



MEMOIRE

Présenté à

L'Institut Nationale des Sciences Appliquées de Lyon
(Ecole Doctorale Informatique et Information pour la Société)

En vue de l'obtention

du Diplôme d'Etudes Approfondies

Par

Haytham Elghazel

**Approches textuelles pour
la recherche d'images**

Encadré par

Khalid Idrissi

Soutenu le 06 juillet 2004

Remerciements

Ce travail de recherche a constitué ma première aventure de recherche en France. Il n'aurait pas été aussi fructueux sans l'aide de plusieurs personnes. Je remercie tous ceux qui ont contribué, de près ou de loin, à réaliser ce travail.

Je tiens à adresser mes remerciements aux membres du Laboratoire d'Informatique en Images et Systèmes d'informations **LIRIS** que j'ai pu côtoyer durant la période de mon stage et qui ont su rendre mon travail agréable.

Plus particulièrement, je tiens à remercier **Mr Khalid IDRISSE** et **Mr Atilla BASKURT** pour leurs encadrements continus, pour les remarques constructives qu'ils m'ont fournies ainsi que pour leurs précieux conseils durant toute la période du stage.

Ma gratitude, mon profond respect et mes remerciements à tous les membres du jury et à tous les rapporteurs pour leur attention consacrée à l'égard de mon travail.

ELGHAZEL Haytham

Table de matières

Table de matières	3
Table des figures	5
Introduction générale	6
Chapitre 1. Etat de l'art	8
1. Approches textuelles	8
1.1. Introduction	8
1.2. Les réseaux sémantiques	9
1.2.1. Les thésaurus.....	9
1.2.2. Les graphes conceptuels	9
1.2.3. Les réseaux de neurones	9
1.3. L'indexation en sémantique latente « LSI ».....	10
1.4. Conclusion.....	12
2. LSI dans la recherche d'images	12
3. Conclusion	12
Chapitre 2. Approche proposée	13
1. Introduction	13
2. Les descripteurs utilisés	13
2.1. Les couleurs dominantes	13
2.2. Le gradient couleur.....	14
2.3. Les filtres de Gabor	15
2.4. Le gradient directionnel	15
2.5. Les mots clés	16
2.6. Conclusion.....	16
3. Adaptation des descripteurs	16
3.1. Les couleurs dominantes	16
3.2. Le gradient couleur.....	17
3.3. Les filtres de Gabor	18
3.3.1. Identification des régions d'intérêt.....	19
3.3.2. Analyse de la texture pour les régions d'intérêt.....	19
3.4. Le gradient directionnel	21
3.5. Les mots clés	22
3.6. Conclusion.....	23
4. Conclusion	23
Chapitre 3. Tests et évaluation	24
1. Introduction	24

2. Fonctionnalités	24
2.1. L'indexation	24
2.2. La recherche	24
2.3. La classification.....	24
3. Evaluation de la recherche.....	25
3.1. Le paramètre « Recall »	25
3.2. Le paramètre « Precision ».....	26
3.3. Evaluation sur les bases d'images	26
3.3.1. Evaluation sur la Base Simplicity.....	26
3.3.2. Evaluation sur la Base MPEG-7.....	28
4. Tests de comparaison avec l'approche Pecenovic	29
5. Quelques résultats de classification	30
Conclusions et perspectives	31
Annexe	32
1. L'espace Lab	32
2. Les bases d'images.....	33
2.1. Simlicity	33
2.2. MPEG-7	33
Références	34

Table des figures

Figure 1.1 Matrice Termes x Documents.....	10
Figure 1.2 Démarche suivie pour l'indexation des documents avec la LSI.....	11
Figure 2.1 Exemples d'histogrammes normalisés du gradient couleur.....	14
Figure 2.2 Exemples d'histogrammes de gradient directionnel.....	16
Figure 2.3 Exemple d'adaptation des couleurs dominantes pour une image I.....	17
Figure 2.4 Comportement des caractéristiques du gradient couleur.....	18
Figure 2.5 Identification des régions d'intérêt.....	19
Figure 2.6 Identification des coefficients de Gabor les plus discriminants.....	20
Figure 2.7 Démarche suivie pour l'adaptation des coefficients de Gabor pour une image I.....	21
Figure 2.8 Comportement des caractéristiques du gradient directionnel.....	22
Figure 2.9 Matrice Descripteurs x Images.....	23
Figure 3.1 Classes d'images considérées dans la base Simplicity.....	26
Figure 3.2 Evaluation de la base Simplicity.....	27
Figure 3.3 Classes d'images considérées dans la base MPEG-7.....	28
Figure 3.4 Evaluation de la base MPEG-7.....	28
Figure 3.5 Comparaison de l'approche Pecenovic à notre approche.....	29
Figure 3.6 Exemples de groupes distincts retournés par l'approche de classification.....	30
Figure 3.7 Exemples d'exceptions dans l'approche de classification.....	30
Figure 4.1 Espace Lab.....	32
Figure 4.2 Base Simplicity.....	33
Figure 4.3 Base MPEG-7.....	33

Introduction générale

De nos jours, les images occupent une place prépondérante au sein de la société. L'apparition des dispositifs d'acquisition, des capacités de stockage (mémoires de masse), des systèmes de transmission et de diffusion (réseaux Internet, réseaux spécialisés à haut débit) nécessite un besoin croissant en traitement de l'image.

Des bases de données ont alors été construites afin de stocker toutes ces images et au fil des années elles ont gagné en importance occupant parfois jusqu'à plusieurs téraoctets de mémoire et ayant des natures très diverses. Afin de pouvoir consulter et retrouver au mieux les images, le problème qui se pose est celui de l'indexation de ces bases d'images.

Depuis le début des années 90, l'accès aux bases par le contenu visuel est devenu un sujet de recherche particulièrement dynamique. Les recherches visent à indexer automatiquement les images par leur propre contenu, par le biais de caractéristiques, souvent de bas niveau, traduisant la couleur, la forme et la texture. Cette approche est habituellement désignée par *Recherche d'images par le contenu (RIPC)*. La recherche dans ce domaine est active et de nombreux travaux ont vu le jour, mais la majorité de ces travaux ne prennent pas assez compte l'aspect sémantique des images, ce qui a pour conséquence de limiter leur efficacité.

D'autre part, de nombreuses activités ont été menées sur les techniques d'indexation et de recherche textuelle, et ceci bien avant l'arrivée de l'Internet pour les applications de base de données, de gestions documentaire, etc. Ce domaine évolue vers une tâche très importante : la classification sémantique qui est la tendance récente et qui permet de regrouper entre eux des documents ayant des thématiques proches. Elle fournit par la suite, en réponse à une requête de l'utilisateur, un document se rattachant au même thème que la requête même s'il n'existe pas (ou peu) de mots communs entre l'un et l'autre.

Par ailleurs, bien que la nature des données soit différente, les problématiques d'indexation et de recherche de textes et d'images présentent des similarités. Dans les deux cas, la pertinence de la recherche est réalisée lorsque celle-ci s'est basée sur la similarité du sens sémantique et non uniquement sur celles des descripteurs de bas niveau (mots clés pour les textes et couleurs, par exemple, pour les images).

Dans ce contexte, l'évolution des approches de recherche d'information textuelles dans le domaine de classification sémantique, se place comme un point de départ pour ce sujet de DEA. Le travail que nous présentons en indexation et recherche d'images par le contenu essaye de :

- * Montrer dans quelle mesure ces approches textuelles peuvent être appliquées aux images.
- * Proposer, dans la mesure où c'est possible, un schéma d'adaptation incluant le choix des descripteurs, leur extraction puis leur utilisation avec la méthode de recherche textuelle, jugée la plus adéquate.

Afin de situer notre approche, nous effectuons dans *le premier chapitre* une synthèse bibliographique traitant deux parties :

- * La première partie présente différents types d'approches textuelles et surtout ceux s'inspirant d'analyseurs sémantiques. Pour chacune, nous indiquerons les avantages et les inconvénients dans le but d'arriver à choisir la technique la plus adéquate pour la problématique de recherche dans les bases d'images : *l'indexation en sémantique latente LSI* dans notre cas.
- * Dans la deuxième partie, notre technique étant choisie, nous détaillerons certaines approches proposées par d'autres auteurs pour l'adaptation de cette technique d'indexation aux images.

Dans *le second chapitre*, nous présenterons notre adaptation de la LSI aux images, en substituant certains descripteurs de l'image aux mots clefs habituellement utilisés. Les choix réalisés seront détaillés, puis justifiés.

Dans *le troisième chapitre*, nous exposons quelques résultats obtenus. Cette démarche va nous permettre d'évaluer les performances de notre approche et de vérifier la faisabilité de l'adaptation des méthodes textuelles à la problématique de recherche dans les bases d'images.

A l'issue du *troisième chapitre*, nous résumons notre travail en dégagant les apports de cette méthode. Nous précisons ses limites et certaines perspectives visant à ouvrir des extensions à ce thème de recherche.

1. Approches textuelles

1.1. Introduction

En *Recherche d'Information (RI)*, domaine qui s'est développé à partir des années 1960, les techniques d'indexation nécessitent de décrire l'ensemble des documents et de sélectionner pour chaque document un ensemble de mots clés.

Pour les premières techniques d'indexations désignées par *texte intégral*, la pertinence d'un document pour une requête s'exprime sous la forme d'un score. Ce score se base sur la présence de mots de la requête dans le texte. Par exemple, pour la requête « le document doit contenir le mot éléphant et herbe », un document contenant le mot « éléphant » et « herbe » est une bonne réponse, contrairement à un document qui ne les contient pas.

De ce fait, le jeu de descripteurs (mots clés) n'est souvent ni suffisant, ni assez précis pour décrire n'importe quel document. Il se peut aussi, que pour des problèmes de synonymie ou de polysémie, certains documents répondant à la question d'un utilisateur puissent ne pas être retrouvés. La description manuelle est donc un processus lent et qui ne garantit pas de bons résultats.

Très vite, des problématiques connexes se sont aussi greffées autour de la RI. Parmi les plus courantes et les plus utiles, l'interaction avec l'utilisateur permet d'obtenir progressivement des documents de plus en plus pertinents. Certains auteurs se sont ensuite essayés à simuler cette interaction, ou au moins en partie, en proposant des techniques permettant de résoudre ces problèmes sémantiques (synonymie, polysémie...). Le problème de synonymie vient du fait que plusieurs utilisateurs peuvent utiliser plusieurs mots pour décrire un même objet : Des études montrent que deux personnes choisissent les mêmes mots-clés pour désigner un même sujet dans seulement 20% des cas. La polysémie au contraire est le fait qu'un même mot peut avoir plusieurs sens.

Ces contributions ont permis la naissance de nouvelles techniques d'indexation textuelles, celles à analyseurs sémantiques désignées par les techniques du *langage contrôlé*.

Comme précisé auparavant, le but de ce travail de DEA est d'arriver à fournir des proximités sémantiques entre des images. C'est pourquoi nous avons choisi, dans un premier temps, de faire le tour sur la deuxième famille de techniques d'indexation vu l'importance qu'elles donnent à la sémantique.

Le panel des techniques d'indexation à analyseurs sémantiques existants est varié. Dans chacun de ces processus, l'utilisateur intervient de manière différente comme nous le décrirons plus loin. Cette intervention peut être plus ou moins importante. Nous allons donc présenter plusieurs techniques (les réseaux sémantiques « thésaurus, graphes conceptuels et réseaux de neurones » et l'indexation en sémantique latente « LSI ») afin de justifier le choix de la technique à partir de laquelle s'inspire notre approche.

1.2. Les réseaux sémantiques

Dans les réseaux sémantiques, le sens d'une donnée est fourni par la place qu'elle occupe dans la structure et les relations qu'elle entretient avec les autres unités. Nous allons présenter trois formes de réseaux sémantiques : les thésaurus, les graphes conceptuels et les réseaux de neurones.

1.2.1. Les thésaurus

Dans [CHA 88], Chaumier définit un thésaurus comme étant « Un dictionnaire de termes (descripteurs) normalisés, organisés de manière conceptuelle, et reliés entre eux par des relations sémantiques ». Ces relations sémantiques sont des trois formes : relation d'équivalence, relation d'association et relation hiérarchique.

Il s'en suit qu'un thésaurus correspond à des listes structurées de descripteurs, éditées par des comités experts en documentation et dans un domaine scientifique et technique : il est entièrement construit par les humains.

D'autre part, les thésaurus permettent de faire des recherches, de l'indexation et de donner des termes équivalents, mais ils ne sont pas capables d'évaluer des proximités sémantiques entre deux mots, deux notions ou deux textes.

De ce fait, ce type de technique ne nous intéresse pas dans notre approche

1.2.2. Les graphes conceptuels

Alors que les thésaurus permettent un travail uniquement sur la terminologie, la théorie des graphes conceptuels, développé par Sowa [SOW 84], est un formalisme général de représentation des connaissances basé sur la linguistique et fondé sur la logique. Il a été conçu dans l'objectif de développer un système de logique qui représente de façon plus simple les structures du langage naturel.

