MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITE MOULOUD MAMMERI, TIZI-OUZOU

FACULTE DE GENIE ELECTRIQUE ET INFORMATIQUE
DEPARTEMENT AUTOMATIQUE

MEMOIRE DE MAGISTER

en ***Automatique***

Option : **Traitement d'Images et Reconnaissance de Formes**

présenté par :

M^{elle} BEDOUHENE Saïda

ingénieur U.M.M.T.O

Thème

Recherche d'images par le contenu

Mémoire soutenu le : devant le jury d'examen composé de :

HAMMOUCHE Kamal	M.C.A à l'U.M.M.T.O	**Président**
DIAF Moussa	Professeur à l'U.M.M.T.O	**Rapporteur**
AHMED OUAMAR Rachid	M.C.A à l'U.M.M.T.O	**Examinateur**
MERAZKA Fatiha	M.C.A à l'ENP Alger	**Examinatrice**
LAHDIR Mourad	M.C.B à l'U.M.M.T.O	**Examinateur**

Avant-propos

Ce mémoire a été effectué au laboratoire « robotique et vision du département Automatique de l'UMMTO.

Mes vifs remerciements vont tout d'abord à Monsieur DIAF Moussa, professeur à l'UMMTO pour m'avoir proposé le thème de ce mémoire et m'avoir dirigée, aidée et conseillée tout le long de notre travail.

Nous exprimons notre reconnaissance à Monsieur HAMMOUCHE Kamal, Maître de Conférences classe A à l'UMMTO, pour m'avoir fait honneur en acceptant de présider le jury de ce mémoire.

Nos remerciements vont également à Monsieur AHMED-OUAMAR Rachid, Maître de Conférences classe A à l'UMMTO pour avoir accepté de faire partie du jury d'examen de notre mémoire.

Nos remerciements vont également à Madame MERAZKA Fatiha, Maître de Conférences classe A à l'ENP Alger pour avoir accepté de faire partie du jury d'examen de notre mémoire.

Que Monsieur LAHDIR Mourad, Maître de Conférences classe B trouve ici nos remerciements pour avoir accepté de faire partie du jury d'examen de notre mémoire.

Sommaire

Introduction générale

Grâce aux avancées récentes de la technologie ces dernières années, en particulier dans le domaine de multimédia et de l'informatique, l'information numérique est devenu le cœur de tous les secteurs d'activités : dans le monde industriel, médical, scientifique, juridique, géographique, etc. Ces progrès se sont accompagnés d'une baisse des coûts des équipements informatiques qui a facilité la diffusion et l'échange de données multimédia numérisées vers le grand public. Cette masse de donnée n'aurait aucun intérêt si l'on ne pouvait pas facilement retrouver les informations concernant un intérêt particulier. Cela a suscité un besoin en développement de techniques de recherche d'information multimédia, et en particulier de recherche d'images.

L'indexation et la recherche d'images par le contenu est une piste prometteuse. Elle offre la possibilité aux utilisateurs d'accéder, d'interroger et d'exploiter directement ces bases d'images en utilisant leur contenu ; ceci explique l'activité de recherche consacrée à ce domaine.

L'idée de faciliter l'accès à des données n'est pas neuve, des techniques de recherche d'images ayant été développées à cet effet depuis la fin des années 70. Parmi ces approches on trouve la technique de recherche d'images à base de texte connue sous le nom « Text-based Image Retrieval » ou TBIR [1] qui est l'approche la plus ancienne utilisée jusqu'à nos jours. Il s'agit d'annoter manuellement chaque image par un ensemble de mots-clés décrivant leur contenu, puis d'utiliser un système de gestion de base de

données pour gérer ces images. À travers des descripteurs textuels, les images peuvent être organisées hiérarchiquement selon les thèmes ou les sémantiques afin de faciliter la navigation et la recherche dans la base. Toutefois, puisque la génération automatique des descripteurs textuels pour un ensemble d'images n'est pas faisable, la plupart de ces systèmes exigent l'annotation manuelle des images [2].

Malgré le grand succès de cette approche pour la recherche des documents, l'annotation textuelle d'images est une tache lourde et coûteuse pour les grandes bases d'images. Elle est souvent confrontée à plusieurs problèmes parmi lesquels, nous pouvons citer : la nécessité de l'intervention d'un humain, leur rigidité, leur subjectivité et les contraintes linguistiques. De plus, l'annotation manuelle des images ne pourra jamais décrire le contenu d'une image d'une manière exhaustive.

L'accroissement spectaculaire du volume des bases d'images oblige à aller au-delà d'une annotation manuelle. Afin de répondre à cette obligation, l'approche basée sur le contenu, visant à extraire directement l'information pertinente à partir de l'image, apparaît comme une alternative à l'approche textuelle. Cette nouvelle modalité a ouvert des possibilités pour les utilisateurs. La recherche d'images par le contenu visuel (Content Based Image Retrieval CBIR ou RIC en français) permet de pallier certains problèmes posés par la description textuelle. Elle s'est révélée efficace et très utile dans de nombreux domaines d'application.

Le CBIR a fait son apparition au début des années 90. Il a un rôle d'aide à la recherche automatique et à la décision. Cette approche, à laquelle nous nous intéressons dans ce mémoire, consiste à représenter chaque image par un ensemble de caractéristiques visuelles de bas niveau telle que la couleur, la texture et la forme. Ces caractéristiques visuelles, calculées de manière automatique, sont ensuite exploitées par le système pour comparer et retrouver des images. Ces dans ce contexte restreint de la recherche d'images par le contenu que se situera notre travail.

Notre manuscrit comportera les parties suivantes :

- *Le premier chapitre,* expose un état de l'art en synthétisant les travaux de recherche réalisés ces dernières années liés au contexte de notre

étude. Puis, il aborde les différentes distances utilisées dans la littérature pour mesurer la similarité entre les descripteurs des images. Enfin il présente quelques systèmes de recherche d'images déjà existants à l'heure actuelle.

- *Le second chapitre*, décrit le principe de la recherche d'images par le contenu et les différents types de requêtes. Puis, il énonce les différentes méthodes d'extraction des descripteurs utilisés dans la recherche CBIR. Ces descripteurs permettent d'extraire des informations pertinentes, en vue d'une exploitation efficace des bases d'images.
- *Le troisième chapitre*, a pour but de décrire les protocoles d'évaluation de nos approches et les résultats que nous avons obtenu. Cette démarche va nous permettre d'évaluer les performances de nos méthodes proposées.
- A l'issue du *troisième chapitre*, nous résumons notre travail en dégageant les apports de ces méthodes. Nous précisons ses limites et certaines perspectives visant à ouvrir des extensions à ce thème de recherche.

Chapitre 1

Etat de l'art

1. Introduction

L'essor des dispositifs d'acquisitions d'images, des capacités de stockage, la baisse des coûts des matériels informatiques et la disponibilité des techniques de numérisation de haute qualité que nous observons ces dernières années, se traduit par une production permanente et considérable d'images numériques dans différents domaines, ce qui a conduit à un développement constant des bases de données d'images.

Pour l'utilisateur de ce type de bases de données, la recherche d'informations est très problématique, elle nécessite une nouvelle technique de traitement des données. Dans ce contexte, la recherche d'images par le contenu s'intéresse à découvrir des connaissances implicitement contenues dans un ensemble de données en s'appuyant sur différentes techniques qui peuvent être mises en œuvre indépendamment ou couplées. Ces techniques visent à explorer les données, à décrire leur contenu, et à en extraire l'information la plus significative.

La recherche d'images par le contenu consiste à caractériser le contenu visuel des images par des descripteurs visuels et d'effectuer des recherches par similarité visuelle à partir de ces descripteurs. Cette nouvelle technique

permet de répondre à de nouveaux besoins dans le domaine de la recherche d'images.

Bien que plusieurs années de travaux scientifiques en recherche d'images par le contenu contemporains, ont d'ores et déjà permis la réalisation d'outils performants, permettant plusieurs formes d'indexation par analyse du contenu mais ce domaine demeure encore aujourd'hui un problème ouvert et très actif.

Dans ce chapitre, nous présentons un panorama des travaux liés aux problèmes de l'accès aux images et de leur manipulation. Plus précisément, nous nous intéressons aux travaux effectués sur l'indexation et la recherche d'images par le contenu. L'objectif est de faire une idée sur quelques approches et systèmes proposés dans la littérature. Nous introduisons aussi les différentes techniques de mesure de similarité utilisées dans différents travaux.

2. Etat de l'art

Les chercheurs dans le domaine de la vision par ordinateur se posent le problème de l'indexation automatique des images par leur contenu, qui permet la recherche d'images par le contenu (CBIR). Cette nouvelle technique pallie les problèmes posés par la recherche textuelle, et permet d'améliorer des applications interrompues et contribue aussi à faire émerger de nouvelles applications dans divers domaines [3].

Le CBIR diffère de la recherche d'informations textuelles essentiellement par le fait que les bases de données d'images sont non-structurées, les images numériques n'˙etant que des matrices d'intensités de pixels, sans signification inhérente les uns par rapport aux autres. Ce qui explique qu'une des questions clé dans tout type de traitement d'images est l'extraction de l'information utile à partir de ces matrices de pixels, avant même de commencer à faire des hypothèses sur le contenu de l'image [4].

