k)}(\mathcal{P}, I_i) = \frac{\sigma^2}{\mu} + \mu$, où μ et σ représentent la moyenne et la variance sur les similarités $s^{(k)}(R_l, I_i)$. Lorsque les images exemples sont approximativement situées à égale distance de l'image cible traitée, σ est faible et l'opérateur tend vers la moyenne. Inversement, lorsque ces valeurs sont très différentes, l'opérateur tend vers le maximum. L'image cible est alors plus directement comparée à l'image requête la plus proche.

1. Dans le document, les lettres en gras font référence aux signatures des images ou du prototype.

Pour illustrer cette fusion, imaginons que l'utilisateur recherche des images contenant des pommes. Sachant que la base peut contenir des photographies de pommes rouges et de pommes vertes, les images pertinentes sont regroupées en deux pôles de l'espace de recherche couleur². Dans l'hypothèse où l'utilisateur n'est intéressé que par retrouver un type d'images de pommes, par exemple les pommes rouges, une moyenne est bien adaptée afin de combiner les exemples du prototype qui sont alors uniquement concentrés dans un seul mode. Dans l'hypothèse de la recherche des deux types de pommes, une fusion favorisant les similarités les plus fortes permet de gérer correctement des exemples issus, à présent, de groupes d'images distincts dans l'espace de recherche.

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.

5. Construction du prototype de catégorie

La construction du prototype de la catégorie d'images recherchée se fait interactivement par accumulation des images types identifiées par l'utilisateur parmi les résultats que lui présente le système. Afin que le prototype code correctement la répartition des images pertinentes dans l'espace de recherche, il est impératif "d'explorer" la base, c'est-à-dire de proposer à l'utilisateur des images variées pouvant être éloignées du prototype de catégorie initial.

La construction du prototype de catégorie se fait durant la phase dite d'exploration. Pour mettre en œuvre la stratégie de recherche, nous nous sommes inspirés des techniques de recuit simulé [CHA 00] [HAJ 88].

Nous souhaitons affecter à chaque image de la base une probabilité d'être pertinente vis-à-vis de la catégorie recherchée. Notre choix s'est porté sur une distribution de Boltzmann de la dissimilarité prototype-cible D ³. Elle permet d'assurer que plus une image est proche du prototype, plus sa probabilité est grande. Toutefois, un paramètre T de contrôle (appelé température) permet de modifier, au cours des itérations, la confiance accordée au modèle :

$$P_{\mathcal{P}}(I = I_i) = \frac{1}{Z_T} \exp\left(\frac{-D(\mathcal{P}, I_i)}{T}\right) \quad (4)$$

où Z_T est une constante de normalisation calculée comme la somme des exponentielles sur toutes les images de la base.

2. Nous supposons que la signature image caractérise parfaitement et uniquement l'objet.

3. La somme des poids $\beta^{(k)}$ intervenant dans la combinaison des modèles (*cf.* eq. 2) est égale à 1 et $s^{(k)}(\mathcal{P}, I_i) \leq 1 \forall k$. Les valeurs de similarités $S(\mathcal{P}, I_i)$ sont donc bornées entre 0 et 1, et la dissimilarité peut s'exprimer en fonction de la similarité normalisée : $D(\mathcal{P}, I_i) = 1 - S(\mathcal{P}, I_i)$.

Le système tire au sort selon ces probabilités un lot d'images pour présentation à l'utilisateur. Ce dernier, par l'annotation "pertinente" ou "non-pertinente" de ces images, valide leur appartenance au prototype de catégorie. Les images pertinentes identifiées viennent compléter le prototype tandis que les autres sont écartées de la recherche.

La construction du prototype de catégorie se fait selon un schéma interactif de type bouclage de pertinence, par succession d'une phase de sélection probabiliste des images candidates, d'une phase de validation par l'utilisateur et d'une phase de mise à jour du prototype. L'algorithme suivant fournit une synthèse de la stratégie de construction du prototype.

- Initialisation : *Choix d'une température $T = T_0$ assez grande*
- Etape 1 : *Pour chaque image cible, calcul de la similarité à la requête composée (eq. 2) puis de la probabilité associée (eq. 4)*
- Etape 2 : *Tirage au sort et affichage des images, en fonction de la loi de probabilité précédente*
- Etape 3 : *Annotation des images pertinentes par l'utilisateur et complément du prototype de catégorie*
- Etape 4 : *Décroissance de T et retour à l'étape 1 jusqu'à fin exploration*

Plus le paramètre de température T est élevé, moins l'influence de la dissimilarité est forte dans le processus de sélection probabiliste, et plus l'horizon de recherche est large autour du prototype de catégorie. Lorsque T décroît, cet horizon diminue et la recherche se recentre autour du prototype appris. Le système tend alors vers la phase d'exploitation.

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

$$T_j = C^j \cdot T_0 \quad (5)$$

pour laquelle les constantes C ($C < 1$) et T_0 sont à déterminer.

5.1. Paramétrage du système

Les paramètres C et T_0 déterminent la distribution de probabilité sur la base, ces derniers doivent donc être réglés en fonction de l'image la plus proche et de l'image la plus éloignée du prototype de catégorie.