En effet, les graphes conceptuels offrent la possibilité de représenter des phrases. Mais « la plupart des approches se fondent sur une analyse syntaxique de chaque phrase pour construire sa représentation conceptuelle » [ZWE et BOU 97]. Ainsi, avant de créer les graphes conceptuels, les phrases sont souvent transcrites sous la forme d'arbre syntaxique.

De ce fait, les opérateurs logiques sont appliqués. Il est ainsi possible de fusionner des graphes, de les projeter les uns sur les autres, d'appliquer les règles de dérivation canonique, d'inférence, etc.

Il est donc possible de comparer des phrases entre elles, mais ces comparaisons sont plus difficiles qu'avec une simple représentation vectorielle. De plus, ce type de technique reste difficilement utilisable dans notre approche. En effet, avec cette forme de représentation, l'intervention de l'humain est importante. Il est impératif de faire une analyse de toutes les images.

1.2.3. Les réseaux de neurones

Les réseaux de neurones, en analogie avec le cerveau humain, sont des outils d'analyse statistique permettant de construire un modèle de comportement à partir de données qui sont des exemples de ce comportement [KM 02]. En effet, les réseaux de neurones apprennent afin de répondre correctement. Une fois cette phase d'apprentissage effectuée, le réseau peut appliquer ce qu'il a appris à de nouvelles données.

De par leur souplesse et leur capacité d'apprentissage, les réseaux de neurones sont utilisés dans des domaines variés et notamment la classification. Construit à partir d'exemples

de chaque classe qu'il a apprise, un réseau de neurones est normalement capable de déterminer à quelle classe appartient un nouvel élément qui lui est soumis.

« Le grand avantage des réseaux de neurones réside dans leur capacité d'apprentissage automatique, ce qui permet de résoudre des problèmes sans nécessiter l'écriture de règles complexes, tout en étant tolérant aux erreurs » [KM 02].

Cependant, ce sont de véritables boîtes noires qui ne permettent pas d'interpréter les modèles construits. En cas, d'erreurs du système, il est quasiment impossible d'en déterminer la cause : C'est pour cela nous éviterons d'utiliser cette technique dans notre approche.

1.3. L'indexation en sémantique latente « LSI »

L'indexation en sémantique latente (*Latent Semantic Indexing LSI*) [DUM 93] [DUM 94] est une technique d'indexation s'inspirant de la *LSA (Latent Semantic Analysis)* développées au début des années 90 [DEE et al 90]. Ce modèle fut conçu comme une nouvelle approche de l'indexation et de la récupération automatique d'informations dans des bases de données constituées de documents textuels. Son but est d'améliorer les techniques d'indexation textuelles et ainsi de permettre le rapprochement sémantique d'un certain nombre de documents au travers des mots les composant. En effet, la LSA permet de résoudre les problèmes de synonymie et de polysémie.

La LSI est une technique d'analyse statistique qui vise à décrire de manière économique les cooccurrences de termes qui surviennent au sein d'un ensemble de documents, puis d'en déduire des proximités sémantiques entre termes. En effet, deux mots peuvent être considérés sémantiquement proches s'ils sont utilisés dans des contextes similaires. Le contexte d'un mot est ici défini comme l'ensemble des mots qui apparaissent conjointement avec lui. Ainsi, les mots « écolier » et « lycéen » sont considérés sémantiquement proches car ils apparaissent tous deux avec des mots comme « classe », « cours », « examen », etc.

En effet, grâce à cette analyse statistique, le sens de chaque mot est caractérisé par un vecteur dans un espace de grande dimension. L'angle formé par deux vecteurs correspondra à la proximité sémantique de ces mots, sans toute fois que celle-ci soit déterminée.

Comme première étape, cette analyse consiste à construire une matrice d'occurrences Termes x Documents (Fig.1.1). Ainsi, pour la représentation d'un ensemble de N documents par M termes, nous aurons une matrice A de taille $M \times N$, dont chaque colonne est associée à un document et contient le nombre d'occurrence des M termes dans le document.

$$A = \begin{array}{c} \begin{array}{c} t_1 \\ t_2 \\ \vdots \\ t_M \end{array} \end{array} \begin{array}{c} D_1 \quad D_2 \quad \dots \quad D_N \\ \begin{array}{|c|c|c|c|} \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline & & & \\ \hline \end{array} \end{array}$$

Figure 1.1 Matrice Termes x Documents.

Une fois la matrice remplie, l'étape suivante consiste à réduire ces dimensions par le biais d'une décomposition en valeurs singulières (*Singular Value Decomposition SVD*). Celle-ci permet de décomposer la matrice A en un produit de trois autres matrices :

$$A = USV^T \quad (1)$$

Où U est une matrice orthogonale de taille $M \times N$ ($UU^T=I_M$) de description des termes, V est une matrice orthogonale de taille $N \times N$ ($VV^T=I_N$) de description des documents et S est une matrice diagonale de taille $N \times N$.

Les colonnes de U et V sont connues sous le nom de vecteurs singuliers respectivement droits et gauches de A . Les éléments diagonaux de S sont appelés valeurs singulières de A : ce sont les racines carrées non nulles des N valeurs propres de AA^T .

A partir d'un certain nombre k , nous nous apercevons de l'existence de valeurs singulières très faibles et qui peuvent être négligées dans la matrice. De ce fait, nous démontrons qu'il y'a une meilleur approximation A_k de A qui est donnée par :

$$A_k = U_k S_k V_k^T$$

Et pour l'obtenir il suffit de :

- * Ranger les éléments de S par ordre décroissant et réarranger les colonnes correspondantes de U et V de façon à garder l'égalité (1).
- * Garder les k premières valeurs de S et les colonnes et lignes correspondantes de U et V^T respectivement.

Note : Il n'y a pas de règle établie pour la détermination de k , le critère de sélection est expérimental et dépend du corpus de documents utilisé. Instinctivement nous pouvons supposer que plus k est faible plus nous éliminons du bruit dans les données, mais que parallèlement nous perdons de l'information : *il faut prendre grand soin lors du choix de ce nombre.*

Cette réduction va nous permettre de ne garder que les k termes les plus significatifs, et afin de calculer la similarité documents-documents, les documents sont représentés dans un espace vectoriel de dimensions k (Fig.1.2). Les colonnes du produit de matrices $S_k \times V_k^T$ sont les coordonnées des vecteurs documents. Par la suite, la LSI s'appuie sur une mesure basée sur le calcul de cosinus de l'angle formé par deux vecteurs, celui-ci représentant la similarité.

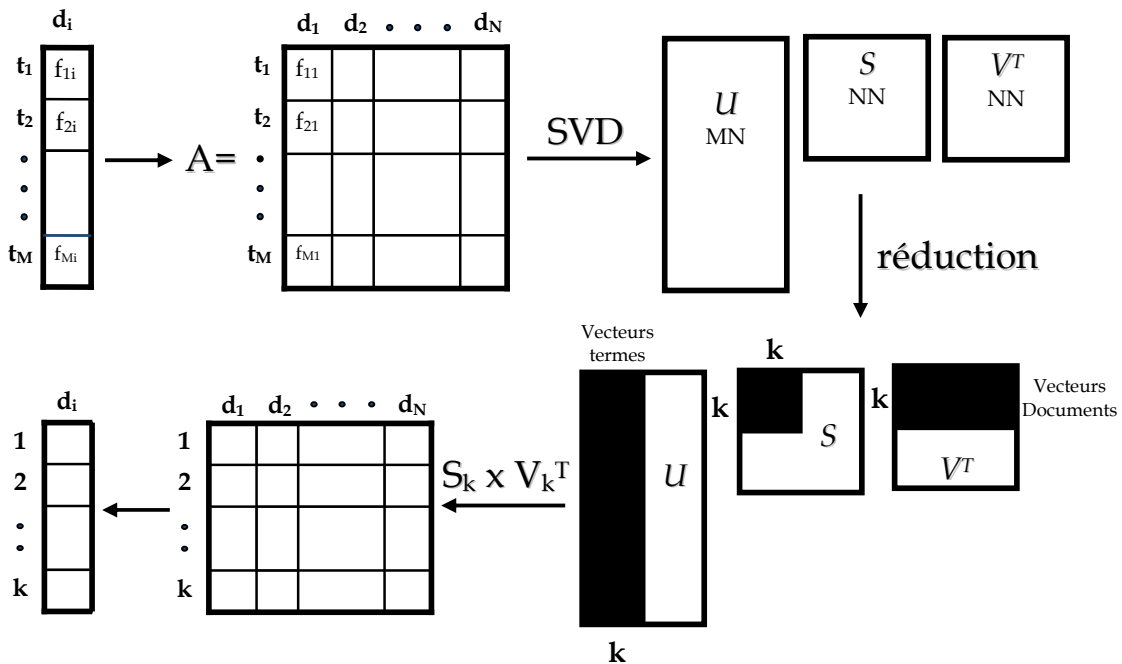


Figure 1.2 Démarche suivie pour l'indexation des documents avec la LSI.

1.4. Conclusion

Grâce à sa capacité d'établir la proximité sémantique entre des mots et entre des documents, à sa souplesse et à sa simplicité, la LSI a été utilisée non seulement dans l'indexation de corpus de textes ou la recherche d'informations mais également dans l'indexation d'images : c'est pourquoi nous avons opté pour cette technique dans notre problématique de recherche d'images.

Dans le paragraphe suivant, nous allons détailler certaines approches d'autres auteurs dans l'adaptation de cette technique d'indexation aux images.

2. LSI dans la recherche d'images

Quelques auteurs proposent d'appliquer la LSI aux images. Pecenovic [PEC 97] [PEC et al 98] propose une adaptation de l'approche LSI aux images en remplaçant la matrice de cooccurrence Termes-Documents par une matrice Images-Descripteurs où les vecteurs descripteurs sont donnés par la combinaison de descripteurs de bas niveau (la couleur et la texture) et des mots clés. De ce fait, il a essayé d'imaginer une nouvelle approche pour traduire la notion d'occurrence pour des descripteurs qui ont habituellement des valeurs numériques.

L'approche consiste à sélectionner pour chaque descripteur, un ensemble d'apprentissage représentatif de la base d'image utilisée, puis de calculer pour cet ensemble la moyenne et l'écart type de la valeur du descripteur. L'occurrence du descripteur pour une image donnée sera donner par le théorème de la limite centrale.

Avec cette approche, Pecenovic a montré des résultats encourageants. Le choix de ses descripteurs a été motivé par sa simplicité comme pour l'histogramme couleur qui reste toutefois sensible au bruit d'illumination et pauvre d'informations spatiochromimétriques.

[ZHA et al 00] utilise LSI en considérant pour chaque image l'histogramme 2D couleur $H \times S$, chaque composante quantifiée à 10 valeurs, comme vecteur descripteur. Il ajoute à ce vecteur un vecteur descripteur de mots clés saisis lors de l'indexation. Les résultats obtenus ne semblent pas satisfaisants, bien que la base de travail utilisée est limitée à 50 images.

Dans [Praks et al 03-1], Pavel Praks utilise une base d'image à niveau de gris et considère que chaque image est vue comme un vecteur de pixels et que chaque pixel est décrit par son niveau de gris. En effet, pour une image de taille 640 x 480, il lui faut un vecteur descripteur de taille 307200 mots clés. Cette approche nécessite beaucoup de capacité en terme de temps et de puissance, c'est pourquoi l'auteur a proposé dans [Praks et al 03-2] un nouvel algorithme rapide de LSI qui ne dépend pas de la dimension du vecteur descripteur. Bien que la base d'images utilisée comporte des images à nuance de gris assez similaires, les résultats ne semblent pas intéressants en terme de classification en raison de l'utilisation direct des valeurs de pixels dans le vecteur descripteur. Ceci conduit à des résultats très sensibles aux moindres changements dans l'image (homothétie, éclairage, etc.).

3. Conclusion

Dans ce chapitre, nous venons de présenter l'intérêt de la LSI dans la présentation de la sémantique fondamentale des documents en décrivant son principe, ses principales caractéristiques et en citant quelques approches existantes dans son adaptation aux images.

Dans le chapitre qui suit, nous allons détailler notre contribution pour l'adaptation de LSI à la problématique de recherche d'images.

Chapitre 2. Approche proposée

1. Introduction

Dans ce chapitre, nous allons présenter notre méthodologie pour l'adaptation de la technique d'indexation en sémantique latente LSI à la problématique de recherche d'images. Ceci est basé sur la substitution des mots clés, habituellement utilisés avec cette technique, par certains descripteurs de l'image. Dans le but d'atteindre notre objectif de classification sémantique des images, l'ensemble de descripteurs qui seront choisis, devrait refléter le plus que possible la similitude perceptuelle humaine.

Ainsi, nous allons présenter dans un premier temps, l'ensemble de caractéristiques d'images privilégiées dans notre contribution afin de passer dans un deuxième temps à l'adaptation de ces caractéristiques à la LSI.