La performance des systèmes de recherche d'images dépend pour une grande partie du choix des descripteurs employés et des techniques associées à leur extraction. Un descripteur est défini comme la connaissance utilisée pour caractériser l'information contenue dans les images [5]. De

nombreux descripteurs sont utilisés dans les systèmes de recherche pour décrire les images. Ceux-ci peuvent être différenciés selon deux niveaux :

1- Les descripteurs bas niveau : décrivent le contenu bas niveau de l'image, principalement en termes de couleurs, textures et formes. Ce sont les descripteurs les plus utilisés dans les systèmes actuels, car les plus simples à mettre en place. Notre travail se focalise sur ces descripteurs de bas niveau.
2- Les descripteurs haut niveau : décrivent le contenu sémantique de l'image, et sont principalement des mots clés fournis par l'utilisateur lors de l'indexation.

La première utilisation du terme "recherche d'images par le contenu" dans la littérature a été faite par T. Kato en 1992 [6]. Il s'agissait de rechercher des images à l'aide des caractéristiques de bas niveau telles que la couleur et la texture. A partir de là, le terme a été utilisé pour décrire le processus de recherche d'images dans une base de données à partir de toutes caractéristiques pouvant être extraites automatiquement des images elles-mêmes.

La forme est l'un des attributs de bas niveau le plus utilisé pour décrire la structure géométrique générique du contenu visuel. Même si la caractérisation du contenu géométrique s'est avérée complexe plusieurs primitives géométriques ont montré leurs intérêts dans des systèmes de recherche.

Pour caractériser les formes dans une image, Jain et Valaya [7] ont proposé d'utiliser un histogramme d'orientation des gradients sur les contours (EOH *Edge Orientation Histogram*). Une première étape de détection des contours est mise en œuvre à l'aide de l'opérateur de Canny-Deriche [8, 9]. Pour chaque pixel appartenant à un contour, on accumule l'orientation de son gradient dans un histogramme. Les orientations sont quantifiées sur n bins. Afin de partiellement atténuer les effets de la quantification, l'histogramme est lissé. A chaque bin est en fait associée la moyenne de sa valeur et de celles des deux bins adjacents. Ce descripteur est invariant à la translation, mais, bien évidemment pas à la rotation.

D'autre part, Ferecatu [10] dans sa thèse a proposé un descripteur de formes inspiré par la transformée de Hough (permettant de détecter les lignes dans une image). Ce descripteur travaille sur l'image en niveaux de gris. Pour chaque pixel, on utilise l'orientation de son gradient ainsi que la taille de la projection du vecteur pixel sur l'axe tangent au gradient. Ces deux informations sont captées dans un histogramme en deux dimensions. Il propose également un descripteur de texture. Ce descripteur est basé sur la transformée de Fourier 2D de l'image. Après avoir obtenu la transformée de l'image, deux histogrammes distincts sont calculés sur l'amplitude de la transformée de Fourier. Ils représentent deux types de distributions de l'énergie. Le premier est calculé sur une partition en disques concentriques. Les rayons sont calculés de façon à avoir un incrément de surface identique entre deux disques successifs. Il permet ainsi d'isoler les basses, moyennes et hautes fréquences. Le second découpe le plan complexe en parts, à la manière d'une tarte. Il se focalise donc plutôt sur les variations selon différentes orientations. Ces deux histogrammes sont utilisés conjointement et ont le même poids dans la signature finale.

Quant à Hu [11], il a proposé plusieurs fonctions non linéaires définies sur les moments géométriques qui sont invariantes à la translation, à la rotation et aux changements d'échelle. Ces descripteurs ont été appliqués avec succès à l'identification d'avions, de navires et de visages. Dans cette catégorie nous trouverons également les descripteurs de Fourier qui décrivent le contour par ses composantes fréquentielles, les moments de Zernike et Zernike modifié, qui ont été adaptés par de nombreux auteurs [12]. Ces moments invariants, qui peuvent être extraits d'une image binaire ou d'une image en niveaux de gris, offrent généralement des propriétés de reconstructibilité, ce qui permet d'assurer que les primitives extraites contiennent la plus grande partie de l'information incluse dans la forme étudiée.

De leur coté, Oliva et Torralba [13], se basent sur la manière dont la vision humaine perçoit la structure générale d'une scène. Il apparaît dans leurs expériences que l'aspect général d'une image est perçu de manière assez grossière, indépendamment des nombreux détails qui peuvent y apparaître.

Ainsi, le fait qu'une image soit particulièrement floue empêche d'en percevoir les détails, mais permet néanmoins de comprendre la structure globale de l'image. Ils proposent donc de calculer un descripteur de forme sur des images réduites en imagettes carrées, d'une taille comprise entre 32x32 pixels et 128x128 pixels. Les images réduites sont ensuite divisées en une grille régulière de 4 régions de hauteur et 4 régions de largeur. Enfin, un descripteur est calculé pour chacune des 16 régions obtenues. Ce descripteur est basé sur des histogrammes d'orientation de gradients, également très utilisés pour la description locale des images qui permettent de capturer de manière compacte mais néanmoins précise la forme globale d'une région d'image en caractérisant l'orientation des différents contours qui y apparaissent. Les descripteurs des différentes régions sont ensuite concaténés pour obtenir un vecteur décrivant l'image dans sa globalité. Ici aussi, cette étape implique la présence d'informations géométriques dans le descripteur, puisque chaque sous-partie du descripteur correspondra à une région donnée de l'image.

Au même titre que la forme, la texture est une caractéristique fondamentale des images car elle concerne un élément important de la vision humaine. De nombreuses méthodes ont été proposées afin de définir des descripteurs permettant de caractériser cette notion aussi riche que complexe. L'une des méthodes de description de la texture la plus utilisée pour la recherche d'images par le contenu se base sur les propriétés fréquentielles comme la transformée de Fourier, la représentation de Gabor, les ondelettes et la transformée en cosinus discrète [14].

Les filtres de Gabor sont largement utilisés aujourd'hui, notamment du fait de leur pertinence en regard du système visuel humain. En effet, Marcelja [15] a montré que les cellules du cortex humain pouvaient être modélisées par des fonctions de Gabor à une dimension. L'idée principale de la méthode de Gabor est de décomposer l'image sur des fonctions analysantes obtenues à partir d'une fonction sinusoïdale orientée sur l'axe des x et modulée par une enveloppe gaussienne dans les directions x et y. Cette décomposition a été utilisée par Manjunath et Ma [16] pour des indexations par les textures.

Cependant, une limitation provient du choix non trivial des paramètres pour déformer la fonction analysante. Une classe plus générale de méthodes de description espace - fréquence est ainsi couramment utilisée en traitement d'images : les analyses multi-résolution par ondelettes.

L'idée d'utiliser les ondelettes dans la recherche d'images par le contenu n'est pas nouvelle. Dans sa thèse, Landré [17] a proposé une technique de décomposition multi-résolution des images en utilisant l'analyse en ondelettes à l'aide de l'algorithme lifting scheme. Il utilise une classification automatique des images afin de construire un arbre visuel de recherche. Cette technique se rapproche de la méthode de Bouman et al. [18] qui utilise un arbre de recherche quaternaire pour la navigation.

D'autres approches de systèmes de recherche d'images utilisant les ondelettes ont déjà été publiées. Dans leur article, Jacobs et al. [19] utilisent l'analyse multi-résolution pour créer un index généré à partir des valeurs les plus élevées (en valeur absolue) des coefficients d'ondelettes. Ils utilisent les ondelettes de Haar et une fonction distance adaptée à leur représentation de coefficients d'ondelettes. Mandar et al. [20], quant à eux, ont proposé une autre technique basée sur le calcul de moments à partir des coefficients d'ondelettes et la construction d'histogrammes.

De leur côté, Idris et al [15] proposent une technique de quantification vectorielle dans laquelle la comparaison est effectuée sur des vecteurs quantifiés issus des coefficients d'ondelette.

Dans [21] l'auteur a présenté quatre familles d'outils de caractérisation de texture. On distingue parmi elles les méthodes statistiques, les méthodes géométriques, les méthodes à base de modèles probabilistes et les méthodes fréquentielles. Parmi les grands classiques des méthodes statistiques, il est impossible de ne pas citer les travaux d'Haralick et de Laws. La Grey Level Co-occurrence Matrix (GLCM) a été proposée par Haralick.

D'autres études qui ont pris le plus grand essor ces dernières années se sont orientées vers l'information couleur. De nombreux travaux ont vu en effet le jour quant à l'utilisation de la couleur pour la recherche d'images par le contenu. Une des premières approches d'écrivant l'information couleur a été d'utiliser les histogrammes de couleurs [22] ou les moments de premiers

ordres des distributions [23] tels que la moyenne et la variance. Ces deux descriptions présentent l'avantage d'être invariantes aux translations et rotations opérées sur les images. Cela permet ainsi de décrire deux images identiques à une transformation géométrique près par la même information couleur. Néanmoins, deux images représentant des contenus différents peuvent partager le même histogramme.

Pour surpasser l'inconvénient majeur des histogrammes de couleurs qui est le manque de l'information spatiale concernant la distribution de la couleur, des histogrammes locaux ont été mis en œuvre.

Dans [24], une nouvelle technique à base d'histogrammes locaux de couleurs a été proposée. Cette technique est insensible à la rotation. Elle divise une image en un ensemble de blocs égaux et calcule leur histogramme de couleurs. Elle utilise un graphe biparti pour calculer la distance ayant le coût minimal entre deux images. Dans ce cas chaque bloc de l'image requête est comparé à tous les blocs des images de la base afin de retrouver les images similaires.