Précisons avant tout que la dissimilarité D utilisée pour le réglage des paramètres a été préalablement égalisée. Considérons la probabilité δ de sélectionner une image dont la dissimilarité est inférieure à un seuil d_{borne} . D'après l'équation (4), nous avons :

$$\delta = \sum_{I_i \in db | D(\mathcal{P}, \mathbf{I}_i) \leq d_{\text{borne}}} P_{\mathcal{P}}(I_i) = \sum_{I_i \in db | D(\mathcal{P}, \mathbf{I}_i) \leq d_{\text{borne}}} \frac{1}{Z_T} \exp\left(\frac{-D(\mathcal{P}, \mathbf{I}_i)}{T}\right)$$

Soit, en notant $d_i = D(\mathcal{P}, \mathbf{I}_i)$, $\forall I_i \in db$:

$$\delta = \sum_{0 \leq d_i \leq d_{\text{borne}}} \frac{1}{Z_T} \exp\left(\frac{-d_i}{T}\right)$$

Pour déterminer des relations explicites entre T , d_{borne} et δ , nous considérons la variable D continue (hypothèse raisonnable dès lors que la base contient beaucoup d'images). D ayant une densité de probabilité constante après égalisation, ceci conduit à l'approximation suivante de δ :

$$\delta \approx \frac{\int_0^{d_{\text{borne}}} \exp\left(\frac{-x}{T}\right) dx}{\int_0^{\infty} \exp\left(\frac{-x}{T}\right) dx}$$

Soit, après calcul :

$$d_{\text{borne}} = -T \cdot \ln(1 - \delta) \quad (6)$$

En considérant que seules les images dont la dissimilarité est inférieure à d_{borne} peuvent être sélectionnées, nous introduisons la notion d'*ensemble de recherche* ou ensemble des images cibles, comme l'ensemble exploré par l'algorithme à l'itération courante du processus. La relation (6) signifie que le paramètre de température fixe la dissimilarité maximale sur l'ensemble des images cibles, en fonction de δ (qui est fixé à $\delta = 1 - 10^{-5}$ dans nos expérimentations). La décroissance de température implique une décroissance de la taille de l'ensemble de recherche.

Nous définissons l'*ensemble maximal de recherche* comme le plus large ensemble d'images exploré sur la base. Il s'agit de l'ensemble de recherche exploré à la première itération de la phase d'exploration, qui peut être la base entière (privée de l'image requête dans le cas où celle-ci appartient à la base). T_0 est donc calculée en fonction de la dissimilarité maximale sur l'ensemble maximal de recherche. En appelant D_{MAX} cette dissimilarité, on obtient (cf. eq. 6) :

$$T_0 = \frac{-D_{\text{MAX}}}{\ln(1 - \delta)} \quad (7)$$

Pour nos expérimentations, la base entière est choisie comme ensemble maximal ; D_{MAX} est dans ce cas la dissimilarité de l'image de la base la plus éloignée de la requête.

De même, nous définissons l'*ensemble minimal de recherche* comme le plus petit ensemble exploré soit l'ensemble correspondant à la dernière itération de l'exploration. L'ensemble minimal de recherche correspond aux $20 \times N_{\text{aff}}$ images (N_{aff} est le

nombre d'images affichées à l'écran à chaque itération) les plus proches du prototype de catégorie. En identifiant d_{\max} la dissimilarité maximale sur l'ensemble minimal de recherche et en choisissant la durée de la phase d'exploration (nous appelons n l'indice de la dernière itération), il est possible de fixer la température finale de l'algorithme :

$$T_{\text{finale}} = T_n = \frac{-d_{\max}}{\ln(1 - \delta)}$$

Par suite, nous calculons la constante C intervenant dans le calcul de T_j (cf. eq. 5) :

$$C = \sqrt[n]{\frac{d_{\max}}{D_{\text{MAX}}}} \quad (8)$$

Ainsi, la phase d'exploration est paramétrée automatiquement par la donnée de la durée de la phase d'exploration (seul choix laissé à l'utilisateur) après détermination de D_{MAX} et d_{\max} à la première itération de la recherche. Nous pensons que 5 à 10 itérations est une fourchette permettant un bon compromis pour une recherche courte et la construction d'un prototype de catégorie pertinent (i.e. une exploration efficace de la base).

REMARQUE. — La taille de l'ensemble minimal de recherche (fixée à $20 \times N_{\text{aff}}$) est paramétrée en fonction du nombre d'images présentées à l'utilisateur. En effet, pour que le tirage probabiliste dans cet ensemble ait un sens, son cardinal doit être largement supérieur au nombre d'images tirées.

5.2. Effet du changement de prototype

D'une itération à l'autre de la phase d'exploration, le prototype de catégorie évolue par l'apport de nouvelles images exemples. La modification du prototype agit sur les dissimilarités requête-cibles. En particulier, la dynamique (i.e. les valeurs extrêmes) des dissimilarités sur la base est réduite du fait de l'effet moyennneur du barycentre de Huet et Philipp (cf. section 4.2).

La figure 3 présente les histogrammes des dissimilarités calculées sur toute la base, évoluant au cours des trois premières itérations d'une recherche sur la base Corel.

Nous remarquons que la forme de l'histogramme des dissimilarités évolue d'une itération à l'autre. Plus précisément, on observe un léger grossissement de la queue gauche de la distribution au cours de l'exploration ainsi qu'un déplacement du mode principal vers la gauche. Ce changement de forme de l'histogramme des dissimilarités correspond au raffinement itéré du prototype pendant la phase d'exploration.