2. Les descripteurs utilisés

Pour traiter le problème de manque de la sémantique dans la problématique de recherche d'images, il faut prendre plusieurs descripteurs qui ne doivent pas être corrélés pour qu'ils ne décrivent pas les mêmes choses. En effet chacun des descripteurs doit contenir une partie de la sémantique.

En raison de contraintes de temps et pour montrer la faisabilité de l'adaptation de la LSI aux images, nous avons commencé par travailler sur un nombre limité de descripteurs mais qui sont porteurs de la sémantique. Nous détaillons dans ce qui suit l'ensemble de ces descripteurs.

2.1. Les couleurs dominantes

La couleur est l'attribut le plus utilisé dans le domaine de l'indexation et de la recherche par le contenu. Nous avons utilisé pour la description de la couleur les couleurs dominantes [IDR et al 02] définies comme des paramètres déduits de l'histogramme couleur de l'image après examen du nombre de pixel les plus important sur l'ensemble des bins. Chaque classe couleur dominante est caractérisée par 3 paramètres (2) :

- * La couleur dominante CD_i définie comme un vecteur à 3 dimensions, correspondant aux trois composantes du système de représentations de couleurs par exemple RGB, Lab, HSV.
- * Le pourcentage P_i que représente la classe de couleurs par rapport à l'image.
- * La cohérence spatiale SCR_i de la classe de couleurs utilisée dans le but de combler, en partie, l'absence d'information spatiale dans l'histogramme couleur : plus la cohérence spatiale sera proche de 0 plus la couleur sera dispersée spatialement dans l'image (et réciproquement).

Pour le calcul des couleurs dominantes, nous avons considérés l'espace couleur Lab (cf. Annexe) comme espace de travail vu son intérêt d'être uniforme au niveau de la perception humaine.

$$\{CD_i(L_i, a_i, b_i), P_i, SCR_i\} \quad (2)$$

Dans [IDR 03], Idrissi montre qu'en moyenne 4 à 6 couleurs suffisent pour modéliser l'histogramme d'une image. Dans ce travail de DEA, nous avons limité le nombre de couleurs dominantes à 4.

2.2. Le gradient couleur

Nous avons introduit ce descripteur pour représenter une information globale de l'image, relative au fait d'avoir dans celle-ci des contours plus ou moins francs. Pour cela nous calculons le maximum de l'histogramme normalisé du gradient couleur (espace couleur choisi est toujours Lab).

$$G = p_{\max} [\nabla(L, a, b)]$$

La figure 2.1 illustre parfaitement notre propos. En effet, une image contenant peu de contours francs (a) verra son gradient (b) prendre essentiellement des faibles valeurs. Par conséquent l'histogramme du gradient (c) possédera un pic vers les faibles valeurs du gradient et dont le cardinal sera relativement important. En revanche pour une image riche en contour (d), le gradient (e) prendra forcément de fortes valeurs et de ce fait l'histogramme (f) sera plus étalé, ce qui entraînera un maximum d'histogramme plus faible.

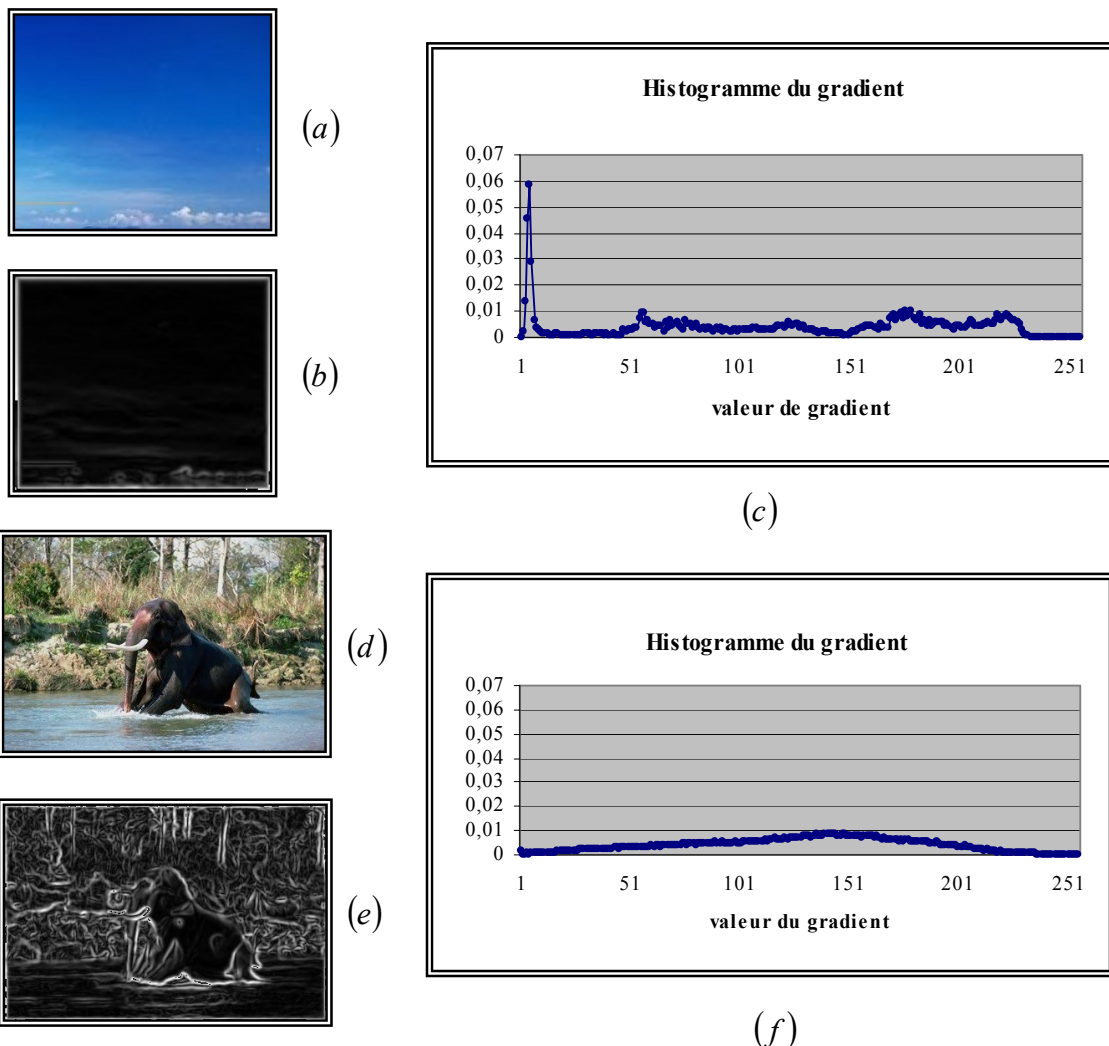


Figure 2.1 Exemples d'histogrammes normalisés du gradient couleur.

2.3. Les filtres de Gabor

Alors que les descripteurs couleur, sont calculés sur les statistiques des pixels, ceux de la texture tiennent compte des relations qui existe entre les pixels se situant dans un voisinage donné. Notre moyen de représenter la texture est les filtres de Gabor.

Les filtres de Gabor sont des techniques de filtrage multi-canal permettant la description de textures localisées en fréquence et en orientation. Autrement dit, les calculs des caractéristiques sont opérés sur chacun des pixels, et plus précisément sur une région qui les entoure dite région d'intérêt. Chaque pixel contient les variations d'intensité sur une petite échelle de fréquence et d'orientation. Pour cela, les vecteurs caractéristiques utilisés pour chaque image d'entrée sont formés pas des observations statistiques sur les 18 images filtrées correspondantes à 3 fréquences et 6 orientations. Ces mesures statistiques caractérisent la répartition d'intensité en utilisant le premier et second moment μ et σ , appliquées sur une image filtrée I de dimension X et Y .

$$\mu = \frac{\sum_{i=1}^Y \sum_{j=1}^Y I(i, j)}{X * Y} \quad \sigma = \sqrt{\sum_{i=1}^Y \sum_{j=1}^Y [I(i, j) - \mu]^2}$$

2.4. Le gradient directionnel

On a choisi la direction de gradient pour mettre le point sur la présence de contours verticaux et horizontaux dans les images. En effet, l'apparition des pics pour les angles 0° , 90° et 180° dans l'histogramme du gradient directionnel montre ce type d'hypothèse. Cette caractéristique ne se retrouve pas dans les images de paysages pour les quelles l'histogramme est plat, mais elle se retrouve plutôt dans les images de villes et des monuments par exemple (Fig.2.2).

En définitive, à chaque image est associé un vecteur descripteur calculé à partir de l'histogramme directionnel de la forme :

$$\{ind_{\max 1}, val_{\max 1}, ind_{\max 2}, val_{\max 2}\}$$

qui contient :

- * Les indices des deux premiers maximums de l'histogramme de gradient directionnel. Ces indices correspondent aux valeurs des angles des deux premiers pics de l'histogramme.
- * Les valeurs de ces deux maximums correspondant à leur niveau dans l'histogramme de gradient directionnel normalisé.

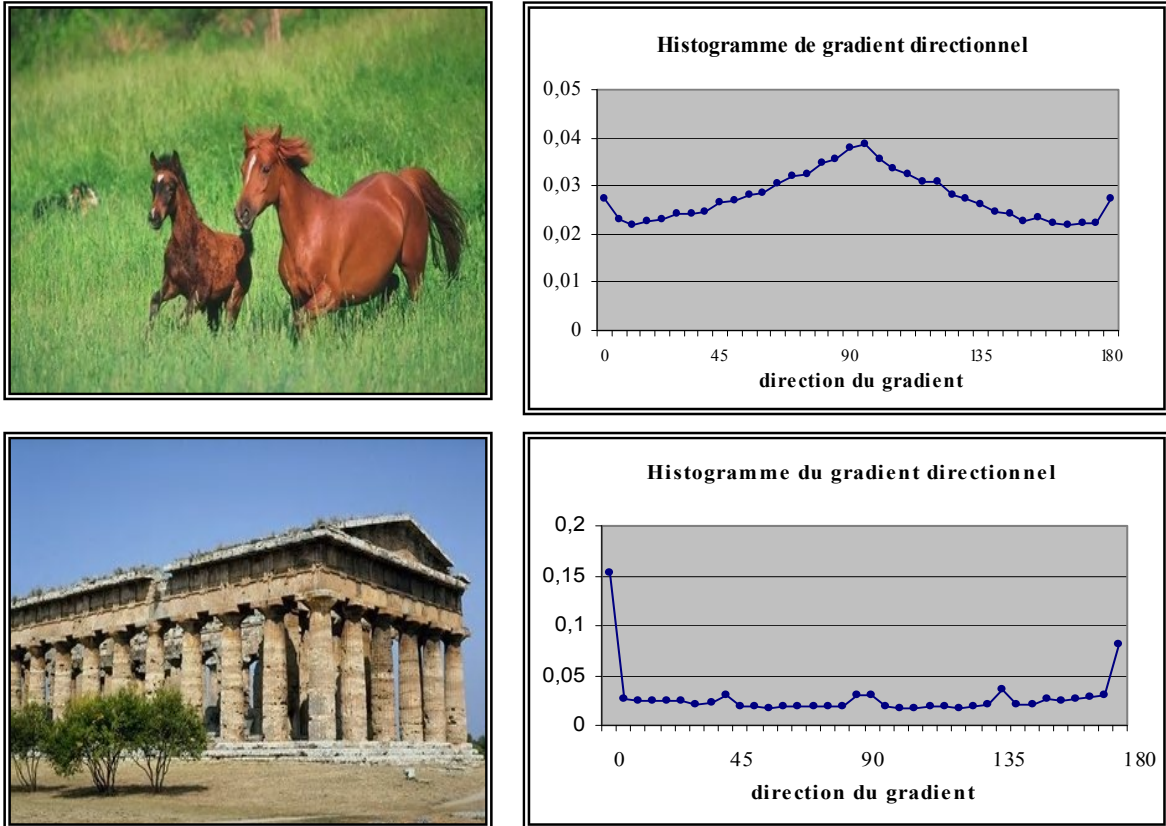


Figure 2.2 Exemples d'histogrammes de gradient directionnel.

2.5. Les mots clés

L'un des avantages les plus intéressants de la technique LSI est qu'elle permet de coupler plusieurs descripteurs de nature différente dans un même vecteur caractéristique. Jusqu'ici, nous avons prévu l'utilisation des descripteurs visuels pour les images, mais il est possible d'intégrer des descripteurs textuels dans le vecteur caractéristique lorsque la base d'images présente des images annotées (titre, auteur, description).

2.6. Conclusion

Nous avons fait dans la première partie de ce chapitre un tour d'horizon des descripteurs sur lesquels se basera notre contribution pour l'adaptation de la LSI à la problématique de recherche d'images. Cette approche d'adaptation est analysée dans le paragraphe suivant.

3. Adaptation des descripteurs

La question qui se pose à ce stade, est la manière que nous devons utiliser pour convertir la notion de fréquence d'apparition pour les mots clés aux descripteurs qui ont habituellement des valeurs numériques.

3.1. Les couleurs dominantes

Notre approche consiste à sélectionner un ensemble d'apprentissage représentatif des

bases d'images utilisées (cf. Annexe) et de traiter par la suite le comportement de ce descripteur couleur.