Il existe d'autres approches pour intégrer les informations spatiales aux histogrammes de couleurs autres que le partitionnement d'une image en régions.

Dans [25], les histogrammes perpétuellement pondérés ont été proposés. Leur principe de fonctionnement consiste à trouver les couleurs représentatives d'une image. Le nombre de couleurs représentatives est égal au nombre des barres de l'histogramme de couleurs.

Dans [26], les auteurs ont proposé les histogrammes joints pour la description des images. D'abord un ensemble d'attributs locaux de pixels est sélectionné. Ensuite un histogramme multidimensionnel est construit. Chaque entrée de cet histogramme contient le nombre de pixels décrits par une combinaison particulière d'attributs. L'histogramme de couleurs calcule la densité des pixels ayant une couleur particulière tandis qu'un histogramme joint calcule la densité jointe de plusieurs attributs de pixels.

Les histogrammes basés sur le contour ont été utilisés dans [27]. Ils servent à décrire, dans ce cas, la longueur des contours entre différentes couleurs

dans le but de prendre en considération les informations géométrique pour l'indexation des images.

Les vecteurs de cohérence de couleurs ont été proposés dans [28]. Ces derniers classifient chaque pixel en cohérent ou non cohérent selon l'appartenance de ce pixel à une région homogène de couleur. Ensuite l'histogramme de couleurs est construit et la valeur de chaque barre de l'histogramme est le nombre de pixels cohérents. Cette approche intègre un peu d'information spatiale. Elle peut être vue comme un raffinement de l'histogramme, cependant, elle présente l'inconvénient d'amplifier la sensibilité aux conditions d'illumination, contrairement au corrélogramme. Partant de cette dernière observation, Huang et al [29] utilisent les corrélogrammes de couleurs comme primitives, où le corrélogramme représente les corrélations de couleurs entre pairs de pixels séparés par plusieurs distances. Cette dernière extraction d'information améliore les résultats de recherche et elle est alors considérée comme meilleure.

Pour leur part, Vertan et Boujemaa [30] ont utilisé l'histogramme de couleurs pondérés. La méthode repose sur la combinaison des informations de couleur et de structure (texture et/ou forme) dans une même représentation. Il est bien connu que les histogrammes de couleurs classiques ne conservent aucune information sur la localisation des pixels dans l'image. Mais on sait également que des pixels ayant la même couleur n'ont pas forcément la même importance visuelle en fonction, justement, de leur localisation. Ainsi est arrivée l'idée d'inclure une information sur l'activité de la couleur dans le voisinage des pixels, mesurant ainsi l'uniformité ou la non-uniformité locale de l'information couleur.

Dans [31], les auteurs proposent l'utilisation d'angles de couleurs. Cette technique permet d'obtenir une invariance selon le modèle diagonal. Dans ce dernier : lors d'une variation d'illumination tous les niveaux de rouge de l'image sont multipliés par une constante, tous les niveaux de vert par une autre constante, et tous les niveaux de bleu par une troisième constante. Ensuite, chaque image est représentée par trois vecteurs, un vecteur pour chaque couleur. Chaque vecteur comporte une dimension pour chaque pixel. Une variation d'illumination se traduit alors par un simple changement

d'échelle du vecteur, sans changement d'orientation. Chaque image est alors caractérisée par les angles formés entre ces trois vecteurs.

D'autres études comme celle de Cordelia Schmid [32] sont orientées à l'utilisation des points d'intérêt pour décrire le contenu visuel des images. Des améliorations ont été apportées par Patrick Gros [33]. Une contribution importante qui permet de trouver des points d'intérêts invariants par transformation affine du plan est apportée par Krystian Mikolajczyk [34]. L'étude d'une description invariante de contours par les transformations affines du plan est réalisée par Stanislaw Matusiak [35]. Des méthodes utilisant la logique floue permettent d'améliorer les résultats en ajoutant la notion de flou dans la requête a été proposé par Patrick Lambert [36].

La thèse de Julien Fauqueur [37] étudie la recherche d'images par composition de catégories de régions issues de la segmentation couleurs des images.

Pour sa part, Scott Cohen [38], aborde le problème de la recherche d'images par le contenu avec une approche de distribution couleurs et de reconnaissance de contours d'objets en utilisant la distance EMD *(Earth Mover Distance*). Il étudie également les requêtes partielles dans lesquelles on cherche des images par comparaison d'attributs de régions similaires.

Quant à Sid-Ahmed Berrani [39], il a utilisé les techniques d'indexation de données multidimensionnelles et une recherche approximative des plus proches voisins afin de déterminer les images les plus proches de l'image requête de l'utilisateur.

Parmi les travaux réalisés en utilisant le standard JPEG2000, nous trouvons: En 1999, *Do et al* [40] ont proposé l'attribut mélange gaussien généralisé utilisé avec la mesure de similarité Kullback. En 2000, *Liu et al* [41] ont proposé l'histogramme bidimensionnel de bits significatifs calculé sur les bandes d'ondelettes. En 2002, *Ziyou et al* [42] ont utilisé la variance, calculée sur les coefficients d'ondelettes, afin de rendre la recherche moins coûteuse en termes de ressource [43].

3. Mesure de similarité entre descripteurs

Après avoir dressé un état de l'art des travaux effectués dans le domaine de la recherche d'images par le contenu, nous pouvons maintenant aborder le problème de mesure de similarité entre les images. Le sujet ayant été souvent traité dans la littérature [44] [45], nous ne dressons pas de liste exhaustive des diverses mesures possibles. En outre, nous nous limitons au cadre de l'indexation vectorielle.

La recherche des contenus visuellement similaires est un thème central dans le domaine de la recherche d'images par le contenu. Pendant la dernière décennie, de nombreuses méthodes ont été proposées pour identifier des contenus visuels semblables du point de vue de la couleur, de la texture et de la forme. Ces méthodes regroupent des caractéristiques mesurées d'une image dans un vecteur multidimensionnel. La similarité entre deux images peut alors être mesurée à l'aide d'une métrique définie sur l'espace vectoriel ainsi défini. Lorsqu'un utilisateur lance une requête, le système effectue une mesure entre le descripteur de la requête et les descripteurs des images de la base dans l'espace des attributs. Les images sont considérées similaires si la distance entre eux est faible. Pour cela, la complexité de calcul de la distance doit être raisonnable puisque la tâche de recherche s'effectue en temps réel.

Le premier type de mesure de similarité correspond aux distances géométriques entre vecteurs. Dans ce cas, on parle de distances car ces mesures ont la propriété de respecter les axiomes des espaces métriques. De nombreuses définitions de distances géométriques ont été proposées, chacune donnant bien entendu des résultats différents, le choix d'une fonction distance est primordial dans ce domaine.

Considérons deux images I_1 (image requête) et I_2 (image dans la base) indexées par des vecteurs dans R^N: $I_1 = \{I_1(i), 1 \leq i \leq N\}$ *et* $I_2 = \{I_2(i), 1 \leq i \leq N\}$

Le calcul de similarité entre I_1 et I_2 passe par une mesure de proximité entre les signatures I_1 et I_2. Dans ce contexte, toute mesure entre vecteurs de R^N est utilisable et nous nous appliquons à citer les plus couramment employées pour la recherche d'images par le contenu.

- **Distances de Minkowski**

L'approche la plus simple pour mesurer la similarité entre deux images correspond aux distances de Minkowski. Cette distance L_r est définie par :

$$L_r(I_1, I_2) = [\sum_{i=1}^{n} |I_1(i) - I_2(i)|^r]^{\frac{1}{r}} \quad (1)$$

Où $r \geq 1$ est le facteur de Minkowski et n la dimension de l'espace caractéristique. Les métriques de Minkowski représentent un bon compromis entre efficacité et performance. Pour cette famille de distances, plus le paramètre *r* augmente, plus la distance L_r aura tendance à favoriser les grandes différences entre coordonnées. Ces distances sont rapides à calculer et simples à implémenter, par contre leur calcul est réalisé en considérant que chaque composante du vecteur apporte la même contribution à la distance.

Pour *r=1*, on obtient la distance de Manhattan ou city block:

$$L_1(I_1, I_2) = \sum_{i=1}^{n} (|I_1(i) - I_2(i)|) \quad (2)$$

Cette norme est aussi connue sous le nom city-block est plus appropriée pour mesurer la similarité entre les données multi-variées ; elle est moins sensible au bruit coloré que la distance euclidienne. Cette distance a entre autres été utilisée par Swain et Ballard [22], ou encore Stricker et Orengo [24].

Pour *r=2,* on obtient la distance Euclidienne :

$$L_2(I_1, I_2) = \sqrt{\sum_{i=1}^{n} (I_1(i) - I_2(i))^2} \quad (3)$$

La distance euclidienne est invariable aux translations et aux rotations des données dans l'espace des attributs et couramment utilisée dans des espaces à 2 ou 3 dimensions, cette métrique donne de bons résultats si l'ensemble des données présente des classes compactes et isolées. Cette distance a été notamment utilisée par Niblack et al. [46] dans le système QBIC d'IBM.