5.3. Sélection probabiliste parmi les images cibles

L'algorithme d'exploration fonctionne par sélection successive d'ensembles de recherche de taille décroissante. L'utilisateur expertise cet ensemble de recherche en annotant pertinente ou non pertinente les images que lui présente le système. Il est donc

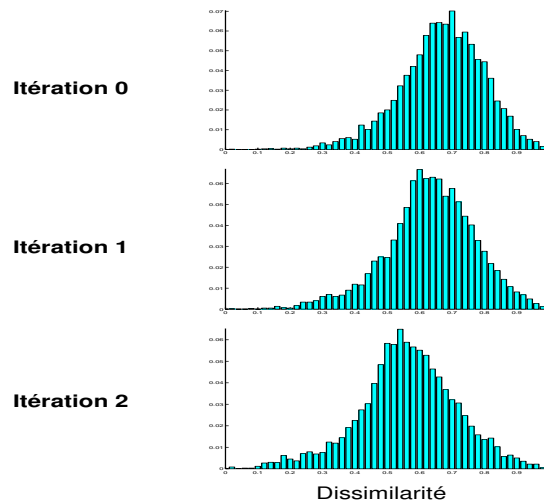


Figure 3. Evolution de l'histogramme des dissimilarités requête-cible pour les trois premières itérations du bouclage de pertinence. La recherche porte sur la catégorie "tigre" de la base Corel.

important que celles-ci "représentent" au mieux la diversité (en terme de contenu) de l'ensemble de recherche. La première stratégie consiste à sélectionner des images en tirant selon la loi de pertinence (cf. eq. 4). Plus le nombre d'images affichées (N_{aff}) augmente, meilleure est cette représentation. Il s'agit donc de trouver un compromis entre le nombre d'images que l'utilisateur doit expertiser et les performances globales de l'exploration.

Etant donné le faible nombre d'images présentées à l'utilisateur (moins de quelques dizaines), la procédure de tirage probabiliste peut être complétée par la prise en compte d'un critère de dispersion portant sur le lot des N_{aff} images à afficher.

L'idée est de tirer plusieurs lots de N_{aff} images, puis de choisir selon un critère donné le meilleur lot à présenter à l'utilisateur. Il est par exemple possible de choisir le lot qui maximise la moyenne des dissimilarités entre les N_{aff} images.

REMARQUE. — Le choix du nombre d'images présentées par le système est laissé à l'utilisateur. Par défaut, nous considérons que $N_{\text{aff}} = 20$ représente un bon compromis entre le nombre d'images à expertiser et la représentativité de l'ensemble de recherche.

6. Exploitation du prototype de catégorie

La stratégie de recherche se compose de deux phases aux buts distincts. La phase d'exploitation a pour objectif de satisfaire la requête utilisateur en fournissant un

maximum d'images pertinentes alors que l'exploration cherche à recueillir des images types (et donc pertinentes) variées sans pour autant maximiser leur nombre. Les deux phases sont complémentaires, le succès de la stratégie de recherche dépend de chacune d'entre elles.

6.1. *Caractéristiques de l'exploitation*

L'exploitation repose sur les principes classiques de la recherche-par-similarité : les images sont classées en fonction de leur similarité à la requête puis les K premières sont présentées à l'utilisateur, dans l'ordre de ce classement. Dans notre cas, la requête correspond au prototype de catégorie appris et "gelé" à l'issue de l'exploration. La similarité d'une image cible au prototype est calculée comme précisé à la section 4.2.

L'étude de la distribution des images pertinentes dans les espaces d'attributs (en rapport à une requête) a montré que celles-ci ne sont pas simplement regroupées en un seul mode (cf. section 4.1, fig. 2). Nous constatons également que cette distribution change fortement d'un espace d'attributs à l'autre. L'exemple présenté montre que l'espace couleur regroupe mieux les images d'avion de la base traitée. Les caractéristiques de couleur sont donc plus discriminantes que la texture pour cette recherche.

La phase d'exploitation repose sur un mécanisme de compétition d'attributs (ou compétition de modèles) permettant d'identifier et de renforcer les espaces de caractéristiques les plus discriminants. Encore une fois, la stratégie est itérative (bouclage de pertinence) mais utilise à présent les images pertinentes ainsi que les images non-pertinentes annotées par l'utilisateur.

6.2. *Raffinement de la mesure de similarité*

La fonction de similarité est une somme pondérée des similarités requête-cibles sur chaque modèle (cf. eq. 2). Il est donc possible de modifier l'influence de certaines caractéristiques en jouant sur les paramètres $\beta^{(k)}$ intervenant dans la fonction. Le réglage des poids peut être réalisé manuellement par l'utilisateur du système [CAR 99] toutefois, de même que Heinrichs et al. [HEI 00], nous préférons opérer une mise à jour semi-automatique à partir de l'information d'interaction. Contrairement à l'approche *heuristique* de Heinrichs et al. mais de manière similaire à celle de Rui et Huang [RUI 00], notre travail se situe parmi les méthodes *basées-optimisation*. Le processus vise à renforcer le poids des attributs les plus discriminants et inversement pour les moins discriminants. C'est le processus de *compétition de modèles*.

Aucun *a priori* n'existe sur l'influence des attributs pour la recherche, donc les poids sont fixés à la même valeur pour tous les modèles. Ensuite, l'interactivité fournit des exemples et des contre-exemples permettant de décider des modèles discriminants et non-discriminants, base de la mise à jour des poids. Le critère optimisé est l'erreur quadratique entre la similarité réelle, mesurée entre l'image cible I_i et la requête, et la

similarité désirée ($S_d^{I_i}$) pour cette même image cible I_i . La règle de minimisation est la règle LMS [WID 60] :

$$\beta^{(k)*} = \beta^{(k)} + \mu(S_d^{I_i} - S(\mathcal{P}, I_i))s^{(k)}(\mathcal{P}, I_i) \quad (9)$$

(M est le nombre de modèles en compétition, $1 \leq k \leq M$, $\mu \in \mathbb{R}^+$)

Les valeurs de la similarité désirée (sachant que les similarités sont bornées entre 0 et 1) sont fixées grâce aux annotations de l'utilisateur :

$$\begin{cases} S_d^{I_i} = 1 & \text{si } I_i \text{ est pertinente} \\ S_d^{I_i} = 0 & \text{sinon} \end{cases}$$

Pour un ensemble d'exemples et de contre-exemples donné, la minimisation est itérée jusqu'à ce que l'erreur quadratique moyenne sur cet ensemble passe sous un seuil fixé empiriquement. La règle LMS augmente le poids des modèles qui "classent" le mieux les images pertinentes et diminue celui des modèles qui "classent" incorrectement les images non-pertinentes [FOU 01].