Nous avons dit précédemment, qu'il est largement suffisant d'utiliser 4 couleurs dominantes pour modéliser l'histogramme d'une image.

En utilisant les valeurs de L, a, b et de la cohérence spatiale SCR, nous avons formé un espace à quatre dimensions correspondants respectivement à L, a, b et SCR.

Une étude d'un ensemble de 500 images a montré la possibilité de quantifier cet espace de manière non uniforme. Cette quantification a permis de définir 40 plages de couleurs qui couvrent une bonne partie mais pas la totalité de cet espace. En effet, chacune des quatre couleurs dominantes extraites CD_i peut être attribué à l'une de ces plages de couleurs, et si ce n'est pas le cas une nouvelle plage est définie en arrondissant à la dizaine les trois composantes L_i, a_i et b_i .

Exemple : Si la couleur dominante CD_i (211, 145, 123) ne figure pas dans la liste des 40 plages identifiées, la plage de couleur définie est {L [210,220]; a [140,150]; b [120,130]}.

L'étape suivante consiste à affecter à chacune des quatre couleurs dominantes CD_i de chaque image de la base, un coefficient tenant compte du fait que la surface qu'occupe une classe de couleur par rapport à la surface totale de l'image peut avoir une signification. Ce coefficient est calculé à partir de la valeur du pourcentage P_i de la couleur dominante et du rang i de cette couleur dans l'ensemble des quatre couleurs comme suit :

$$coef = P_i * (4 - i + 1)$$

A l'issue de cette étape, une matrice, formée par les plages de couleurs ayant servi à la description des images pour les lignes et des images de la base pour les colonnes, est construite. Cette matrice est ensuite remplie en fonction des poids attribués à chaque plage. En effet, si la plage i contient une des quatre couleurs dominantes d'une image j alors la case (i,j) est remplie avec le coefficient correspondant à la couleur dominante. Sinon elle est remplie par un 0.

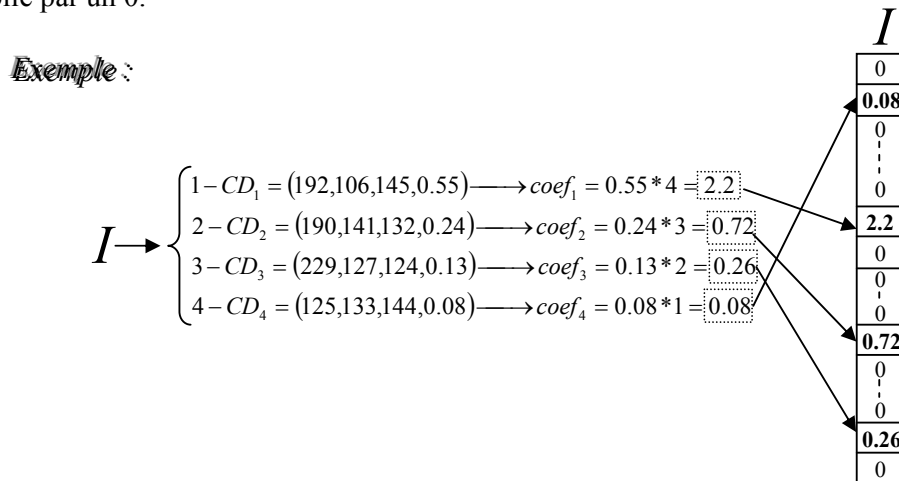


Figure 2.3 Exemple d'adaptation des couleurs dominantes pour une image I.

3.2. Le gradient couleur

Notre méthodologie d'adaptation du gradient couleur à la LSI consiste à quantifier l'histogramme normalisé du gradient couleur sur 36 valeurs, à savoir 9 pour l'axe des

abscisses (valeur du gradient) $\{[0,30[, [30,60[, [60,90[, [90,120[, [120,150[, [150,180[, [180,210[, [210,240[, [240,255]\}$ et 4 pour l'axe des ordonnées (cardinal) $\{[0,0.01[, [0.01,0.05[, [0.05,0.07[, [0.07,+0.07]\}$. Par conséquent, le nombre de caractéristiques liées au gradient couleur est de l'ordre de 36. Ce choix de quantification est déterminé d'une manière expérimentale sur notre base de test de 500 images de sorte que chaque combinaison d'un intervalle des abscisses avec un intervalle des ordonnées présente un groupe significatif d'images (Fig.2.4).

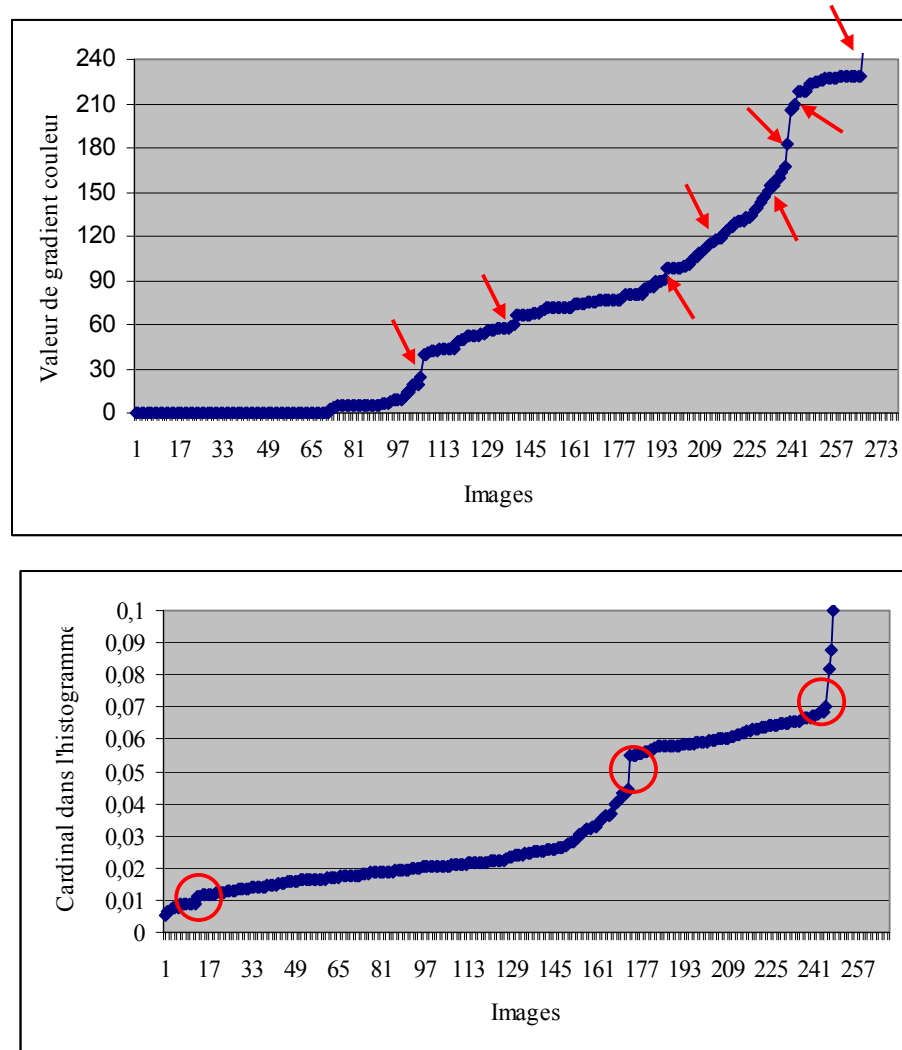


Figure 2.4 Comportement des caractéristiques du gradient couleur.

Ainsi, une nouvelle matrice, formée par les 36 caractéristiques pour les lignes et les images de la base pour les colonnes, est construite. Cette matrice est ensuite remplie en fonction des poids attribués à chaque caractéristique. En effet, pour chaque image j de la base nous calculons le maximum pour son gradient couleur, puis nous cherchons dans les 36 caractéristiques celle d'indice i qui comporte la valeur de ce maximum pour finir par l'attribution d'un coefficient égal 1 pour la case (i,j) et 0 pour les autres cases.

3.3. Les filtres de Gabor

En général, les bases d'images dont nous disposons sont hétérogènes (nature,

monuments, plages, etc.). La texture n'est pas toujours un élément majeur dans les images de telles bases. Une application des filtres de Gabor sur toute l'image serait non représentative de la classe de l'image. Nous avons donc jugé intéressant d'utiliser les filtres de Gabor sur des régions d'intérêt de l'image.

Ainsi, nous présentons dans une première étape la technique utilisée pour l'identification des régions d'intérêt, alors que l'extraction des descripteurs associée aux filtres de Gabor pour ces régions ainsi que l'adaptation de ces descripteurs à la LSI fera l'objet d'une deuxième partie.

3.3.1. Identification des régions d'intérêt

Pour déterminer les régions d'intérêt dans une image, nous avons procédé à une quantification de l'image basée sur la prise en compte des couleurs dominantes. En effet, en utilisant trois couleurs dominantes, l'image issue de la quantification présente trois classes d'intérêt correspondantes à trois classes de couleurs (Fig.2.5).

Pour extraire ces trois classes, l'image quantifiée fera l'objet d'une segmentation par seuillage sur les trois couleurs dominantes utilisées. La dernière étape consiste à identifier pour chacune des trois classes d'intérêts la plus grande fenêtre englobée ayant une taille minimale 32×32 .

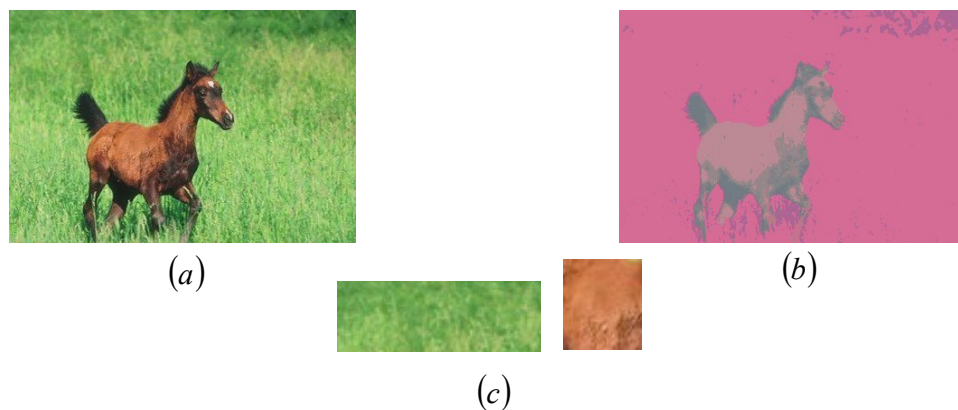


Figure 2.5 Identification des régions d'intérêt : (a) Image originale (b) Image quantifiée (c) Fenêtres englobées.

La figure 2.5 illustre les résultats d'identification des régions d'intérêts pour une image de nature (cheval). Les images (a) et (b) correspondant respectivement à l'image originale et quantifiée. Des trois classes d'intérêt de (a), nous avons pu identifier seulement deux fenêtres englobées (c) vu que l'une de ces trois classes ne peut pas contenir une fenêtre homogène de taille minimale 32×32 . Les deux fenêtres identifiées (c) présentent ainsi les textures les plus importantes dans (a).

3.3.2. Analyse de la texture pour les régions d'intérêt

Comme précisé auparavant, la texture est décrite à l'aide de la transformée en ondelettes de Gabor par une matrice formée par les 18 filtres pour les lignes et les deux caractéristiques μ et σ pour les colonnes comme suit $((\mu_{0,0}, \sigma_{0,0}), (\mu_{0,1}, \sigma_{0,1}), \dots, (\mu_{2,5}, \sigma_{2,5}))$.

En général, pour deux images (fenêtres d'intérêts dans notre cas) présentant des textures similaires, les valeurs de μ et σ calculées par les filtres de Gabor sont assez proches. Nous avons effectué une étude statistique sur un ensemble de fenêtres d'intérêts extraites de 500 images des bases utilisées (cf. Annexe), pour identifier les coefficients de Gabor qui sont les plus discriminants.

Cette étude, consiste à identifier, pour chaque classe de fenêtres d'intérêt (représentant

les textures les plus importante dans les 500 images) et pour chaque coefficient de Gabor, l'écart type intra classe des valeurs de μ et σ . Nous avons intérêt à ne sélectionner que les filtres fournissant des valeurs de μ et σ assez homogènes c'est-à-dire ayant des écarts type assez faibles pour toutes les classes.

En conséquence, cette étude a montré que les 18 coefficients de Gabor ne sont pas tous discriminants. En effet il suffit de ne garder que les valeurs correspondant aux six premiers coefficients ce qui est équivalent à six couples de caractéristiques. La figure 2.6 illustre parfaitement notre propos.