Pour *r=∞*, on obtient la distance Chebyshev ou la distance de maximum:

$$L_\infty(I_1, I_2) = \lim_{r \to \infty} \sqrt[r]{\sum_{i=1}^{n} |I_1(i) - I_2(i)|^r} = sup_i(|I_1(i) - I_2(i)|) \quad (4)$$

Cette distance est adaptée aux données de grande dimension, elle est souvent employée dans les applications où la vitesse d'exécution est importante. Cette distance examine la différence absolue entre les différents

pairs des vecteurs, elle est considérée comme une approximation de la distance Euclidienne mais avec moins de calcul.

Afin de rendre compte de l'importance relative des composantes du vecteur les unes par rapport aux autres, les distances de Minkowski pondérés (équation.5) ont été proposées.

$$L_r^w(I_1, I_2) = [\sum_{i=1}^{n}(w_i|I_1(i) - I_2(i)|^r)]^{\frac{1}{r}} \quad (5)$$

Où w est un vecteur de pondération à n composantes.

- **Distance quadratique**

La distance de Minkowski traite les éléments du vecteur caractéristique d'une manière équitable. La distance quadratique en revanche favorise les éléments les plus ressemblants [47]. Sa forme générale est donnée par :

$$Dist_q(I_1, I_2) = \sqrt{(I_1 - I_2)^T A (I_1 - I_2)}$$
$$= \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}(I_1(i) - I_2(i))(I_1(j) - I_2(j))a_{ij}} \quad (6)$$

Où $A = [a_{ij}]$ est la matrice de similarité, a_{ij} représente la distance entre deux éléments des vecteurs I_1 et I_2 :

$$a_{ij} = 1 - \frac{d_{ij}}{d_{max}} \quad (7)$$

où d_{ij} est la distance dans l'espace couleurs considéré et d_{max} le maximum global de cette distance. Les propriétés de cette distance la rendraient proche de la perception humaine de la couleur, ce qui en fait une métrique attractive pour l'application CBIR.

- **Distance de Mahalanobis**

Cette distance prend en considération la corrélation entre les données ; de plus elle n'est pas dépendante de l'échelle de données. Elle est ainsi définie par :

$$Dist_{Mah}(I_1, I_2) = (I_1 - I_2)\,\Sigma^{-1}(I_1 - I_2)^T \quad (8)$$

Où Σ est la matrice covariance entre l'ensemble des descripteurs d'images. La distance de Mahalanobis tient compte de la distribution statistique des données dans l'espace, c'est ce qui la différencie des autres distances.

- **Distance d'Earth Mover (EMD)**

La distance EMD a été utilisée dans les systèmes CBIR. Elle définie une mesure quantitative de travail minimale pour changer une signature en une autre. Le calcul cette distance se ramène à la solution d'un problème de transport résolu par optimisation linéaire. Elle est alors définie comme :

$$Dist_{EMD}(I_1, I_2) = \frac{\sum_{i=1}^{n_{I_1}} \sum_{j=1}^{n_{I_2}} g_{ij} d_{ij}}{\sum_{i=1}^{n_{I_1}} \sum_{j=1}^{n_{I_2}} g_{ij}} \quad (9)$$

Où d_{ij} indique la distance entre les deux composantes i et j, et g_{ij} est le flot optimal entre les deux distributions. Le coût total $\sum_{i=1}^{n_{V^1}} \sum_{j=1}^{n_{V^2}} g_{ij} d_{ij}$ est minimal sous les contraintes suivantes :

$$g_{ij} \geq 0, \forall i, j \quad (10)$$

$$\sum_{i=1}^{n_{V^1}} g_{ij} \leq I_2(j), \forall j \quad (11)$$

$$\sum_{j=1}^{n_{V^2}} g_{ij} \leq I_1(i), \forall i \quad (12)$$

$$\sum_{i=1}^{n_{V^1}} \sum_{j=1}^{n_{V^2}} g_{ij} = min(I_1(i), I_2(i)) \quad (13)$$

La première contrainte n'autorise que des mouvements des composantes de I_1 vers I_2. Les deux contraintes suivantes limitent la quantité de composantes déplacée de I_1, et la quantité de composantes reçue par I_2. La dernière contrainte implique le maximum de déplacement de composantes possible. Cette méthode de comparaison est très coûteuse car elle requiert la résolution d'un problème d'optimisation linéaire soluble de façon itérative.

- **Distance de Bhattacharyya**

La distance de Bhattacharyya [48] peut être utilisée pour comparer la similarité entre deux histogrammes Q et V de deux images. On définie cette distance par la relation suivante :

$$Dist_{Bha}(Q, V) = 1 - \sum_i \sqrt{Q_i} \sqrt{V_i} \quad (14)$$

D'autres mesures ont été proposées, comme les distances issues de la distribution. Le problème de mesure de similarité se ramène dans ce cas à une mesure entre distributions de probabilités. Ces méthodes sont efficaces pour la mesure de similarité mais elles ne traitent pas le problème de la non linéarité des données, de plus elles sont coûteuses au niveau temps de calcul particulièrement dans le cadre de l'apprentissage. On trouve :

- ***Distance de Kullback-Leibler***

La divergence de Kullback-Leibler, issue de la théorie de l'information, permet de mesurer la dis-similarité basée sur l'entropie mutuelle de deux distributions de probabilités, sa forme générale est donnée par :

$$Dist_{Kul}(I_1, I_2) = \sum_{i=1}^{n} I_1(i) \log \frac{I_1(i)}{I_2(i)} \quad (15)$$

- ***Divergence de Jeffrey***

La divergence de Jeffrey est symétrique et plus stable que la divergence de Kullback-Leibler, elle est définie par :

$$Dist_{JD}(I_1, I_2) = \sum_i \left(I_1(i) \log \frac{I_1(i)}{\hat{V}(i)} + I_2(i) \log \frac{I_2(i)}{\hat{V}(i)} \right) \quad (16)$$

Où $\hat{V}(i) = (I_1(i) + I_2(i))/2$.

- ***Distance de Kolmogorov-Smirnov***

La distance de Kolmogorov-Smirnov est appliquée aux distributions cumulées $I^c(i)$, elle est définie par:

$$Dist_{KS}({I_1}^c, {I_2}^c) = max_i(|{I_1}^c(i) - {I_2}^c(i)|) \quad (17)$$

- ***Distance de Cramer-Von Mises***

Cette distance s'applique également sur des distributions cumulées, elle est définie par :

$$Dist_{CVM}({I_1}^c, {I_2}^c) = \sum_i \left({I_1}^c(i) - {I_2}^c(i) \right)^2 \quad (18)$$

- **Intersection d'histogrammes**

Cette mesure est l'une des premières distances utilisée dans les systèmes CBIR. Elle a été proposée par Swain et Ballard [22]. Cela permet en fait d'évaluer le recouvrement de deux histogrammes normalisés H^1 et H^2. La distance de Swain et Ballard s'exprime ainsi :

$$d(H^1, H^2) = 1 - \frac{\sum_{j=1}^{n} min(H_j^1, H_j^2)}{\sum_{j=1}^{n} H_j^2} \quad (16)$$

Où n est le nombre de valeurs de chaque histogramme. Deux images présentant une intersection normalisée d'histogrammes proche de 1 sont considérées comme similaires. Cette mesure n'est pas une métrique parce que non symétrique. L'intersection des histogrammes n'est pas invariante aux changements d'illuminations.

4. Quelques systèmes de recherche d'images

Ces dernières années, de nombreux systèmes d'indexation et de recherche d'images par le contenu, ont vu le jour. La plupart de ces systèmes, permettent de naviguer au sein de la base d'images, et/ou d'effectuer des recherches par l'exemple et d'exprimer des requêtes au moyen d'une interface graphique conviviale et adéquate. Voici une liste des systèmes les plus réputés.

- **QBIC**

Le logiciel QBIC (Query By Image Content) d'IBM, est le premier système commercial de recherche d'images par le contenu. Il supporte différents types de requêtes : par l'exemple, par croquis…etc. Les auteurs de ce logiciel fut rassembler un large panel de descripteurs d'informations contour, couleur et texture. Le système utilise la moyenne pour caractériser la couleur dans les espaces RGB, YIQ, Lab et Munsell. La texture est représentée par une version améliorée des caractéristiques de Tamura. La forme est assurée par des signatures classiques comme la circularité, la surface, l'excentricité et les moments invariants. La distance Euclidienne est utilisée pour comparer les images et la distance quadratique pour comparer les histogrammes.

- **Virage**

Virage est le moteur de recherche d'images développé par la société Virage Inc. Son objectif est de construire un environnement dédié à la recherche d'images, principalement composé de primitives. Similairement à QBIC, Virage propose des requêtes portant sur la couleur, la localisation des couleurs, la texture et la structure de l'image. L'interface de Virage offre la possibilité d'ajouter et de pondérer les différentes primitives ainsi que d'utiliser le bouclage de pertinence. L'avantage de Virage par rapport à QBIC est qu'il autorise une combinaison entre les différents modes de recherche. L'utilisateur définit le poids qu'il veut attribuer à chaque mode.

- **Photobook**

Photobook est un système d'indexation d'images développé par le MIT Media Laboratory. Ce système se base sur la couleur, la texture et la forme pour

définir les signatures d'une image. La requête s'effectue classiquement par choix d'une image candidate. La distance Euclidienne est utilisée pour mesurer la similarité.