A l'issue de la phase d'exploration, une première mise à jour des poids est effectuée à partir de tous les exemples composant le prototype de catégorie. Les images non-pertinentes expertisées ne sont alors pas utilisées car elles sont trop nombreuses pour garantir la stabilité de la mise à jour [FOU 02]. Pour les itérations suivantes, le mécanisme de compétition se base sur tous les exemples et contre-exemples accumulés au cours de l'exploitation.

7. Synthèse et discussion

Notre stratégie de recherche se différencie des techniques classiques selon trois points (cf. fig. 4). Tout d'abord, la phase d'exploration repose sur une sélection probabiliste des images présentées à l'utilisateur (module 1) et sur la construction par accumulation, d'un prototype de la catégorie d'images recherchée (module 2). Ensuite, la phase d'exploitation est caractérisée par le raffinement de la mesure de similarité (module 3).

Nous détaillons à présent l'algorithme complet de la stratégie de recherche.

NOTATIONS. — I_i : image quelconque de la base,
 les lettres en gras désignent les signatures, i.e. une image I_i est indexée par I_i ,
 \mathcal{P} : prototype de catégorie,
 db : ensemble des images de la base,
 $\mathbb{1}$: fonction indicatrice,
 $\{I_{\text{pert}}\}^{(j)}$, $\{I_{\text{non-pert}}\}^{(j)}$: ensembles d'images annotées pertinentes et non-pertinentes par l'utilisateur parmi les images affichées à l'itération j ,
 \mathcal{E} : ensemble d'apprentissage servant à la compétition de modèles.

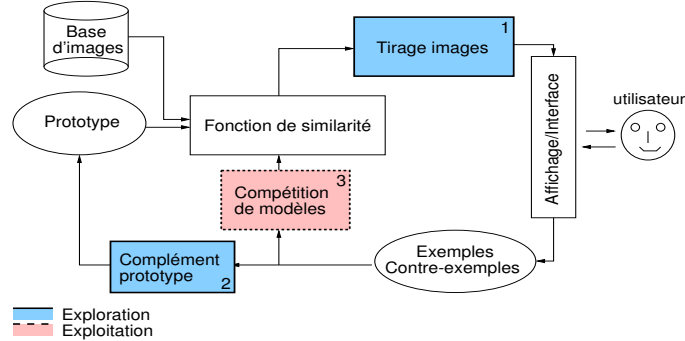


Figure 4. Schéma synoptique de notre stratégie de recherche.

Entrée: Nombre d'images à afficher N_{aff} ,
image requête R ,
nombre d'itérations de la phase d'exploration n

1. Mode exploration

$j \leftarrow 0$; $\mathcal{P} \leftarrow \{R\}$;

Initialisation uniforme des $\beta^{(k)}$;

Calcul de $S(\mathcal{P}, I_i) \forall I_i \in db$ (cf. eq. 2);

Calcul de T_0 (cf. eq. 7);

Calcul de C (cf. eq. 8);

Calcul de $P_{\mathcal{P}}(I = I_i) \forall I_i \in db$ (cf. eq. 4);

pour ($j = 1..n$) **faire**

 Tirage au sort des images résultats $\rightarrow \mathcal{A}^{q*}$ (cf. eq. ??);

 Annotation des résultats $\rightarrow \{I_{\text{pert}}\}^{(j)}, \{I_{\text{non-pert}}\}^{(j)}$;

$\mathcal{P} \leftarrow \mathcal{P} \cup \{I_{\text{pert}}\}^{(j)}$;

 Calcul de $S(\mathcal{P}, I_i) \forall I_i \in db$ (cf. eq. 2);

 Calcul de $T_j = C^j \cdot T_0$;

 Calcul de $P_{\mathcal{P}}(I = I_i) \forall I_i \in db$ (cf. eq. 4);

fin

2. Passage au mode exploitation

$\mathcal{E} \leftarrow \mathcal{P}$;

Mise à jour des $\beta^{(k)}$ à partir de \mathcal{E} (cf. eq. 9);

Calcul de $S(\mathcal{P}, I_i) \forall I_i \in db$ (cf. eq. 2);

tant que (*Utilisateur = non satisfait*) **faire**

 Affichage des images les plus proches de la requête;

 Annotation des résultats $\rightarrow \{I_{\text{pert}}\}^{(j)}, \{I_{\text{non-pert}}\}^{(j)}$;

$\mathcal{E} \leftarrow \mathcal{E} \cup \{I_{\text{pert}}\}^{(j)} \cup \{I_{\text{non-pert}}\}^{(j)}$;

 Mise à jour des $\beta^{(k)}$ à partir de \mathcal{E} (cf. eq. 9);

 Calcul de $S(\mathcal{P}, I_i) \forall I_i \in db$ (cf. eq. 2);

$j \leftarrow j + 1$;

fin

Affichage des images pertinentes $\bigcup_j \{I_{\text{pert}}\}^{(j)}$

Nous avons discuté, dans la section 3, des approches interactives en séparant les méthodes basées similarité et celles basées statistiques. Des comparaisons et des rapprochements sont cependant possibles, en particulier lorsque, comme dans notre système, la similarité n'est plus définie entre deux images mais entre une image et un ensemble d'images.