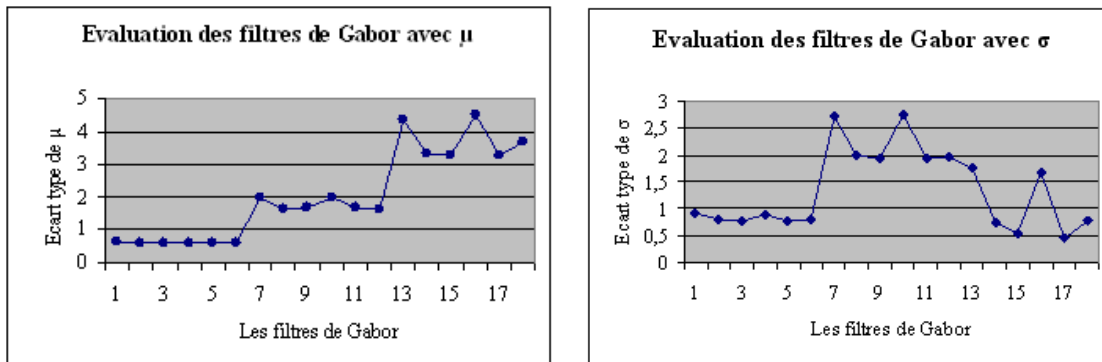


Figure 2.6 Identification des coefficients de Gabor les plus discriminants.

D'autre part, les valeurs de μ et σ calculées pour les six coefficients choisis, prennent dans la majorité des cas des valeurs entre 0 et 20. Nous avons associé ces valeurs de μ et σ à deux composantes nommées $C\mu$ et $C\sigma$ formant une base.

Ainsi pour l'adaptation de ces valeurs à la LSI, nous avons adopté, comme une première approche, après une analyse et un encadrement de ces valeurs, une quantification uniforme de la base $(C\mu, C\sigma)$ sur 121 valeurs, à savoir 11 pour $C\mu$ et 11 pour $C\sigma$. Chaque composante est partitionnée en 11 intervalles $[0,2[$, $[2,4[$, $[4,6[$, $[6,8[$, $[8,10[$, $[10,12[$, $[12,14[$, $[14,16[$, $[16,18[$, $[18,20[$, $[20,+20[$. Nous obtenons donc 121 couples $(C\mu, C\sigma)$ notés respectivement de 0 à 120 ce qui donne pour six filtres $121*6=726$ couples.

Etant donné une image j , à chaque fenêtre d'intérêt identifiée de j , nous associons un vecteur dont les composantes sont les 726 valeurs correspondantes à chacun des 726 couples (Fig.2.7). Pour calculer ces quantités, nous calculons les six paires (μ, σ) issue de l'application des six filtres de Gabor sur cette fenêtre, puis nous chercherons pour chaque paire calculée l'indice i du couple $(C\mu, C\sigma)$ correspondant parmi les 726. Enfin nous associons à la case i un coefficient égal à 1.

Pour une image donnée j , nous associons un vecteur de taille 726 résultant de la sommation vectorielle des vecteurs correspondant aux fenêtres identifiées de cette image. Ce qui nous donne pour N images une matrice de taille $726 * N$.

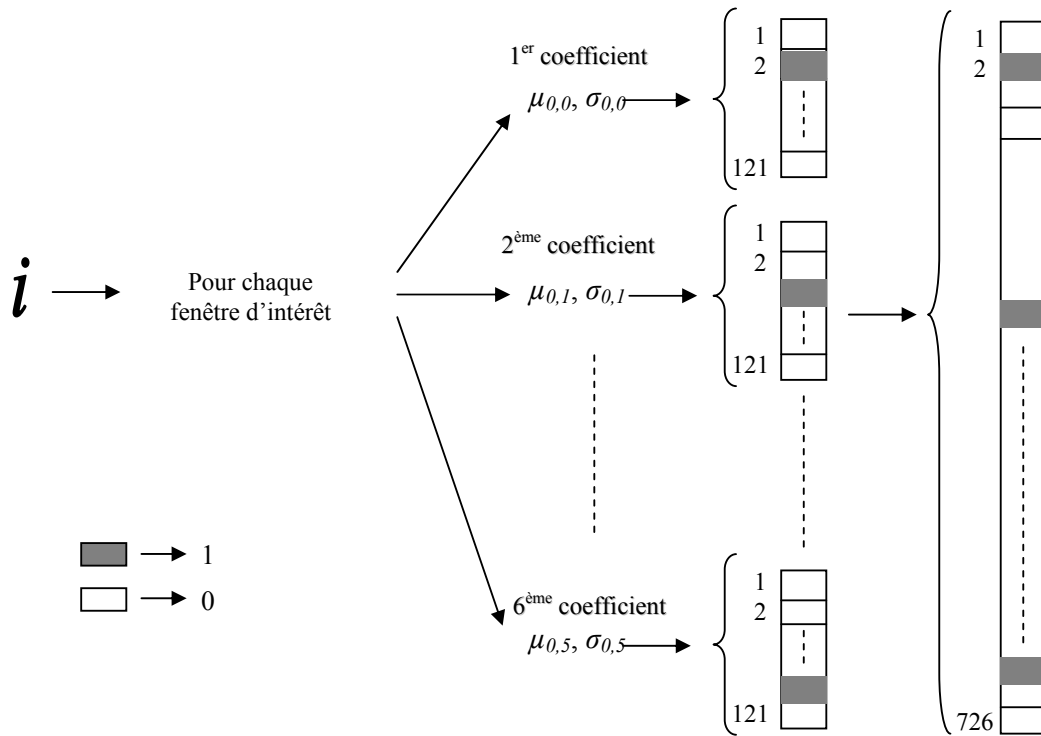


Figure 2.7 Démarche suivie pour l'adaptation des coefficients de Gabor pour une image I .

3.4. Le gradient directionnel

Le dernier descripteur de bas niveau que nous avons utilisés est lié à l'histogramme de gradient directionnel. De cet histogramme nous avons identifié 4 caractéristiques : les indices de ses deux premiers maximums (ind_{max1} et ind_{max2} : les angles $[0^\circ, 180^\circ]$) et leur valeur (val_{max1} et val_{max2} : les niveaux dans l'histogramme $[0, 0.1]$).

Dans notre approche d'adaptation de ce descripteur à la LSI, nous nous sommes contenté des indices des deux maximums et de la valeur du premier maximum. Pour décrire le gradient directionnel, nous avons formé un espace à trois dimensions une pour ind_{max1} , une pour ind_{max2} et la dernière pour val_{max1} .

Une étude de la base de test de 500 images, consistant à voir le comportement des trois caractéristiques choisies, a montré que cet espace peut être quantifier d'une manière non uniforme sur 125 valeurs, à savoir :

- * 5 pour ind_{max1} correspondants respectivement à $0^\circ, 45^\circ, 90^\circ, 135^\circ$ et 180° : en effet la figure 2.8 (a) montre que la plupart des images de la base de test ont des indices de maximum appartenant à l'ensemble $\{0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ\}$.

- * Pareil pour ind_{max2} .

- * 5 pour val_{max1} comme suit $[0, 0.03[$, $[0.03, 0.04[$, $[0.04, 0.08[$, $[0.08, 0.1[$ et $[0.1, +0.1[$: la figure 2.8 (b) présente la dispersion des valeurs du coefficient val_{max1} dans l'intervalle $[0, 0.15]$ pour les images de la base de test, cette figure montre nettement l'existence des cinq intervalles définis ci-dessous où chaque intervalle représente un ensemble d'images.

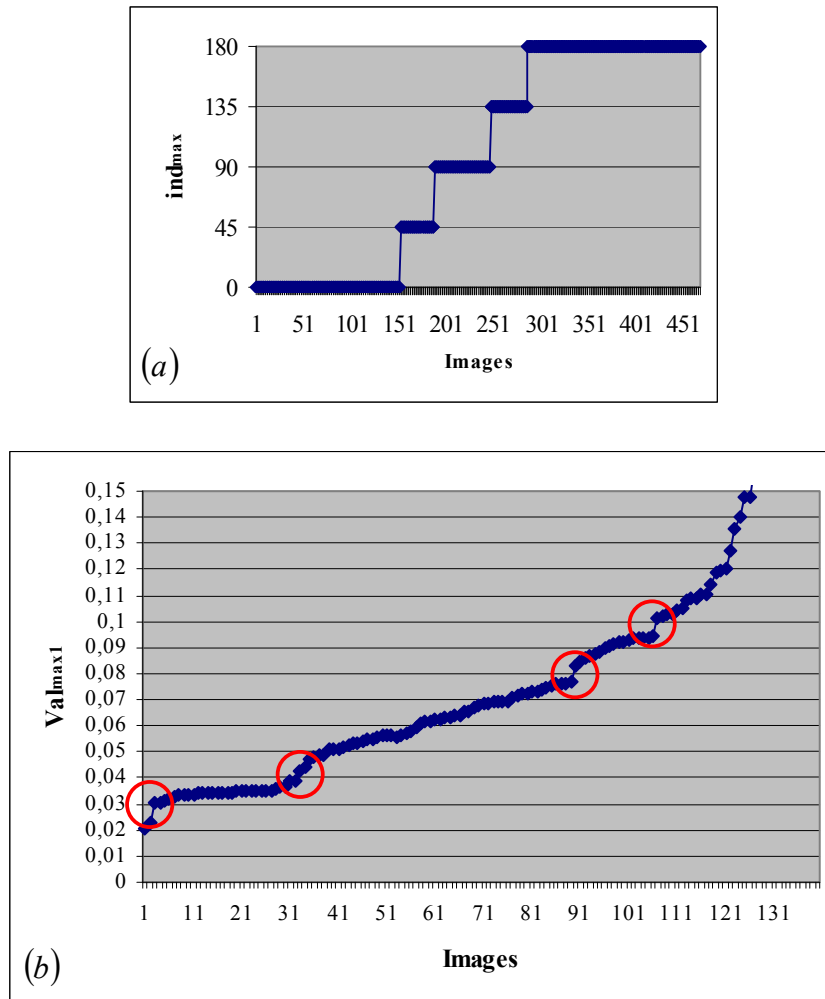


Figure 2.8 Comportement des caractéristiques du gradient directionnel.

En définitive, une nouvelle matrice, formée par les 125 caractéristiques pour les lignes et les images de la base pour les colonnes, est construite. Cette matrice est ensuite remplie en fonction des poids attribués à chaque caractéristique. En effet, pour chaque image j de la base nous calculons le triplet $(ind_{max1}, ind_{max2}, val_{max1})$, que nous chercherons dans les 125 caractéristiques celle d'indice i qui lui correspond, nous associons alors à la case (i,j) un coefficient égal à 1 et 0 pour les autres.

3.5. Les mots clés

Les annotations de chacune des images sont enregistrées dans un fichier texte pour chacune d'entre elles. Un coefficient est associé à chacun des mots pour chaque image. Ce coefficient est calculé en fonction de l'apparition du mot dans l'annotation de l'image.

Une matrice, formée des mots ayant servi à l'annotation des images pour les lignes et des images de la base pour les colonnes, est construite. Cette matrice est ensuite remplie en fonction des poids attribués à chacun des mots. Si le mot i figure dans l'annotation de l'image j alors la case (i, j) est remplie avec la valeur 1. Sinon elle est remplie par un 0.

3.6. Conclusion

Après l'analyse de tous ces descripteurs, nous avons formé cinq matrices avec le même nombre de colonnes. Ces cinq matrices vont être juxtaposées pour former une nouvelle matrice A nommée matrice Descripteurs x Images ayant comme nombre de lignes la somme des nombres de caractéristiques de chacune des cinq matrices et qui vaut $M=Nbpc+36+726+125+Nbmc$ (Fig.2.9).

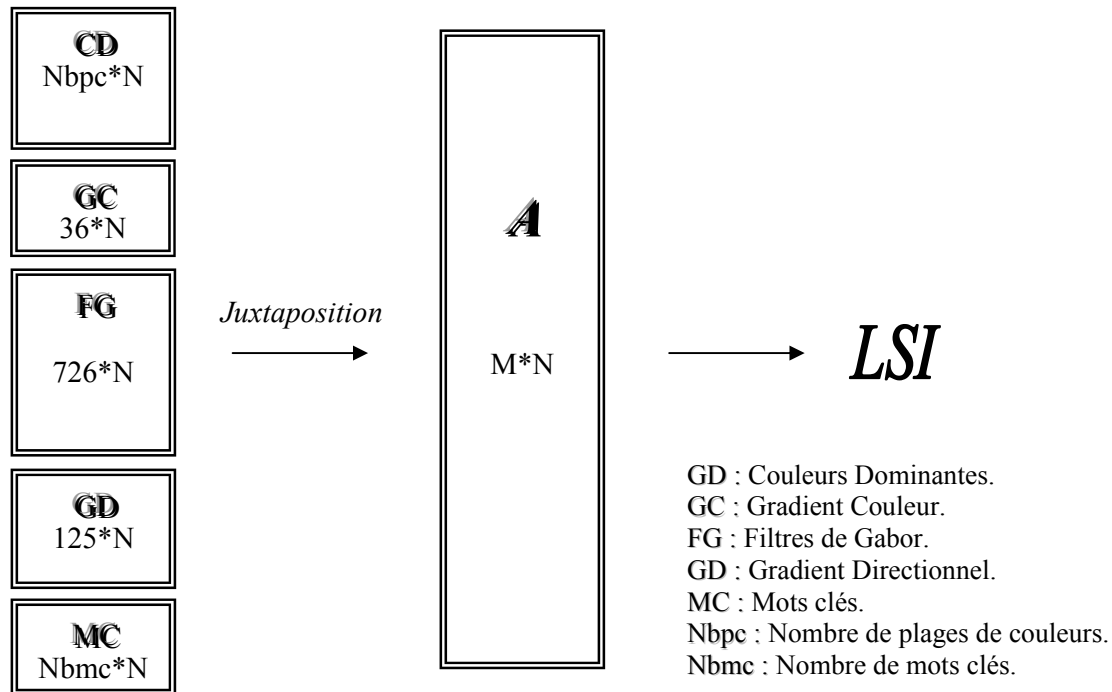


Figure 2.9 Matrice Descripteurs x Images.