- **Blobworld**

Blobworld a pour but de retrouver, à partir d'une image requête, des régions similaires en couleur et texture, appelés blobs, dans les images de la base. La couleur est décrite par un histogramme de 218 cases dans l'espace Lab. Le contraste et l'anisotropie ont été utilisés pour représenter la texture. Pour caractériser la forme, Blobworld utilise la surface, l'orientation et l'excentricité. La requête consiste à sélectionner une région qu'il juge l'utilisateur importante dans une image segmentée. La distance utilisée combine la distance quadratique (pour la couleur) et la distance euclidienne (pour la forme et la texture).

- **BDLP**

BDLP (Berkelely Digital Library Project) a été développé à l'université de Californie, Stanford, USA. Il permet à l'utilisateur d'interroger une base d'images par le contenu visuel et certains mots clés comme : la collection, la localisation spatiale et le nom du photographe. Pour extraire le contenu visuel, les couleurs de chaque image sont quantifiées en treize couleurs. Six valeurs sont associées à chaque couleur : le pourcentage de cette couleur dans l'image et le nombre de régions qui ont une taille *très petite*, *petite*, *moyenne*, *grande*, et *très grande* de cette couleur dans l'image.

A l'interrogation, l'utilisateur peut sélectionner treize couleurs au plus et indiquer la quantité (« n'importe quelle », « une partie », « la plupart ») de chaque couleur dans l'image. En outre, pour les régions qui ont cette couleur, l'utilisateur peut indiquer leurs tailles (n'importe laquelle, petite, moyenne, grande) et le nombre de régions (n'importe laquelle, peu, quelques-unes, beaucoup).

- **Netra**

Netra est un système de recherche d'images développé au sein de l'université de Santa Barbara. Netra utilise une segmentation pour calculer les attributs de couleur, texture et forme. Les filtres de Gabor avec ondelettes est

l'approche adoptée pour représenter la texture. Le système utilise courbures de la forme comme descripteurs. Pour mesurer la similarité Netra emploie la distance euclidienne.

- **FRIP**

FRIP (Finding Regions in the Pictures) a été développé à l'université de Yonsei, Korea. Les images, dans ce système, sont segmentées en régions en utilisant des filtres circulaires. Pour chaque région, la couleur, la texture et la forme sont extraites. La couleur est représentée par la couleur moyenne dans l'espace RVB. La texture est extraite en employant une ondelette bi-orthogonale. La forme est représentée par un vecteur de 12 valeurs, représentant les distances entre les centres de régions et les 12 points de la frontière prélevés d'une manière uniforme après la fixation de l'orientation. En outre, la superficie, les coordonnées du centre et la longueur de l'axe majeur et mineur sont stockés. L'utilisateur choisit une image pour la requête, cette image sera segmentée. De cette dernière, l'utilisateur sélectionne la région d'intérêt. Les images retrouvées sont présentées dans l'ordre décroisant de leur similarité.

- **VisualSeek**

Dans VisualSeek, les caractéristiques sont un ensemble de couleurs dominantes définies dans l'espace HSV, les ondelettes pour la texture ainsi que les relations spatiales entres régions. La recherche consiste à retrouver des régions de même couleur dominante avec la même distribution spatiale. Les distances euclidienne et quadratique sont les deux mesures employées.

- **SIMPLIcity**

SIMPLIcity (Semantics -Sensitive Integrated Matching for Picture LIbraries) vise à réduire le fossé sémantique dans les systèmes de recherche d'images par le contenu. Les images sont segmentées en régions, la caractérisation de chaque région est basée sur les ondelettes. Ainsi les images sont automatiquement triées suivant des critères sémantiques simples, ce qui permet ensuite d'accélérer et d'aider la recherche d'images similaires.

- **MARS**

MARS est système inter disciplinaire qui implique plusieurs domaines de recherche: traitement d'image, gestion de base de données et recherche d'information. Pour la caractérisation visuelle, l'image est découpée en blocs de 5*5. Des indices de texture et de couleur sont calculés pour chaque bloc d'image. La couleur est représentée par un histogramme 2D (coordonnées HS), les coefficients d'ondelettes décrivent la texture. La segmentation des images se déroule en deux procédures. La première est l'algorithme des k-moyennes sur l'espace couleur/texture, la deuxième une détection de régions par regroupement selon un modèle d'attraction. Le système est paramétrable : par exemple on peut choisir la palette de couleurs. De plus Mars offre un nombre d'operateurs logiques pour formuler la requête. La ressemblance entre couleurs est mesurée par l'intersection d'histogrammes. Pour les textures le système applique une distance euclidienne.

- **PicHunter**

PicHunter est basé sur l'histogramme et la distribution spatiale de la couleur pour construire le vecteur descripteur de l'image. La distance utilisée est de type Minkowski L_1. Le système PicHunter incorpore une boucle de retour de pertinence probabiliste qui prédit les images cible d'après l'interaction système/usager.

- **IKONA**

IKONA est un système de recherche et de navigation interactive dans de grandes bases de données multimédia développées par l'équipe IMEDIA de l'INRIA est basé sur une architecture client/serveur. IKONA procède à l'extraction des descripteurs de la couleur à l'aide de l'histogramme de couleurs pondéré. L'extraction de la texture est effectuée en utilisant le spectre de Fourier. Pour la forme, il utilise l'histogramme d'orientation des contours.

- **Cortina**

Cortina [49] utilise des descripteurs issus de la norme MPEG-7 et des mots issus du texte autour des images dans les pages web pour construire son index d'images. Le regroupement d'images est réalisé avec l'algorithme des k

plus proches voisins. La requête est soit une requête par mot-clé, soit une requête par image exemple.

- **Kiwi**

Kiwi [17] (Key-points Indexing Web Interface) est un système développé à l'INSA de Lyon. Il est basé sur une analyse des images et l'extraction de points d'intérêts multi-résolution des images en utilisant les ondelettes. Les histogrammes de couleurs sont construits pour un sous-ensemble de pixels, définis par une matrice 3X3 autour des points clés. Ensuite, les trois premiers moments (moyenne, variance, dissymétrie) de l'histogramme sont stockés et utilisés pour décrire la distribution de couleurs dans l'image.

- **WINDSURF**

WINDSURF [50] est basé sur la décomposition en ondelettes des images, suivie par une segmentation des régions à l'aide des nuées dynamiques et par l'extraction d'attributs colorimétriques et de texture. Les régions de l'image requête sont ensuite comparées selon la distance de Mahalanobis pour donner les images les plus proches de la requête.

- **RETIN**

Le système RETIN [51] (Recherche et Traque Interactive) a été développé à l'ENSEA de Cergy-Pontoise. Dans ce système, un ensemble de pixels est sélectionné au hasard dans chaque image. Il utilise des attributs de couleur dans l'espace Lab. La texture de ces pixels est obtenue par l'application des douze filtres de Gabor (quatre différentes directions et trois différentes fréquences). L'utilisateur formule sa requête par une image exemple.

5. Domaines d'applications

Les applications des systèmes de recherche d'images par le contenu sont variées. Citons les plus importantes :

Des applications judiciaires : les services de police possèdent de grandes collections d'indices visuels (visages, empreintes) exploitables par des systèmes de recherche d'images.

Les applications militaires, bien que peu connues du grand public, sont sans doute les plus développées: reconnaissance d'engins ennemis via images radars, systèmes de guidage, identification de cibles via images satellites.

Le journalisme et la publicité sont également d'excellentes applications. Les agences de journalisme ou de publicité maintiennent en effet de grosses bases d'images afin d'illustrer leurs articles ou supports publicitaires. Cette communauté rassemble le plus grand nombre d'utilisateurs de recherche par le contenu (davantage pour les vidéos) mais l'aide apportée par ces systèmes n'est absolument pas à la hauteur des espoirs initiaux.

D'autres applications incluent : le diagnostic médical, les systèmes d'information géographique, la gestion d'œuvres d'art pour explorer et rechercher des peintures similaires, .. . Architecture pour retrouver des bâtiments ou des aménagements intérieurs,. . .

6. Conclusion

Dans cette recherche bibliographique sur l'indexation et la recherche d'images par le contenu, nous avons exploré, dans un premier lieu, les différents travaux effectués. En second lieu, nous avons présenté les différentes distances employées pour mesurer la similarité entre les images. En troisième lieu, nous avons dressé une liste de quelques systèmes existant à l'heure actuelle.

Les systèmes de recherche d'images par le contenu permettent aux utilisateurs d'accéder à l'information visuelle de manière plus adaptés que la manière utilisée par les systèmes de recherche d'information traditionnels. Le problème majeur lié au développement d'un système d'indexation et de recherche d'images par contenu dans un contexte générique consiste à la mise en place des techniques capables de décrire le contenu visuel des images et de prendre compte des besoins des utilisateurs et leurs différents points de vue en ce qui concerne l'interprétation des images. Ainsi, le choix de la mesure de similarité doit être réalisé en prenant en compte toutes les informations disponibles sur l'ensemble des données car certaines d'entre elles peuvent favoriser les attributs d'un ordre d'échelle plus important et par conséquence, la contribution des attributs moins significatifs sera

négligée. C'est le cas de la distance euclidienne et des métriques de Minkowski d'ordre supérieure.

Malgré leur utilité et leur popularité, les systèmes de recherche actuels souffrent de certaines limitations, comme le manque de sémantique dans le traitement des requêtes, l'imprécision des résultats, une faible interactivité, ou encore un manque d'intégration de techniques de traitement d'images.