Le premier point de comparaison concerne la construction et la mise à jour de la fonction de pertinence (ou la densité de probabilité) utilisée dans les systèmes statistiques. Sous l'hypothèse d'indépendance des espaces de représentation (couleur, texture, ...), la probabilité totale s'exprime alors comme un produit des probabilités calculées dans chaque espace [VAS 00]. Ceci revient à ne pas considérer de compétition explicite entre les espaces. En revanche, en considérant une combinaison linéaire des similarités (d'autres types de compétitions ont également été étudiés [RUI 00]), il est possible de gérer explicitement cette compétition de modèles de représentation durant la recherche. L'adaptation des poids peut alors se faire selon des heuristiques ou des critères d'optimisation à partir des exemples collectés. La méthode proposée dans la section 6.2 s'inscrit dans cette logique.

Conscients de cette limitation, Pauwels et al. proposent un système probabiliste où les espaces de représentation sont évalués et seuls ceux jugés suffisamment discriminants sont conservés pour le calcul de la probabilité globale [CAE 00]. Geman propose également une compétition entre les modèles via des poids calculés pour chaque image ([GEM 00], troisième méthode).

Un autre point concerne le calcul ou la mise à jour des poids intra-classes. Dans notre système, la similarité est obtenue selon un critère de fusion (*cf.* eq. 3) des similarités simples $s^{(k)}(\mathbf{R}_l, \mathbf{I}_i)$. Dans un contexte statistique, des techniques d'estimation de densités de probabilité peuvent être utilisées pour évaluer la probabilité en tout point \mathbf{I}_i de l'espace considéré en fonction des observations \mathbf{R}_l [CAE 00].

Le choix de l'architecture des systèmes est donc important et influe sur les possibilités d'adaptation au cours de l'interaction utilisateur. Les méthodes statistiques et géométriques, généralement séparées, présentent pourtant plusieurs points de rapprochement qu'il nous semble intéressant d'approfondir.

8. Expérimentations

Les tests sont réalisés sur deux bases généralistes composées d'images couleurs de divers types. La première base est appelée *Généraliste* et comprend 1200 images récupérées sur Internet. La vérité-terrain établie manuellement sur cet ensemble regroupe les images en 15 classes séparées du type : avion, image aérienne, voiture, ours, éléphant, *etc.* La seconde base est 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 *rappel* (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.

8.1. Quantification des espaces d'attributs

Afin d'évaluer notre schéma d'indexation et en particulier l'aspect quantification des espaces d'attributs, nous comparons les performances obtenues par l'utilisation de notre signature couleur (quantification dynamique dans l'espace couleur HSV) à celles obtenues par l'histogramme couleur de Smith et Chang [SMI 96]. Cet histogramme est issu d'une quantification statique sur 166 classes de l'espace HSV. Nous évaluons en particulier l'influence du nombre de classes utilisé dans notre approche. Le critère de qualité retenu est la précision moyennée sur trois valeurs de rappel (0.2, 0.5 et 0.8) [SMI 98]. Les résultats sont réunis dans le tableau 5.

	Sign25	Sign49	Sign100	Sign166	Hist166
Eléphant	0.3	0.3	0.29	0.27	0.15
Coucher de soleil	0.23	0.23	0.24	0.25	0.13
Voiture	0.08	0.09	0.1	0.1	0.04

Figure 5. Précisions moyennées sur 3 valeurs de rappel, obtenues pour des recherches sur la base Généraliste, pour 3 catégories d'images différentes. Performances comparées entre notre signature couleur (HSV) de taille variant entre 25 et 166 classes (Sign25 à Sign166) et l'histogramme de Smith et Chang (Hist166).

Nous observons, tout d'abord, que la précision moyenne est approximativement deux fois plus grande pour notre signature que pour celle de Smith et Chang, quel que soit le nombre de classes pour la quantification et quelle que soit la catégorie de la requête. Ceci signifie que les images pertinentes sont mieux classées et donc mieux discriminées par notre signature couleur. La seconde observation concerne la taille de la signature. Il n'y a pas de tendance nette quant à l'évolution des performances avec le nombre de classes utilisées pour la quantification de l'espace couleur. Nous constatons plutôt une réelle robustesse vis-à-vis de ce paramètre : les performances sont stables sur une large échelle de valeurs du nombre de classes. Nous utilisons donc préférentiellement les signatures de taille minimale, c'est-à-dire 25.

Afin de fournir une analyse comparative plus fine, nous présentons à présent les courbes de capacité obtenues sur nos deux bases d'images, pour l'histogramme de

Smith et Chang et pour notre signature couleur (cf. fig. 6). La courbe de capacité est définie comme l'histogramme des dissimilarités sur tous les couples d'images de la base [STR 94]. Celle-ci nous permet d'évaluer la dispersion des signatures au sein de l'espace de recherche : plus cette dispersion augmente, meilleure est la qualité discriminatoire de la signature. La mesure de capacité dépend de la métrique, la distance L_2 est utilisée pour le test présenté.

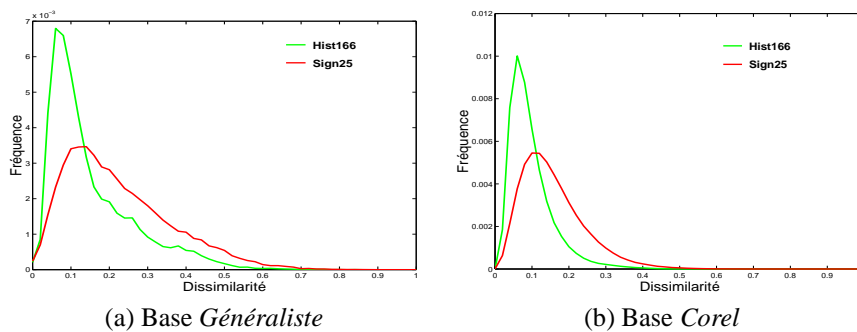


Figure 6. Comparaison des courbes de capacité de l'histogramme couleur 166 classes de Smith et Chang (*Hist166*) à notre signature couleur (espace couleur HSV) de 25 classes (*Sign25*) pour deux bases d'images. La métrique utilisée est L_2 .