Cette matrice contient trop de redondance inutile vu le grand nombre de 0 qu'elle peut contenir. En effet, elle peut comporter des lignes ne contenant que des 0. Pour réduire cette redondance, nous avons proposé d'appliquer une nouvelle opération sur cette matrice qui consiste à sélectionner dans cette matrice les lignes comportant plus de deux cases non nuls.

La nouvelle matrice construite remplace la matrice Termes x Documents habituellement utilisée avec la LSI.

4. Conclusion

Nous avons présenté dans ce chapitre notre approche proposée pour l'adaptation de la technique LSI à la problématique de recherche dans les bases d'images. Elle est constituée essentiellement de deux étapes principales : le choix des descripteurs et leur adaptation.

Les descripteurs que nous avons choisis sont peu nombreux vu les contraintes de temps mais ils sont bien porteurs de la sémantique. Nous avons utilisés comme descripteur les couleurs dominantes, le maximum de gradient de couleur, l'histogramme du gradient directionnel, les filtres de Gabor et les mots clés.

L'étape d'adaptation des descripteurs à la LSI fait appel à une approche bien connue dans le domaine d'imagerie : *la quantification*. Cette approche nous a permis de traduire les valeurs numériques des descripteurs en notion de fréquence d'apparition pour former une matrice Descripteurs x Images sur laquelle nous allons appliquer la LSI.

Chapitre 3. Tests et évaluation

1. Introduction

Nous présentons dans ce chapitre notre système intégrant les fonctionnalités d'indexation, de recherche et de classification. Nous détaillons dans une première partie, ces fonctionnalités, ensuite nous présentons quelques uns des résultats obtenus lors d'une série de tests sur des parties de bases d'images décrites en Annexe. Ceci nous permettra par la suite d'évaluer les performances de notre approche d'adaptation de la LSI aux images.

2. Fonctionnalités

Notre système offre trois fonctionnalités principales, l'indexation, la recherche et la classification :

2.1. L'indexation

L'indexation consiste en l'application de l'indexation en sémantique latente LSI sur la matrice Descripteurs x Images adaptée à ce type de technique. A partir de cette matrice, la LSI permet d'extraire pour chaque image de la base, un vecteur descripteur qui capture la structure sémantique latente fondamentale de l'image. L'indexation est exécuté une seule fois en hors ligne selon les étapes suivantes :

- * Calcul des descripteurs.
- * Adaptation des descripteurs dans la matrice Descripteurs x Images.
- * Décomposition en valeurs singulières SVD.
- * Réduction de l'espace vectoriel de représentation des images à k dimensions : k est une valeur définie par le système après de nombreux tests, mais nous pouvons également laisser à l'utilisateur la possibilité de refaire l'étape de réduction en ligne pour déterminer la valeur de k la plus adéquate avec sa base d'image.

2.2. La recherche

Elle est lancée à partir d'une image clé par l'utilisateur. Le moteur de recherche mesure alors le cosinus entre les vecteurs descripteurs de l'image clé et ceux des images de la base. Les images faisant un angle de cosinus supérieur à un seuil donné sont considérés comme les plus similaires à l'image requête et seront retournées à l'utilisateur.

2.3. La classification

Après application de la LSI, chaque image de la base est représentée par un index. La classification de ces images est réalisée en utilisant l'algorithme du K-MeansCosinus [WAR 03] qui est une adaptation de l'algorithme des K-Means à la distance cosinus.

C'est un algorithme de classification non supervisée très utilisé en matière de clustering. Il nécessite la connaissance du nombre « N » de clusters a priori. En effet, nous constituons une base de référence (aussi appelée base d'apprentissage) qui servira de modèle. Cette base contient « N » éléments qui définiront les N premiers centroïdes de l'algorithme. Le choix de ces éléments peut se faire au hasard, ou suivant différentes règles (choisir les N observations les plus représentatives de l'ensemble des classes, ...). Il faut prendre grand soin lors de la

construction de cette base car sa construction est un facteur important de la qualité des résultats obtenus.

Dans notre cas, nous identifions les N centroïdes en fonction d'une pré-classification de la base d'image. En effet nous commençons par la sélection d'une image de toute la base et puis l'extraction des images qui lui sont similaires selon une distance donnée par la valeur du cosinus entre leur vecteur descripteur. Cette étape est reproduite plusieurs sur le reste de la base correspondant à l'itération précédente (c'est-à-dire les images non classées) jusqu'à ce que toutes les images soient classées. Cette approche fournit les N classes les plus représentatives et pour avoir les N centroïdes de l'algorithme il suffit de sélectionner une image de chaque classe (la première image par exemple).

Les centroïdes étant calculés, nous calculons pour chaque image sa distance par rapport à tous ces centroïdes. Chaque image sera affectée à son centre le plus proche. Ainsi nous avons défini des classes, entités regroupant un centre et toutes les images qui lui ont été affecté. La distance utilisée pour les K-MeansCosinus est donnée par la valeur du cosinus de l'angle formé par les vecteurs descripteurs.

L'étape suivante consiste à calculer pour chaque classe, son centre de gravité qui devient son nouveau centroïde.

Les deux dernières étapes de réallocation et de re-centrage sont répétées jusqu'à ce que la distance entre deux centroïdes, correspondant à deux itérations successives, soit nulle.

3. Evaluation de la recherche

L'évaluation des systèmes de recherche d'images n'est pas aisée, principalement en raison de la perception subjective du contenu de celles-ci, alors que les autres disciplines de recherche d'information ont réussi à développer des critères d'évaluation sophistiqués et efficaces. C'est d'ailleurs à ces disciplines que nous les avons empruntés les paramètres Recall (Rappel) et Precision (Précision) qui sont les deux paramètres les plus couramment utilisés dans les systèmes de recherche d'images.

Pour mesurer ces paramètres, nous supposons que la base d'images est constituée de classes d'images disjointes dont nous connaissons les cardinaux respectifs. Toutes les images de la base sont prises successivement en tant qu'image clé et nous observons les images retournées jusqu'à un rang r variable.

Une autre approche d'évaluation est également proposée dans MPEG-7, consistant à considérer le nombre d'images positives retournées pour un rang fixe, égal à 2 fois le cardinal de la classe de l'image clé.

3.1. Le paramètre « Recall »

Ce paramètre évalue le rang r pour lequel le système est capable de fournir toutes le images appartenant à la classe de l'image clé. Ceci revient à une mesure de rapidité. Le Rappel R_r est donné pour le rang r par :

$$R_r = \frac{A_r}{A_r + C_r} \quad (3)$$

avec

$$A_r = \sum_{i=0}^{r-1} V_i \quad \text{et} \quad C_r = \sum_{i=0}^{N-1} V_i - A_r$$

où $i \in [1, r]$ représente le rang d'une image, $V_i \in \{0,1\}$ selon que l'image de rang i appartient à la classe (1) ou non (0) et N le cardinal de la classe de l'image clé.

En d'autres termes, R_r est égal au rapport entre le nombre d'images positives retournées pour le rang r et le nombre total d'images positives dans la base. Nous traçons alors la courbe $R_r = f(r)$. Plus vite cette courbe tendra vers 1 et plus le système sera performant.

3.2. Le paramètre « Precision »

Comme son nom l'indique, il mesure la précision du système. Il est donné par :

$$P_r = \frac{A_r}{A_r + B_r}$$

avec

$$B_r = \sum_{i=0}^{r-1} (1 - V_i)$$

où P_k représente en fait le rapport entre le nombre d'images positives rendues au rang r et ce même rang r . Avec ce paramètre, plus le système est précis et plus la courbe $P_r = g(r)$ décroît lentement.

3.3. Evaluation sur les bases d'images

Nous avons effectués nos tests sur 2 types de bases : d'une part sur la base Simplicity [SIM] qui est une base généraliste présentant une très grande diversité d'images, et d'autre part sur la base MPEG-7, avec l'idée de pouvoir se comparer à d'autres approches de recherche. De part son hétérogénéité, la base MPEG-7 devrait présenter plus de difficulté pour la recherche, les classes n'étant pas suffisamment disjointes. Nous présentons avec plus de détails ces bases ainsi que des échantillons de leurs classes dans l'annexe.

Nous nous sommes basés sur les paramètres Rappel et Précision décrits précédemment pour les évaluations.

3.3.1. Evaluation sur la Base Simplicity

Nous avons réalisé notre expérimentation sur une partie de la base Simplicity en considérant cinq classes composées chacune de 30 images tirées aléatoirement de la base (Fig.3.1).

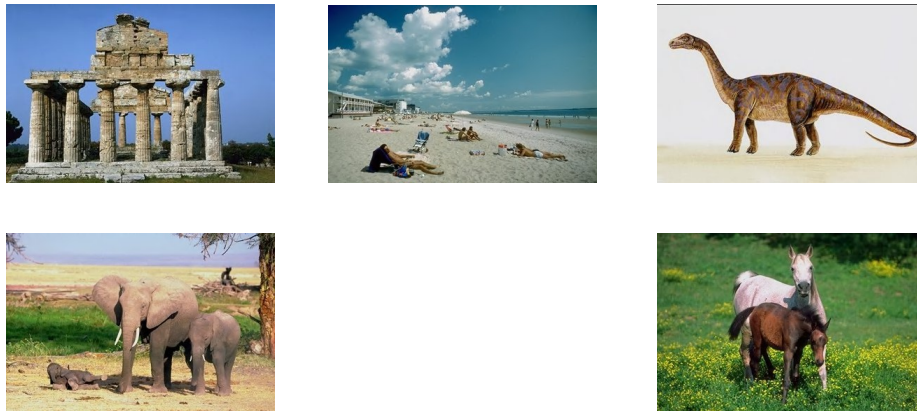


Figure 3.1 Classes d'images considérées dans la base Simplicity.

Comme indiqué précédemment, le choix du nombre de dimensions k dont nous allons tronquer les matrices obtenues par SVD est un problème important. Il n'y a pas de consensus

général pour déterminer ce nombre. Ainsi, sa détermination est expérimentale et dépend du corpus de documents utilisé. Pour cela nous avons cherché à estimer ce nombre k pour effectuer une recherche optimale. Nous avons effectués plusieurs tests en modifiant à chaque fois la valeur de k et nous avons pu constater qu'à partir d'une valeur de $k=7$ le système est stable.

Le principe des tests consiste, à faire autant de requête de recherche qu'il y'a d'images appartenant à l'une des classes de la base (c'est-à-dire 150), en considérant comme image clé chacune des images appartenant aux classes. On comptabilise alors, pour chaque requête, les images appartenant à la même classe que l'image clé (images positives) parmi les r premières images retournées par le système. r représentera ainsi le rang. Pour chaque rang k , nous déterminons le nombre moyen d'images positives retournées par le système (noté A_r dans (3)). La paramètre « Rappel » est alors donné par le rapport entre A_r et le cardinale des classes qui vaut 30, alors que le paramètre « Précision » est donné par le rapport entre A_r et le rang r .