Le chapitre suivant se focalisera à une étude détaillée du principe de la recherche d'images par le contenu et les différents attributs visuels pouvant être extraire des images. Le choix d'un meilleur ensemble de descripteurs visuels promet une bonne caractérisation du concept de couleur, de texture et de forme.

Chapitre 2

Principe des systèmes CBIR et les différents descripteurs d'image

1. Introduction

L'indexation et la recherche d'images par le contenu sont l'une des solutions possibles et prometteuses pour gérer les bases de données d'images numériques, qui ne cessent d'augmenter. Elles visent à résoudre ce problème en se basant sur un paradigme de représentation de bas niveau du contenu de l'image, par la couleur, la texture, la forme, etc.., et d'autres par une combinaison de celles-ci. Il est donc indispensable de développer des outils permettant de sélectionner les images les plus pertinentes par leur aspect visuel. Dans ce domaine, plusieurs méthodes d'indexation peuvent êtres recensées. Elles convergent toutes vers une description ou une caractérisation de l'image dans un domaine descriptif mais complexe, et adaptatif.

L'objectif de ce chapitre est de présenter le principe général des systèmes CBIR. Puis on s'intéressera à l'étude des différentes approches d'indexation et de recherche d'images par le contenu. La littérature dans ce domaine étant très riches, nous allons présenter dans ce qui suit les techniques les plus pertinentes.

2. Architecture d'un système d'indexation et de recherche d'images

Deux aspects indissociables coexistent dans notre problème, l'indexation et la recherche. Le premier concerne le mode de représentation informatique des images et le second concerne l'utilisateur de cette représentation dans le but de la recherche. L'architecture classique d'un système d'indexation et de recherche d'images par le contenu, présentée en figure 1, se décompose en deux phases de traitement : une phase d'indexation dit, hors ligne est une étape de caractérisation où les attributs sont automatiquement extraits à partir des images de la base, et stockés dans un vecteur numérique appelé descripteur visuel. Ensuite, ces caractéristiques sont stockées dans une base de données. Et une autre phase de recherche dit, en ligne consiste à extraire le vecteur descripteur de l'image requête proposer par l'utilisateur et le comparer avec les descripteurs de la base de données en utilisant une mesure de distance. Le système renvoi le résultat de la recherche dans une liste d'images ordonnées en fonction de la similarité entre leurs descripteurs visuels et le descripteur visuel de l'image requête.

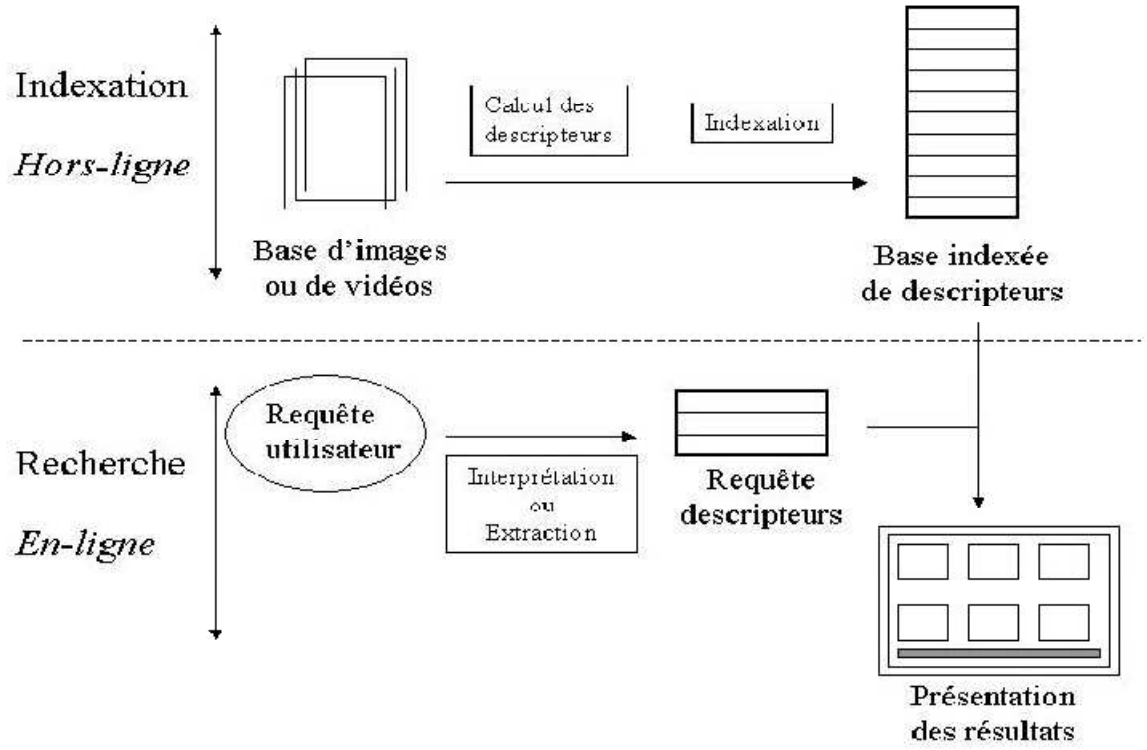


Fig.1**.** Architecture d'un système d'indexation et de recherche d'images

3. Les types de requête

La première étape de la recherche d'images est la constitution de la requête. Elle doit permettre au système de retrouver les images désirées par l'utilisateur. Suivant les besoins de l'utilisateur et le type de base de données images, plusieurs requêtes sont proposées. Requête par mots clés, requête par esquisse (croquis), requête par exemple et la combinaison de celles-ci afin d'accéder à un niveau d'abstraction supérieur. La figure [52] suivante illustre les trois grandes catégories de requête:

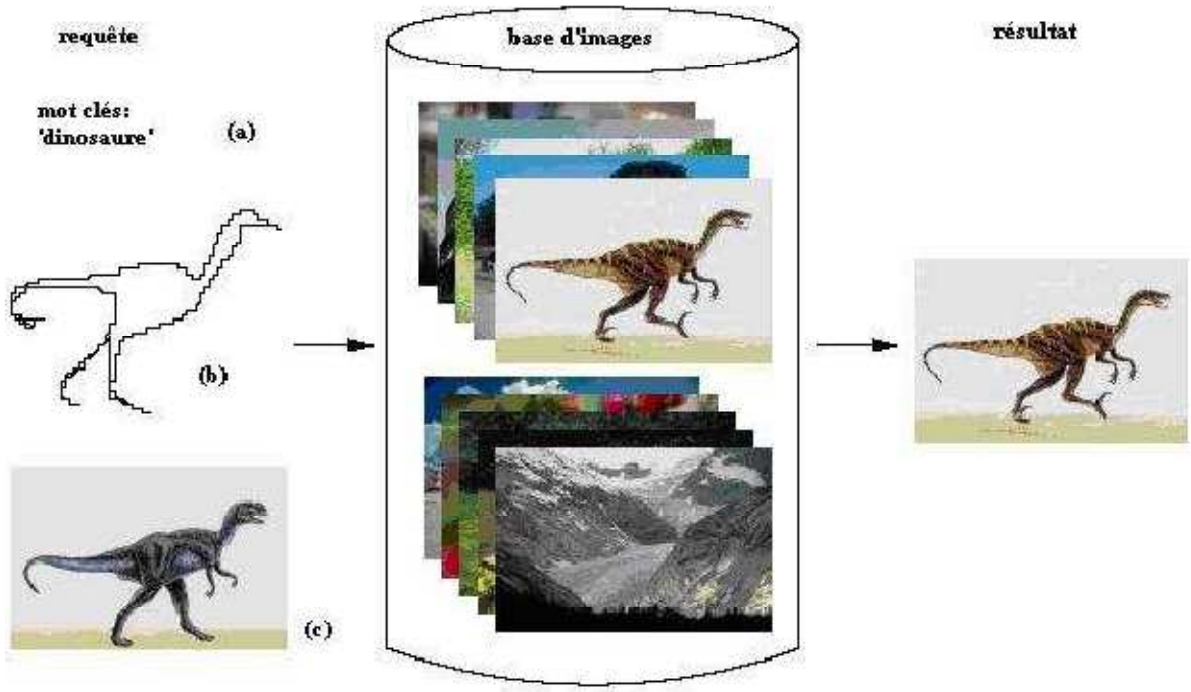


Fig.2. Différents types de requêtes

3.1. Requête par mots clés

Les images sont recherchées suivant un ou plusieurs critères, par exemple trouver les images contenant 80% de rouge. Donc, le système se base sur l'annotation manuelle et textuelle d'images.

3.2. Requête par esquisse

Dans ce cas, le système fournit à l'utilisateur des outils lui permettant de constituer une esquisse (dessin) correspondant à ses besoins. L'esquisse fournie sera utilisée comme exemple pour la recherche. L'esquisse peut être une ébauche de forme ou contour d'une image entière ou une ébauche des couleurs ou textures des régions d'une image. L'utilisateur choisira, en fonction de la base d'images utilisée, de ses besoins et préférences, l'une ou l'autre de ces représentations. Cette technique présente l'inconvénient majeur qu'il est parfois difficile pour l'utilisateur de fournir une esquisse, malgré les outils qui lui sont fournis.

3.3. Requête par l'exemple

Pour les systèmes de recherche d'images à base d'exemples, l'utilisateur, pour représenter ses besoins, utilise une image (ou une partie d'image) qu'il considère similaire aux images qu'il recherche. Cette image est appelée image exemple ou requête. L'image exemple peut soit être fournie par l'utilisateur, soit être choisie par ce dernier dans la base d'images utilisée. Cette technique est simple et ne nécessite pas de connaissances approfondies pour manipuler le système.