Nous observons que la dispersion des images dans l'espace de recherche est plus forte avec notre signature couleur qu'avec l'histogramme de Smith et Chang, pour les deux bases d'images considérées. La capacité discriminatoire de notre approche est donc meilleure, ce qui confirme les résultats précédemment énoncés.

Notre approche par quantification dynamique des espaces d'attributs permet donc, par rapport à un histogramme statique, une indexation plus efficace de la base : les signatures images sont mieux dispersées dans l'espace de recherche et les résultats des requêtes sont meilleurs. Un avantage majeur de l'approche dynamique est de permettre une réduction de la taille de la signature sans dégradation des performances. Bien que l'évaluation présentée soit concentrée sur la signature couleur, nous avons également évalué notre approche pour les autres modèles de représentation et pour l'utilisation combinée de ces derniers [FOU 01].

8.2. Stratégie de recherche exploration/exploitation

Nous nous intéressons, à présent, à la validation de notre stratégie de recherche exploration/exploitation. La figure 7 illustre le fonctionnement de la stratégie globale en présentant les résultats fournis par le système durant la première étape des deux phases du processus de recherche.

La requête est relative aux images de la catégorie "porte" de la base *Corel*. Les résultats présentés par le système à la première itération de l'exploration ne sont pas

(a) Image requête

(b) Images présentées par le système à la première itération de l'exploration

(c) Images présentées par le système à la première itération de l'exploitation

Figure 7. *Illustration de la stratégie d'exploration/exploitation pour la recherche d'images de la catégorie "porte" dans la base Corel.*

classés, les images sont sélectionnées par tirage au sort selon les probabilités définies à la section 5. Nous observons peu d'images pertinentes (1 seule) parmi cet ensemble, toutefois, la variété des résultats fournis par l'exploration permet d'identifier des individus pertinents "éloignés" de la requête initiale. Les résultats à la première itération de la phase d'exploitation montrent que la construction du prototype de catégorie est efficace dans l'objectif de retrouver de nombreuses images pertinentes. Précisons, en outre, que les "bonnes images" de ce dernier ensemble viennent s'ajouter à celles déjà obtenues.

La figure 8 fournit une seconde illustration de la diversité des images résultats retrouvées par notre approche. La requête porte maintenant sur la classe "tigre". Les images présentées correspondent à un extrait des images pertinentes accumulées durant la recherche, soit cinq itérations d'exploration et une itération d'exploitation.

(a) Image requête

(b) Images pertinentes accumulées durant 6 itérations
(5 itérations d'exploration et 1 itération d'exploitation)

Figure 8. Exemple d'ensemble d'images pertinentes accumulées après 6 bouclages, pour une recherche d'images de la catégorie "tigre".

Il est intéressant de comparer le comportement de notre approche à une stratégie "classique" de recherche par bouclage de pertinence. La figure 9 propose les per-

performances de notre stratégie par exploration/exploitation comparées à la stratégie de compétition de modèles seule. Dans ce dernier cas, le système fonctionne sur le principe classique de la recherche-par-similarité : le système propose uniquement à l'utilisateur les images les plus proches de la requête (requête unique). Le bouclage a alors pour but de raffiner la fonction de similarité par compétition de modèles (*cf.* section 6.2). Pour référence, nous calculons également les performances d'une stratégie de recherche totalement aléatoire. Les critères de qualité sont moyennés sur 20 requêtes distinctes de la catégorie "portrait" sur la base et *Généraliste* et 20 requêtes distinctes de la catégorie "fleur" sur la base *Corel*. La recherche se base sur des attributs de couleur, de texture et de topologie (*cf.* section 2). 20 images sont présentées à l'utilisateur, pour annotation, à chaque itération. Le critère de performance retenu est le rappel tracé en fonction du nombre d'itérations (*cf.* fig. 9 (a) et (b)) et la précision tracée en fonction du rappel à la septième itération⁴ de la recherche (*cf.* fig. 9 (c) et (d)).

(a) Base *Généraliste* (catégorie "portrait") (b) Base *Corel* (catégorie "fleur")

(c) Base *Généraliste* (catégorie "portrait") (d) Base *Corel* (catégorie "fleur")

Figure 9. Comparaison des performances obtenues par la stratégie de recherche complète, pour un bouclage simple par compétition de modèles et pour une recherche aléatoire.

4. Le choix de la septième itération est réalisé au hasard. Par le biais de ces courbes précision-rappel, nous évaluons l'apport de la requête multiple à une itération quelconque de la stratégie de recherche.

Nous constatons qu'avec notre approche le taux de rappel est faible durant les premières itérations puis augmente rapidement après quatre ou cinq bouclages. Cette tendance est différente de celle observable pour la recherche-par-similarité classique pour laquelle la croissance du rappel est plus régulière. Cette singularité du comportement de notre stratégie vient des deux modes distincts de fonctionnement. Durant la phase d'exploration, le système construit le prototype de catégorie et ce n'est que durant l'exploitation que ce dernier cherche à maximiser le nombre d'images pertinentes retrouvées, en mettant à profit l'information collectée. Les performances sont alors supérieures à l'approche classique. L'inflexion de la courbe de rappel aux environs de cinq itérations correspond à la transition (passage automatique) entre les deux modes. La stagnation de performance observable vers les dernières itérations est due au fait qu'en fin de recherche, les images pertinentes encore non-détectées demeurent peu nombreuses et très délicates à isoler.

La réelle différence de qualité des résultats entre les trois méthodes et plus particulièrement la supériorité de notre approche est confirmée par les courbes de précision-rappel (*cf.* fig. 9 (c) et (d)). En conséquence, la stratégie par exploration/exploitation est effective pour construire une requête multiple variée permettant d'accumuler, après quelques itérations, un nombre plus important d'images pertinentes que la recherche-par-similarité avec bouclage classique.