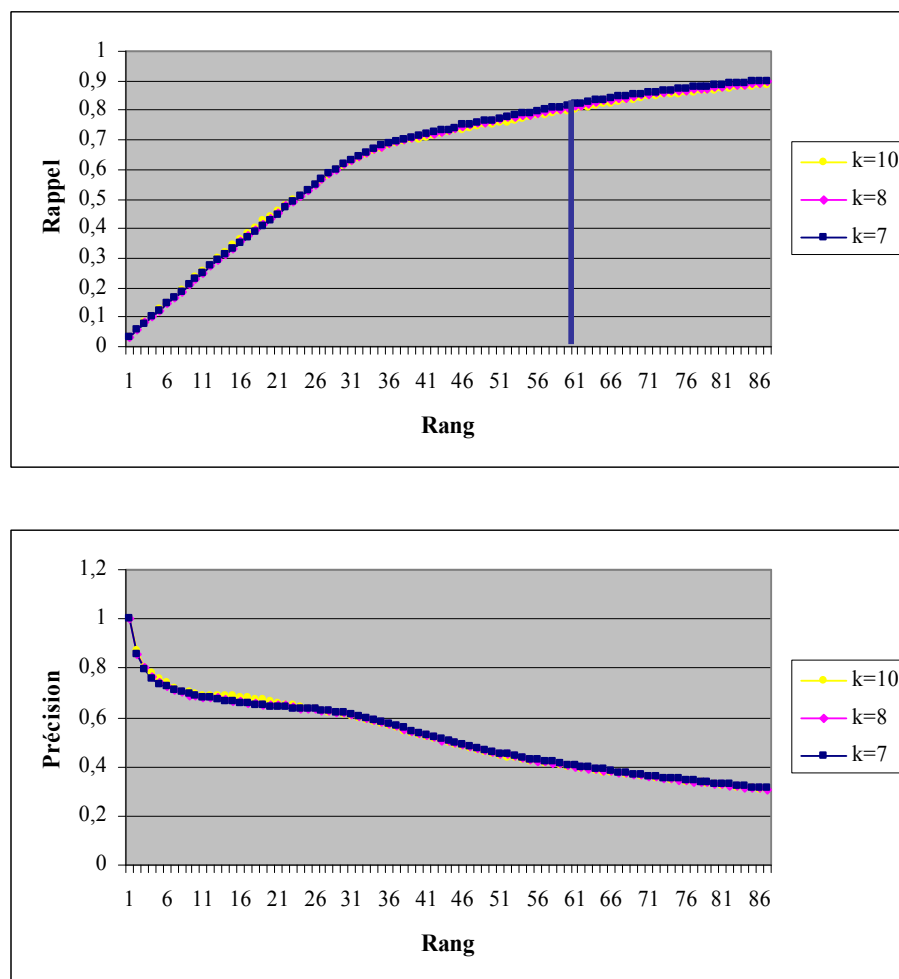


Figure 3.2 Evaluation de la base Simplicity.

La figure 3.2 donne les courbes de Rappel et de Précision fournies par le moteur de recherche pour chacun des 3 valeurs de k : 7,8 et 10, pour la base Simplicity. Ces courbes illustrent parfaitement notre propos sur le fait qu'à partir de $k=7$ (courbes en bleu) le système devient stable.

En plus, les courbes de Rappel montrent que pour un rang égal au double du cardinal de la classe de l'image c'est-à-dire 60 (trait en bleu), le système retourne plus de 80% des images positives. C'est-à-dire que pour une image clé donnée de la base Simplicity de cardinal 30, sur 60 images retournées par le système 24 sont des images appartenant à la même classe que l'image clé.

3.3.2. Evaluation sur la Base MPEG-7

Nous avons cherché à vérifier la validité des résultats précédents en utilisant une base complètement différente. Nous avons considérés dans la base MPEG-7, 3 classes composées chacune de 35 images tirées aléatoirement de la base (Fig.3.3).



Figure 3.3 Classes d'images considérées dans la base MPEG-7.

Nous allons réaliser notre expérimentation sur cette base en se basant toujours sur les mêmes principes que pour la base Simplicity.

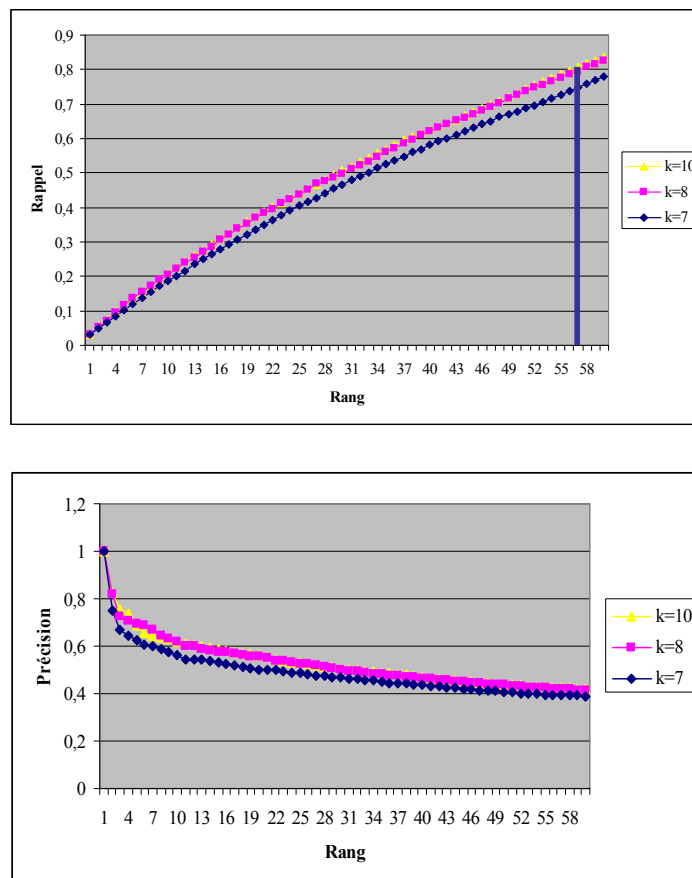


Figure 3.4 Evaluation de la base MPEG-7.

La figure 3.4 montre la performance du système pour 3 valeurs de k : 7, 8 et 10. Il apparaît clairement qu'une valeur de $k=8$ (courbes en rose) est largement suffisante pour que la recherche soit optimale. Les courbes de Rappel dans 3.3 viennent confirmer le résultat illustré par les courbes 3.2 selon lequel pour un rang égal au double du cardinal de la classe de l'image ou même moins (57: trait en bleu), le système retourne plus de 80% des images positives.

4. Tests de comparaison avec l'approche Pecenovic

Nous avons, à travers ce test, comparé les résultats de recherche rendus par notre approche d'adaptation et par celle proposée par Pecenovic présentée dans le chapitre 1.2. Nous avons pour cela travaillé sur la base Simplicity. Nous avons considéré pour Pecenovic seulement le descripteur couleur défini par l'histogramme 3D couleur Lab. En effet, l'approche consiste à calculer pour la base de 150 images sélectionnées de la base Simplicity, la moyenne et l'écart type des trois histogrammes couleur. A chaque image de la base, est associé un vecteur caractéristique de taille $256*3=768$. Pour remplir ce vecteur, nous appliquons pour chacune des 768 caractéristiques le théorème de la limite centrale comme suit :

$$V[i] = \begin{cases} \frac{val_{car}(i) - \mu_i}{\sigma_i} & \text{if } val_{car}(i) > \mu_i \\ 0 & \text{else} \end{cases} \quad \text{où } \mu_i \text{ et } \sigma_i \text{ correspondent respectivement à la moyenne et l'écart type de la caractéristique d'indice } i \text{ parmi les 768.}$$

Ainsi, pour la base de 150 images, nous obtenons une matrice Descripteurs x Images de taille $768*150$ sur laquelle nous appliquons la LSI.

Nous avons cherché à comparer cette approche Pecenovic pour l'adaptation de la LSI aux images à notre approche. Pour cela, nous avons effectué deux tests : un test où que le descripteur couleurs dominantes a été considéré pour notre approche et l'autre en utilisant tous les descripteurs présentés au chapitre 2 sans les mots clés (base d'images non annotées).

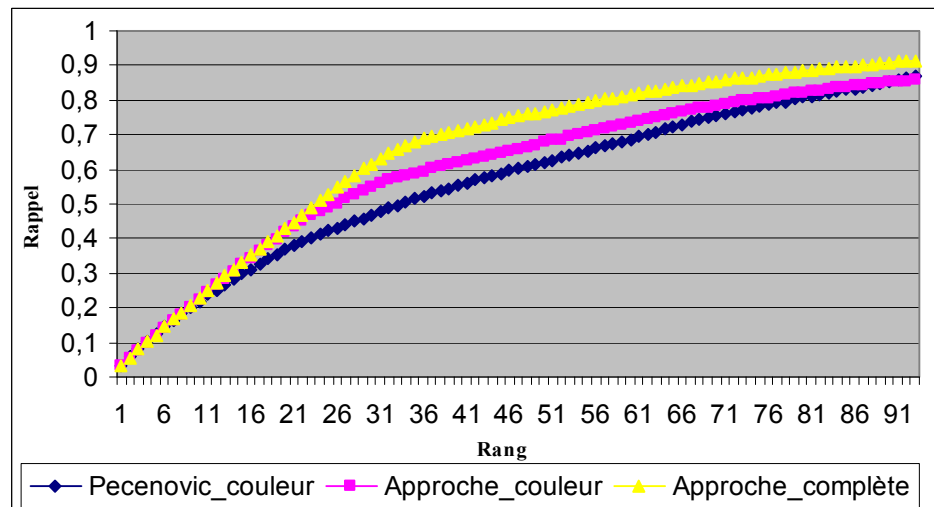


Figure 3.5 Comparaison de l'approche Pecenovic à notre approche.

La figure 3.5 présente les courbes de Rappel pour les 3 méthodes appliquées {Pecenovic (courbe en bleu), notre approche en considérant que le descripteur couleur (courbe en rose) et en considérant tous les descripteurs (courbe en jaune)}.

Les résultats entre les 2 premières méthodes sont relativement proches avec un léger avantage pour notre approche. Alors qu'en appliquant notre approche en utilisant tous les descripteurs (courbe en jaune), les résultats sont de plus en plus avantageux. Toutefois il convient de noter que la taille du vecteur descripteur pour l'approche Pecenovic est de l'ordre de 768 alors que pour notre approche est de l'ordre de 237 (après réduction de la redondance dans la matrice Descripteurs x Images), cela signifie qu'en prenant plus de descripteurs les résultats vont être plus intéressants.

5. Quelques résultats de classification

Nous allons présenter dans ce paragraphe quelques résultats de classification sur la partie de Simplicity présentée en 3.3.1.

Après lancement de la LSI avec une valeur de $k=7$, nous avons passé à l'application de notre approche de classification basée sur l'algorithme de K-MeansCosinus. Nous nous sommes retrouvés avec 12 groupes. 8 groupes se retrouvent plus distincts (Fig.3.6). Les autres groupes présentent quelques exceptions (Fig.3.7) : On y trouve quelques éléphants chez les chevaux mais il n'y a rien d'illogique car ce sont des images d'éléphants entourés de verdure tout comme les chevaux. Quelques images d'éléphants se retrouvent aussi avec de la plage ou avec des monuments mais là encore ce n'est pas trop grave car ce sont des images où l'eau ou le ciel ou le sable prend une place importante, c'est le même cas parfois pour quelques images de monuments qui se trouvent avec des images de plage.

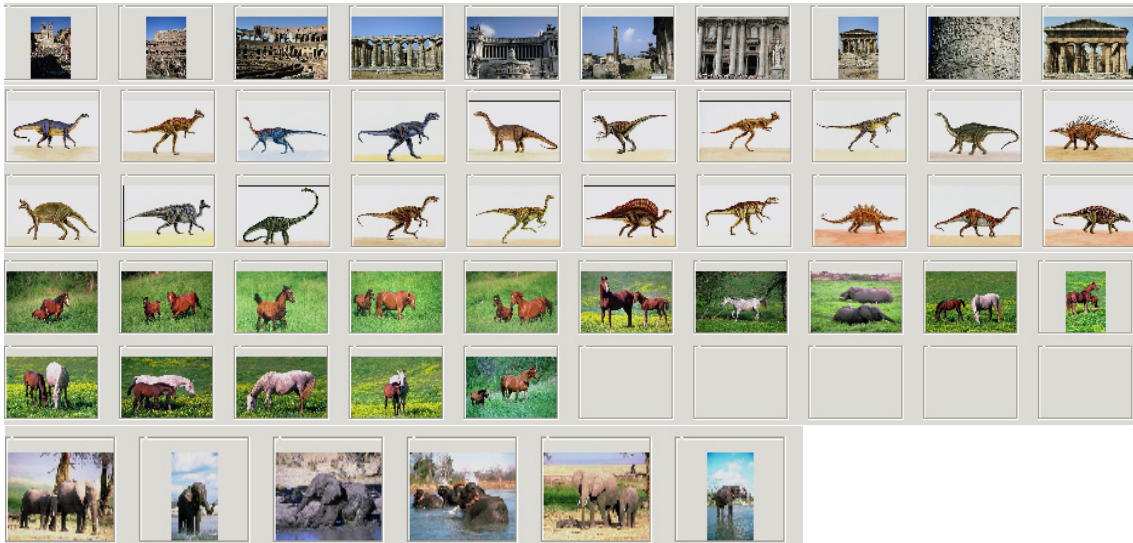


Figure 3.6 Exemples de groupes distincts retournés par l'approche de classification.