4. Extraction des caractéristiques

Le but de l'indexation est de fournir une représentation image permettant des recherches efficaces. Il ne s'agit pas de coder toute l'information portée par l'image mais de se concentrer sur l'information qui permet de traduire efficacement une similarité proche des besoins exprimés par un utilisateur. Une des clés de l'indexation efficace est l'extraction des caractéristiques primaires en accord avec le type et le but des recherches visées par le système. Ces caractéristiques sont généralement simples, intuitives et génériques.

L'analyse faite du signal se focalise généralement autour des attributs de bas niveau tel que la couleur, la texture et la forme. L'extraction de ces attributs constitue le premier pas de toutes les procédures d'analyse d'images qui visent à un traitement symbolique de leur contenu [53].

Il y a principalement deux approches pour les caractéristiques qui peuvent être extraites. La première est la construction de descripteurs globaux à toute l'image. Dans ce cas, il s'agit de fournir des observations sur la totalité de l'image. L'avantage des descripteurs globaux est la simplicité des algorithmes mis en œuvre, et le nombre réduit d'observations que l'on obtient. Cependant, l'inconvénient majeur de ces descripteurs est la perte de l'information de localisation des éléments de l'image. La seconde approche est locale consiste à calculer des attributs sur des portions restreintes de l'image. L'avantage des descripteurs locaux est de conserver une information localisée dans l'image, évitant ainsi que certains détails ne soient noyés par le reste de l'image. L'inconvénient majeur de cette technique est que la

quantité d'observations produite est très grande, ce qui implique un gros volume de données à traiter.

Le choix des caractéristiques extraites est souvent guidé par la volonté d'invariance ou de robustesse par rapport à des transformations de l'image.

4.1. Les descripteurs de couleurs

La couleur est une caractéristique riche d'information et très utilisée pour la représentation des images. Elle forme une partie significative de la vision humaine. La couleur est devenue la première signature employée pour la recherche d'images par le contenu en raison de son invariance par rapport à l'échelle, la translation et la rotation [54]. Ces valeurs tridimensionnelles font que son potentiel discriminatoire soit supérieur à la valeur en niveaux de gris des images. Une indexation couleur repose sur deux principaux choix : l'espace colorimétrique et le mode de représentation de la couleur dans cet espace [4].

4.1.1 Les espaces de couleurs

Avant de sélectionner un type de description du contenu couleur, il convient de choisir un espace de couleurs. Une couleur est généralement représentée par trois composantes. Ces composantes définissent un espace de couleurs. Plusieurs études ont été réalisées sur l'identification d'espaces colorimétriques le plus discriminants mais sans succès car il n'existe pas d'espace de couleurs idéal [55]. Il existe plusieurs espaces colorimétriques qui ont chacun certaines caractéristiques intéressantes.

L'espace RGB est très simple à utiliser, car c'est celui employé par de nombreux appareils de capture d'images qui effectuent leurs échanges d'informations uniquement en utilisant les triplets (*R*ouge, *V*ert, *B*leu). On parle d'espace colorimétrique orienté matériel. Cette manière de représenter la couleur est extrêmement basique, puisqu'aucun traitement n'est nécessaire. Cependant, ces trois composantes sont fortement corrélées, cet espace est sensible aux changements d'illumination, et ne correspond pas au processus de perception humaine. La représentation des couleurs dans cet espace donne un cube de Maxwel comme illustré dans la figure.3.

En vision par ordinateur, les trois composantes sont représentées par trois octets (24 bits). On obtient avec ce choix une palette de 256^3 couleurs possibles. Les systèmes de recherche d'images par le contenu utilisent une quantification pour réduire le nombre de couleurs.

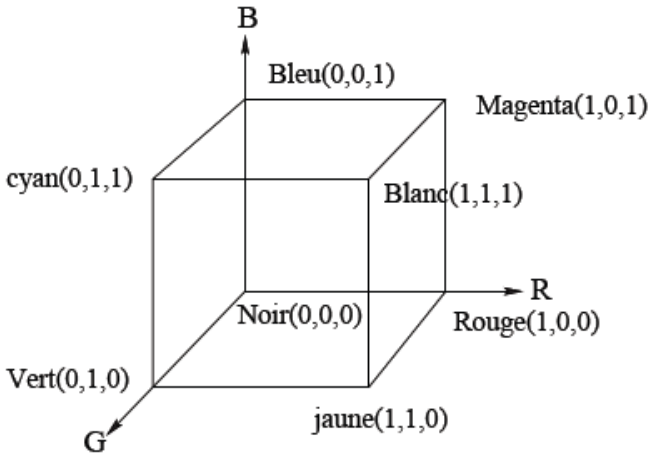


Fig.3. Cube de Maxwel

HSI cette représentation est plus sophistiquée. En effet, elle nécessite un changement de base de l'espace RGB vers l'espace HSI. Le principe de cet espace est de caractériser les couleurs de façon plus intuitive et de se rapprocher de la perception humaine. Les trois composantes de cet espace sont la Teinte, la Saturation et l'Intensité. Le cône de la figure.4 illustre un modèle de cet espace. Dans ce modèle, la teinte est un angle de 0 degrés à 360 degrés, elle représente la couleur pure. La saturation indique la gamme de gris dans l'espace couleur. Elle varie de 0 à 100%. Parfois, la valeur est calculée à partir de 0 à 1. Lorsque la valeur est 0, la couleur est grise et lorsque la valeur est 1, la couleur est une couleur primaire. L'intensité est la mesure de la luminosité de la couleur, qui doit varier entre le noir et le blanc.

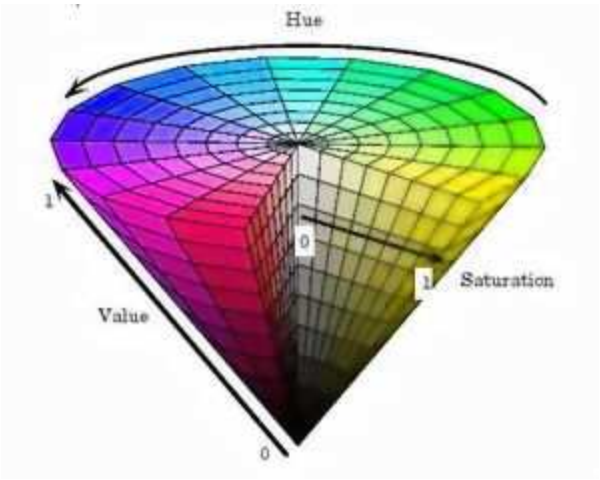


Fig.4. Representation de l'espace couleur HSI

Les avantages de l'espace HSI sont :

- La représentation de la couleur est faite par des parties perceptuellement distinctes.
- L'espace HSI est facilement quantifiable.
- La teinte est l'information la plus signifiante pour la couleur.

Un des principaux inconvénients de cet espace est qu'il n'est pas uniforme, la distance calculée dans cet espace entre les couleurs visuellement proches peut être très grande.

L'espace *CIE XYZ* prend en compte la sensibilité de l'œil. La composante Y représente la luminance, X et Z contient l'information de chrominance. La figure.5 donne le diagramme de cet espace. Il est rarement utilisé en recherche d'images car il n'est pas uniforme de point de vue humaine. De plus, il n'est pas facile d'interpréter les valeurs de tristimulus *X*, *Y* et *Z* et d'interpréter les couleurs qu'il représente [56]. La distance couleur associée à cet espace est la distance euclidienne.

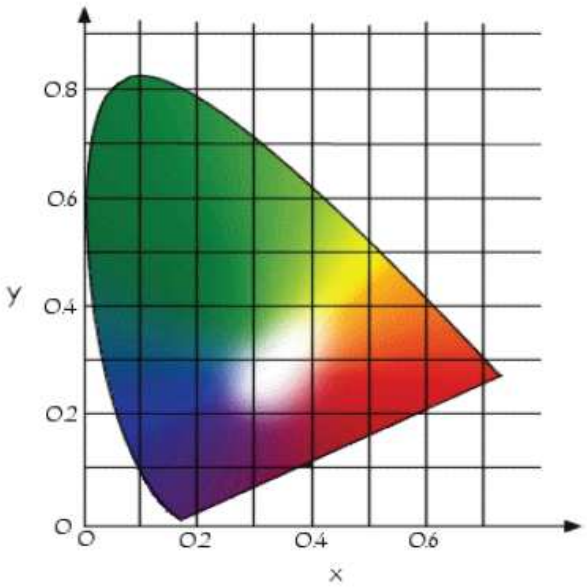


Fig.5. Representation de l'espace couleur XYZ

L'espace *CIE Lab* est une alternative au modèle XYZ. Il caractérise une couleur à l'aide d'un paramètre d'intensité L correspondant à la luminance et les deux autres paramètres de chrominance décrivent la couleur. La composante *a* permet de parcourir l'axe de couleur rouge-vert, pour une gamme de couleurs allant de vert (-128) au rouge (+127), et la composante *b* parcourt l'axe de couleur jaune-bleu, pour une gamme de couleurs allant du bleu (-128) au jaune (+127). La figure.6 représente une illustration de cet espace. L'espace Lab est perceptuellement uniforme, il posséde une bonne propriété de respecter les distances entre couleurs visuellement proches, il est independant vis-à-vis du materiel utilisé. Mais lorsque il s'agit d'estimer avec precision les caractéristiques chromatiques, les variations de couleurs sur les axes *a* ou *b* sont cinq fois moins visibles que les variations de luminance dans les applications de traitement d'images.