9. Conclusion

Nous avons proposé une nouvelle méthode pour l'indexation et la recherche d'images par le contenu dans des bases généralistes. Notre signature code des informations de couleur, de texture et intègre une notion de composition spatiale. Afin d'optimiser la taille des index sans perte de qualité sur les résultats, notre signature est construite à partir d'une quantification des espaces d'attributs adaptée à la base. Les expériences menées sur la taille des signatures et le type de quantification nous amènent à conclure qu'une quantification adaptée à la base est plus efficace qu'un découpage régulier des espaces d'attributs ; elle permet de réduire la taille des index (jusqu'à 25) sans dégradations des performances.

La phase de recherche est basée sur un algorithme interactif utilisant le principe de l'exploration de la base. Celui-ci se divise en deux phases de traitement : l'exploration et l'exploitation. La première a pour objectif de construire un prototype de la catégorie d'images recherchée, contenant de nombreuses images de la catégorie, même si celle-ci présente une grande variabilité sur la base. L'algorithme, inspiré du recuit simulé, est de type stochastique et gère l'exploration par diminution progressive de l'espace de recherche. La phase d'exploitation s'emploie à raffiner, par compétition de modèles, la fonction de similarité utilisée pour la comparaison du prototype aux images cibles. L'évaluation réalisée sur deux bases généralistes distinctes a permis de valider notre stratégie d'indexation et de recherche d'images.

Du point de vue des perspectives, une signature globale reste restrictive sémantiquement parlant, et nous nous intéressons, par conséquent, au calcul de signatures

partielles plus proches de l'objet dans l'image. D'autre part, notre stratégie d'exploration offre de nombreuses pistes de travail et un point d'entrée pour la comparaison aux méthodes statistiques de bouclage de pertinence.

Remerciements

Les auteurs remercient D. Declercq (équipe Signal du laboratoire ETIS) pour sa collaboration sur les aspects d'analyse statistique.

10. Bibliographie

- [BIM 98] BIMBO A. D., MUGNAINI M., PALA P., TURCO F., « Visual querying by color perceptive regions », *Pattern Recognition*, vol. 31, n° 9, 1998, p. 1241-1253.
- [BIM 99] BIMBO A. D., *Visual information retrieval*, Morgan Kaufmann Ed, San Francisco, USA, 1999.
- [CAE 00] CAENEN G., FREDERIX G., KUIJK A., PAUWELS E., SCHOUTEN B., « Show me what you mean! PARISS : A CBIR-interface that learns by example », *International Conference on Visual Information Systems (Visual'2000)*, vol. 1929, 2000, p. 257-258.
- [CAR 99] CARSON C., THOMAS M., BELONGIE S., HELLERSTEIN J., MALIK J., « Blobworld : A System for Region-Based Image Indexing and Retrieval », *Third Int. Conf. on Visual Information Systems*, June 1999.
- [CHA 00] CHALMOND B., "Eléments de modélisation pour l'analyse d'images", Springer, 2000.
- [CIN 99] CINQUE L., LEVIALDI S., PELLICANÒ A., « Color-Based Image Retrieval Using Spatial-Chromatic Histograms », *IEEE International Conference on Multimedia Computing and Systems*, 1999.
- [COX 00] COX I., MILLER M., MINKA T., PAPATHOMAS T., YIANILOS P., « The Bayesian Image Retrieval System, PicHunter : Theory, Implementation and Psychophysical Experiments », *IEEE Transactions on Image Processing*, vol. 9, n° 1, 2000, p. 20-37.
- [FOU 01] FOURNIER J., CORD M., PHILIPP-FOLIGUET S., « RETIN : A content-based image indexing and retrieval system », *Pattern Analysis and Applications Journal, Special issue on image indexation*, vol. 4, n° 2/3, 2001, p. 153-173.
- [FOU 02] FOURNIER J., CORD M., « Long-term similarity learning in content-based image retrieval », *International Conference in Image Processing (ICIP'02)*, Rochester, New-York, USA, September 2002.
- [GEM 00] GEMAN D., MOQUET R., « A Stochastic Feedback Model for Image Retrieval », *RFIA'2000*, vol. III, Paris, France, February 2000, p. 173-180.
- [HAJ 88] HAJEK B., « Cooling schedules for optimal annealing », *Mathematics of Operations Research*, vol. 13, n° 2, 1988, p. 311-329.
- [HEI 00] HEINRICH A., KOUBAROULIS D., LEVIENAISE-OBADIA B., ROVIDA P., JO-LION J., « Image Indexing and Content based Search using Pre-attentive Similarities », *RIA02000*, vol. 2, Paris, France, April 2000, p. 1616-1631.