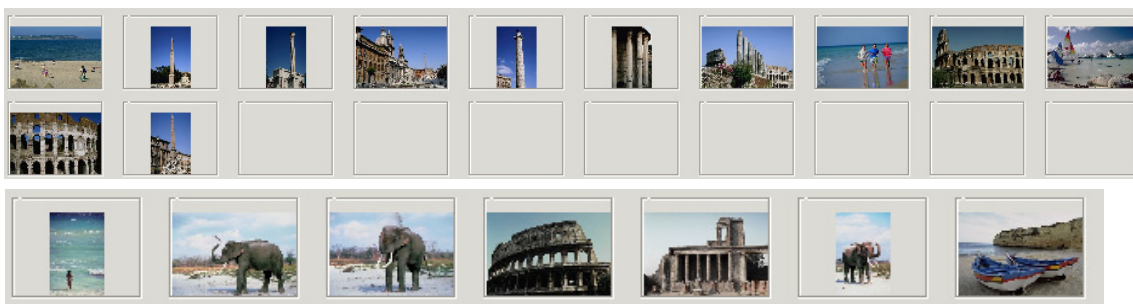


Figure 3.7 Exemples d'exceptions dans l'approche de classification.

En résumé, nous trouvons parfois quelques exceptions mais les résultats sont encourageants.

Conclusions et perspectives

Le problème le plus fréquent dans la problématique de recherche et de classification d'images est la présence d'un *gap sémantique* représenté par la différence entre la description des images et la sémantique portées par elles. Pour remédier à ce problème, nous avons essayé, tout le long de ce stage, de bénéficier des avantages des techniques d'indexation textuelles dans la classification et la recherche sémantique et montrer ainsi la faisabilité de leur adaptation aux images.

Après avoir fait un tour d'horizon des techniques d'indexation textuelles existantes, nous avons opté pour la technique d'*indexation en sémantique latente LSI*.

Notre approche pour l'adaptation de la LSI aux images, proposée dans ce rapport, se distingue par son originalité et sa simplicité : elle s'appuie sur une technique bien connue en imagerie : *la quantification*.

De nombreux tests ont permis d'évaluer les performances de cette adaptation pour la recherche et la classification d'images. Ils ont mis en évidence la faisabilité de l'adaptation. Cependant, ils ont souligné les limites de cette approche face au manque de l'utilisation de descripteurs « sémantiques ». En effet, comme nous l'avons indiqué précédemment, pour traiter le problème de manque de la sémantique dans la recherche, il faut prendre plusieurs descripteurs.

Plusieurs idées ont été envisagées pendant ce stage. Nous avons voulu essayer une exécution de la plupart de ces idées. De toute façon, les idées qui suivent méritent, à notre avis, une attention plus particulière pour une amélioration de l'approche.

D'une part, le choix de la manière de quantification repose sur l'interprétation du comportement des descripteurs dans une base de test, des améliorations et des évolutions sont envisageables pour l'automatisation de la manière de quantification en proposant une étude plus spécifique et plus efficace qui peut être basée sur un apprentissage.

D'autre part, l'aspect forme a été totalement ignoré dans notre choix de descripteurs. Ainsi, des approches comme la représentation ICV (*Images Context Vectors*) [GAL and JHO 95] ou l'utilisation des points d'intérêts peuvent être intégrée pour ajouter plus d'efficacité à notre système.

D'un autre côté, la LSI est une technique efficace pour extraire des descripteurs qui capture la proximité sémantique dans les documents. Dans [SHI et al 04], Shima montre que le sous-espace choisi par cette méthode, après l'étape de réduction, peut ne pas être le plus approprié pour la classification des documents (les images dans notre cas). En effet, les descripteurs sont choisis selon leur variance et non pas selon leur puissance de classification. Afin de choisir les descripteurs les plus adéquats pour la classification, Shima propose d'appliquer une méthode d'organisation des descripteurs basée sur la SVM (Support Vector Machines). Les résultats expérimentaux trouvés suggèrent que la méthode améliore la classification avec une représentation considérablement plus compacte. De telles approches peuvent être utilisées pour améliorer la démarche proposée dans ce travail.

Cette annexe a pour but de présenter de façon succincte l'espace couleur Lab que nous avons utilisé dans l'identification des couleurs dominantes ainsi que les bases d'images utilisées dans notre approche et nos tests (Simplicity et MPEG-7).

1. L'espace Lab

Les espaces couleurs classiques tels que (RGB, CMJ...) sont issus d'une approche purement physique et ne prennent pas en compte les aspects psychologiques de la perception humaine. D'autres espaces ont été mis au point soit pour se rapprocher de la vision humaine soit pour distinguer la composante luminance des composantes chrominances décorrélant ainsi l'espace couleur. Parmi ces espaces, l'espace Lab tend à uniformiser la perception de la couleur et est utile en analyse d'image car il permet de calculer une distance entre couleurs proportionnelle à la différence couleur entre celles-ci.

Cet espace couleur a été proposé par la Commission Internationale de l'Eclairage CIE en 1976. Dans cet espace, L représente la luminance et a et b définissent la composante chromatique de la couleur. Dans ce système, la métrique euclidienne est significative des différences de couleur perçues, ce qui permet de mesurer les variations de couleurs en rapport avec la perception humaine.

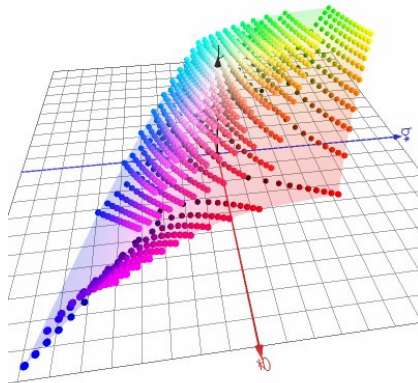


Figure 4.1 Espace Lab.

Le passage RGB vers Lab nécessite au préalable de passer par l'espace XYZ.

$$\begin{cases} L = 116f\left(\frac{Y}{Y_0}\right) - 16 \\ a = 500 \left[f\left(\frac{X}{X_0}\right) - f\left(\frac{Y}{Y_0}\right) \right] \\ b = 200 \left[f\left(\frac{Y}{Y_0}\right) - f\left(\frac{Z}{Z_0}\right) \right] \end{cases} \quad \text{avec } f(x) = \begin{cases} x^{\frac{1}{3}} & \text{si } x > 0.008856 \\ 7.787x + \frac{16}{161} & \text{si } x \leq 0.008856 \end{cases}$$

où les trois composantes trichromatiques (X_0, Y_0, Z_0) représente le blanc de référence dans l'espace XYZ.

2. Les bases d'images

2.1. Simlicity

La base Simplicity (Semantics-sensitive Integrated Matching for Picture LIbraries) [SIM] est une base composée de 1000 images de la collection *COREL CD-ROM*.

Ces images ne sont pas annotées et sont réparties sur 10 classes de 100 images chacune.

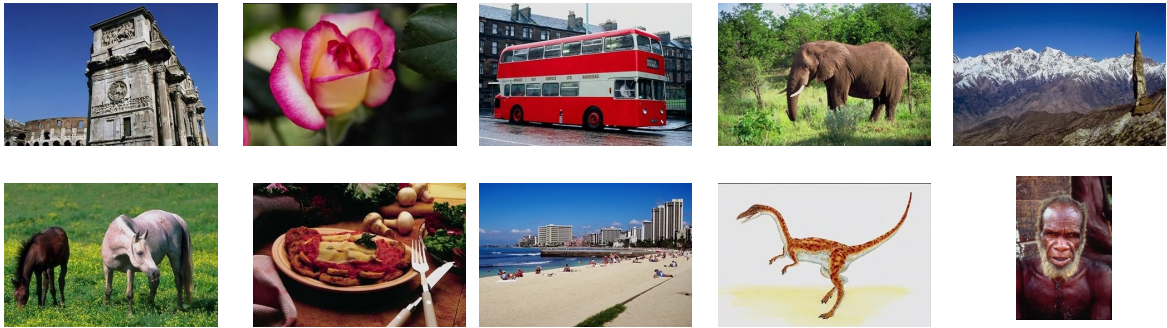


Figure 4.2 Base Simplicity.

2.2. MPEG-7

La base MPEG-7 utilisé est une base composée par 300 images réparties sur 3 classes. Les images de la base ne sont pas annotées.



Figure 4.3 Base MPEG-7.

Références

- [CHA 88] J. Chaumier, « Le traitement linguistique de l'information », Paris : Entreprise moderne d'édition, 1988. 3eme édition mise à jour et augmentée. 186 p. (Systèmes d'information et nouvelles technologies).
- [DEE et al 90] S. Deerwester et al, « Indexing by Latent Semantic Analysis ». Journal of the American Society for Information Science, vol 41, pp. 391-407, 1990.
- [DUM 93] S. T. Dumais, « LSI meets TREC: A status report », In: D. Harman (Ed.), The First Text REtrieval Conference (TREC1), National Institute of Standards and Technology Special Publication 500-207 , pp. 137-152, 1993.
- [DUM 94] S. T. Dumais, « Latent Semantic Indexing (LSI) and TREC-2 », In: D. Harman (Ed.), The Second Text REtrieval Conference (TREC2), National Institute of Standards and Technology Special Publication 500-215, pp. 105-116, 1994.
- [GAL and JHO 95] Stephan I. Gallant and Michael F. Jhonston, « Image Retrieval using Image Context Vectors : first results», in Storage and Retrieval for Image and Video Databases III, SPIE Vol. 2420 :82-94, 1995.
- [IDR et al 02] K. Idrissi, J. Ricard and A. Baskurt. « An objective performance evaluation tool for color based image retrieval systems ». ICIP-2002, Rochester, USA, September 2002
- [IDR 03] K. Idrissi, « Segmentation et description de la couleur pour l'indexation, la navigation visuelle et la recherche dans les bases d'images » thèse de doctorat, Université Claude Bernard de Lyon, Mars 2003.
- [KM 02] KM Center, « Gestion des connaissances, outils et applications du knowledge management », <http://kmcenter.free.fr/technologies/IA/rn.htm> 2003.
- [PEC 97] Z.Pecenovic, « Image retrieval using Latent Semantic Indexing » Bs.C. thesis, EPFL, Lausanne, 1997.
- [PEC et al 98] Pecenovnic, M.N. Do, S. Ayer and M. Vetterli, « New methods for image retrieval », Proc. of ICPS'98 Congress on Exploring New Tracks in Imaging, pp. 242-246, Antwerp, Belgium, September 1998.
- [Praks et al 03-1] P. Praks, J. Dvorsky and V. Snasel, « Latent Semantic Indexing for Image Retrieval Systems », In SIAM Linear Algebra Proceedings 2003, Philadelphia: International Linear Algebra Society (ILAS), July 2003.
- [Praks et al 03-2] P. Praks, J. Dvorsky, V. Snasel and J. Cernohorsky, « On SVD-free Latent Semantic indexing for image retrieval for application in a hard industrial environment », In Proceedings of the IEEE International Conference on Industrial Technology - ICIT, pp. 466-471 Maribor, Slovenia, December 2003.

[SHI et al 04] K. Shima, M. Todoriki and A. Suzuki, « SVM-based feature selection of latent semantic features », Pattern Recognition Letters Volume 25, Issue 9, pp. 1051-1057, April 2004.

[SIM] SIMPLiCity « Semantics-sensitive Integrated Matching for Picture Libraries », <http://wang.ist.psu.edu/IMAGE/>.

[SOW 84] J.F. SOWA, « Conceptual Structures: Information Processing in Mind and Machine », London: Addison-Wesley, 1984.

[ZHA et al 00] R. Zhao, W.I. Grosky, « From features to semantics: some preliminary results », In Proc. of International Conference on Multimedia & Expo, pp. 679-682, New York, USA, July, 2000.

[ZWE et BOU 97] P. Zweigenbaum, J. Bouaud, « Construction d'une représentation sémantique en Graphes Conceptuels à partir d'une analyse LFG ». In Damien Genthial, Actes de TALN'97, Grenoble, pp. 30-39.

[WAR 03] D. Wartel, « GALILEI : Algorithmes de Clustering », Technical Report, Université Libre de Bruxelles, May 2003.

RESUME

La recherche par le contenu dans les bases d'images est d'actualité et de nombreux systèmes de recherche ont vu le jour, indexant généralement les bases à l'aide de descripteurs bas niveau (couleur, texture et forme). Les moteurs de recherche calculent alors les distances entre descripteurs pour trouver les images de la base les plus similaires de l'image clef. Ce travail de DEA essaye de bénéficier des avantages des méthodes classiques de recherche d'information textuelle pour les adapter à la problématique de recherche dans les bases d'images. Ceci consistera d'une part, à faire le tour des méthodes textuelles existantes, et d'autre part à proposer un schéma d'adaptation incluant le choix des descripteurs, leur extraction puis leur utilisation avec la méthode jugée la plus adéquate.

Mots clés : approche textuelle, descripteurs de bas niveau, couleur, texture, forme, mots clés, classification.

ABSTRACT

The Content Based Image Retrieval and many retrieval systems were conceived generally to indexing the image bases using low-level features (colour, texture and shape). The search engines then calculate the distances between descriptors to find, in the base of images, the most similar to the key image. This Master training try to profit by the benefits of the traditional methods of retrieval for textual information to adjust them with the problems of retrieval in the bases of images. This will consist on the first hand, on making the turn of the existing textual methods, and on the other hand, to propose a scheme of adaptation including the choice of the descriptors, their extraction, then their use with a method considered to be most adequate.

Keywords: textual approach, low-level features, colour, texture, shape, key word, classification.