- [HUA 97] HUANG J., KUMAR S. R., MITRA M., ZHU W.-J., ZABIH R., « Image Indexing Using Color Correlograms », *Conference on Computer Vision and Pattern Recognition (CVPR'97)*, 1997.
- [HUA 01] HUANG T., ZHOU X., « Image Retrieval with Relevance Feedback : From Heuristic Weight Adjustment to Optimal Learning Methods », *International Conference in Image Processing (ICIP'01)*, vol. 3, Thessaloniki, Greece, October 2001, p. 2-5.
- [HUE 98] HUET F., PHILIPP S., « Fusion of images interpreted by a multi-scale fuzzy classification », *Pattern Analysis and Applications Journal*, vol. 1, 1998, p. 231-247.
- [MA 95] MA W., MANJUNATH B., « Image indexing using a texture dictionary », *SPIE Conference on Image Storage and Archiving System*, vol. 2606, Philadelphia, Pennsylvania, October 1995, p. 288-298.
- [MEI 99] MEILHAC C., NASTAR C., « Relevance Feedback and Category Search in Image Databases », *IEEE International Conference on Multimedia Computing and Systems (ICMCS'99)*, Florence, Italy, June 1999, p. 512-517.
- [MOJ 01] MOJSILOVIC A., ROGOWITZ B., « Capturing image semantics with low-level descriptors », *International Conference in Image Processing (ICIP'01)*, vol. 1, Thessaloniki, Greece, October 2001, p. 18-21.
- [PHI 02] PHILIPP-FOLIGUET S., VIEIRA M. B., SANFOURCHE M., « Fuzzy segmentation of color images and indexing of fuzzy regions », *CGIV'2002*, Poitiers, France, 2002, p. 507-512.
- [ROC 71] ROCCHIO J., « Relevance feedback in information retrieval », *G. Salton editor, The SMART retrieval system : Experiments in Automatic Document Processing*, Prentice Hall Inc., 1971.
- [RUB 98] RUBNER Y., TOMASI C., GUIBAS L., « A Metric for Distributions with Applications to Image Databases », *IEEE International Conference on Computer Vision*, Bombay, India, January 1998, p. 59-66.
- [RUI 00] RUI Y., HUANG T., « Optimizing Learning In Image Retrieval », *Conf on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, Hilton Head, SC, June 2000, p. 236-243.
- [SCH 00] SCHETTINI R., CIOCCA G., ZUFFI S., « Color in databases : Indexation and Similarity », *First International Conference on Colour in Graphics and Image Processing CGIP'2000*, Saint-Etienne, France, October 2000.
- [SMI 96] SMITH J., CHANG S., « VisualSEEK : a fully automated content-based image query system », *ACM Multimedia Conference*, Boston, USA, November 1996, p. 87-98.
- [SMI 98] SMITH J., « Image Retrieval Evaluation », *IEEE Workshop on Content-based Access of Image and Video Libraries (CBAIVL'98)*, June 1998, p. 112-113.
- [SMI 99] SMITH J., LI C., « Image Classification and Querying Using Composite Region Templates », *Computer Vision and Image Understanding*, vol. 75, n° 1-2, 1999, p. 165-174.
- [STR 94] STRICKER M., SWAIN M., « The Capacity of Color Histogram Indexing », *Conference on Computer Vision and Pattern Recognition (CVPR'94)*, 1994, p. 704-708.
- [VAS 00] VASCONCELOS N., « Bayesian models for visual information retrieval », PhD thesis, Massachusetts Institute of Technology, 2000.
- [VAS 01] VASCONCELOS N., KUNT M., « Content-based retrieval from image databases : current solutions and future directions », *International Conference in Image Processing*

(*ICIP'01*), vol. 3, Thessaloniki, Greece, October 2001, p. 6-9.

[VEN 99] VENDRIG J., WORRING M., SMEULDERS A. W. M., « Filter Image Browsing, Exploiting Interaction in Retrieval », HUIJSMANS D. P., SMEULDERS A. W. M., Eds., *Visual Information and Information Systems*, vol. 1614 de *Lecture Notes in Computer Science*, Berlin, 1999, Springer-Verlag, p. 147-154.

[WID 60] WIDROW B., HOFF M., « Adaptive switching circuits », *IRE WESCON*, New-York, 1960, p. 96-104.

Matthieu Cord a soutenu sa thèse de doctorat en *Traitement des Images* en 1998. Il a effectué un séjour post-doctoral à la *Katholieke Universiteit Leuven (Belgique)* sur le thème de la modélisation 3D en imagerie aérienne. Actuellement Maître de Conférences au laboratoire ETIS (UMR 8051 CNRS), ses activités concernent l'indexation et la recherche d'images ainsi que l'analyse d'images aériennes.

Jérôme Fournier a soutenu sa thèse en *Sciences Traitement des Images et du Signal* en octobre 2002. Il a effectué son doctorat au sein du laboratoire ETIS sur le thème de l'indexation et de la recherche d'images par le contenu. Ses travaux portaient sur une caractérisation compacte, souple et performante des images ainsi que sur le développement d'algorithmes de recherche interactifs.

Sylvie Philipp-Foliguet a enseigné les mathématiques en lycée de 1977 à 1984. Elle a obtenu sa thèse de doctorat en informatique de l'Université de Paris 6, en 1988. Sa thèse portait sur l'analyse de texture pour la détection de défauts dans les images radiographiques de soudures. Elle est actuellement Professeur des Universités à l'ENSEA de Cergy-Pontoise où elle dirige l'équipe Image du laboratoire ETIS. Ces travaux ont principalement porté sur la segmentation, l'analyse d'images et l'interprétation mais récemment, ses activités se sont recentrées en indexation et recherche d'images par le contenu.

Annexe pour le service de fabrication

Article pour la revue :

RSTI - TSI

Auteurs :

Matthieu Cord — Jérôme Fournier — Sylvie Philipp-Foliguet

Titre de l'article :

*Approche interactive de la recherche
d'images par le contenu*

Titre abrégé :

Recherche interactive d'images

Traduction du titre :

*Interactive search in content-based image retrieval : a browsing retrieval
strategy*

Date de cette version :

9 janvier 2004

Coordonnées des auteurs :

- téléphone : 01.30.73.62.83
- télécopie : 01.30.73.62.87
- Email : {cord,philipp}@ensea.fr

Logiciel utilisé pour la préparation de cet article :

L^AT_EX, avec le fichier de style `article-hermes.cls`,
version 1.10 du 2000/03/28.

Formulaire de copyright :

Joindre le formulaire de copyright signé, récupéré sur le web à l'adresse
<http://www.hermes-science.com>