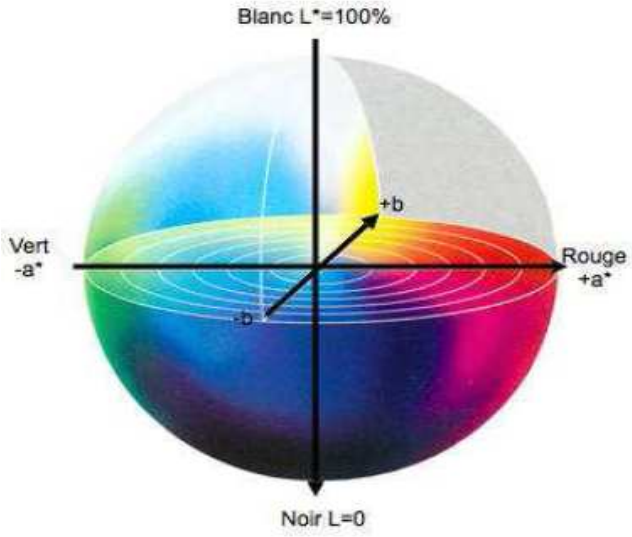


Fig.6. L'espace de couleur CIELab

4.1.2. L'histogramme

Le descripteur de couleurs le plus courant et le plus rencontré dans la littérature est l'histogramme de couleurs. Pour cette raison que nous avons choisi d'utiliser ce descripteur pour la realisation de notre système de recherche d'images et nous en parlerons plus en detail ulterieurement.

4.1.3. Histogramme de couleur-structure CS

Le descripteur de couleur-structure CS étend et enrichit la notion d'histogramme en introduisant dans la représentation un minimum d'information spatiale locale. Un élément structurant, définie un masque binaire, est translaté en chaque pixel de l'image. Tous les intervalles de l'histogramme correspondant aux couleurs présentent à l'interieur du masque sont alors incrémentés. Ainsi, l'histogramme CS représente-t-il la fréquence relative des éléments structurants contenant une couleur donnée. Les mesures de similarités adoptées pour l'histogramme CS sont la distance Minkowski et la distance euclidienne.

Pour assurer une certaine invariance par rapport aux homothéties, les images sont normalisées à une taille fixe avant l'extraction du descripteur.

4.1.4. Les moments statistiques

La méthode d'histogramme utilise la distribution complète de la couleur. On doit stocker de nombreuses données. Au lieu de calculer la distribution complète, dans les systèmes de recherche d'images, on calcule seulement des caractéristiques dominantes de couleur tel que l'espérance, la variance et d'autres moments. Ce descripteur sera vu en détail au troisième chapitre.

4.1.5. Corrélogramme de couleurs

Le corrélogramme de couleurs a été proposé pour qualifier non seulement la distribution de couleurs des pixels, mais aussi la correlation spatiale entre les pairs de couleurs. Ces informations sont représentées sous la forme d'un histogramme à trois dimensions : les deux premières dimensions représentent les combinaisons possibles de pairs de pixels et la troisième dimension représente leurs distances spatiales. Le corrélogramme est un tableau indexé par des pairs de couleurs, où le k^{ieme} entré *(i,j)* spécifie la probabilité de trouver un pixel de couleur j à une distance k d'un pixel de couleur i dans une image. Soit I l'ensemble de pixels d'une image et $I_{c(i)}$ l'ensemble de pixels dont la couleur est $c(i)$. Alors le corrélogramme de couleurs est défini par :

$$\gamma_{i,j}^{(k)} = Pr_{p_1 \in I_{c(i)}, p_2 \in I}\left[p_2 \in I_{c(i)}, |p_1 - p_2| = k\right] \quad (4)$$

Où $i, j \in \{1,2,\dots,N\}, k \in \{1,2,\dots,d\}$ *et* $|p_1 - p_2|$ est la distance entre les pixels p_1 et p_2. Si on considère toutes les combinaisons de paires de couleurs alors la taille du corrélogramme de couleurs est considérable ($O(N^2d)$).

Les corrélogrammes donnent des meilleurs résultats en termes de pouvoir de discrimination. Cependant, son implémentation est très délicate. Même si l'espace mémoire n'est pas forcement plus élevé que celui utilisé par les histogrammes, leur création est très coûteuse en temps de calcul, ce qui nécessite une implémentation très optimisée pour qu'ils soient utilisables avec un temps de calcul raisonnable.

4.1.6. Vecteur de Cohérence de couleurs

Le vecteur de cohérence de couleurs CCV, représente une autre variante, plus détaillée, de l'histogramme de couleurs. Il a été proposé par Pass [57]. Dans cette technique, chaque rang de l'histogramme peut être partitionné en deux catégories :

- Cohérent, s'il appartient à une région de couleur uniforme,
- Incohérent, sinon.

On note α_i le nombre de pixels cohérents dans la $i^{éme}$ rang de couleur et β_i le nombre de pixels incohérents. Le CCV d'une image est défini par le vecteur$[(\alpha_1, \beta_1), (\alpha_2, \beta_2), \dots\dots, (\alpha_N, \beta_N)]$, tel que la somme :$(\alpha_1 + \beta_1, \alpha_2 + \beta_2, \dots, \alpha_N + \beta_N)$ donnera l'histogramme de couleurs de l'image. L'avantage de cette approche réside dans l'ajout de l'information spatiale à l'histogramme et cela à partir de leur raffinement, mais elle présente l'inconvénient d'amplifier la sensibilité aux conditions d'illumination.

4.1.7. Descripteur par couleurs dominantes

Ce descripteur fournit une description compacte des couleurs représentatives dans une image. Il est défini par :

$$F = \{(c_i, p_i, v_i, s)\}, (i = 1,2, \dots, N)$$

Où N désigne le nombre de couleurs dominantes, c_i la valeur de la $i^{ème}$ couleur dominante, p_i un poids exprimant sa fréquence relative d'apparition dans l'image et v_i sa variance. La cohérence spatiale s représente l'homogénéité spatiale des couleurs dominantes dans l'image. Pour le calcul des couleurs dominantes, il faut choisir un espace de couleurs uniforme au niveau de la perception humaine.

Par définition, le descripteur par couleurs dominantes est intrinsèquement invariant aux transformations de similarité.

Considérons les deux descripteurs ci-dessous :

$$F_1 = \{(c_{1i}, p_{i1}, v_{1i}, s_1)\}, (i = 1,2, \dots, N_1)$$

$$F_2 = \{(c_{2i}, p_{2i}, v_{2i}, s_2)\}, (i = 1,2, \dots, N_2)$$

Le degré de similarité entre deux couleurs c_{1k} et c_{2l}, noté $a_{1k,2l}$, est défini par :

$$a_{1k,2l} = \begin{cases} 1 - \frac{\delta(c_{1k},c_{2l})}{\delta_{max}} & si\ \delta(c_{1k},c_{2l}) \leq T_d \\ 0 \end{cases} \quad (5)$$

Où $\delta(c_{1k},c_{2l})$ désigne la distance euclidienne entre deux couleurs $c_{1k}\ et\ c_{2l}$, T_d est un seuil prédéfini.

4.1.8. Distribution spatiale de couleurs

Ce descripteur vise à capturer la disposition spatiale (color layout) des couleurs dans une image. L'image est tout d'abord divisée en 64 (8 × 8) blocs rectangulaires. Chaque rectangle est représenté par sa couleur dominante, qui est par définition la moyenne des couleurs des pixels constituant le rectangle considéré. On construit ainsi trois matrices (8 × 8), une par composante de couleur. Les coefficients quantifiés des transformées en cosinus discrète 2D (DCT - Discrete Cosine Transform) de chacune de ces matrices sont ensuite calculés. Ces coefficients sont parcourus en zigzag, à partir des fréquences les plus basses [58].

La mesure de similarité recommandée pour ce descripteur est une distance euclidienne pondérée entre les coefficients DCT ainsi déterminés. Notons que ce descripteur n'est pas invariant aux rotations. Son utilisation est donc pertinente uniquement pour des applications concernant des requêtes globales, où le positionnement des différents couleurs dans l'image doit être pris en compte.

4.1.9. Cohérence spatiale

Ce descripteur est calculé pour chaque classe couleur identifiée. Un histogramme de connexité est calculé par la formule suivante :

$$H_s(c) = \sum_{i=0}^{Y-1} \sum_{j=0}^{X-1} \delta(I(i,j),c)\alpha(i,j) \quad (6)$$

I est l'image segmentée de taille *(X, Y), c* est la couleur du pixel *(i, j), δ* est le symbole de Kronecker et $\alpha(i,j)$ est défini par :

$$\alpha(i,j) = \begin{cases} 1\ si\ \ \forall k,k' \in [-W,W]\ I\big(i+k,j+k'\big) = I(i,j) \\ 0 \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad sinon \end{cases} \quad (7)$$

La fenêtre $(2W+1)*(2W+1)$ représentant le degré de compacité souhaité. La cohérence spatiale est donnée par le rapport suivant :

$$SCR(c) = \frac{H_s(c)}{H(c)} \quad (8)$$

Où H représente l'histogramme de couleurs et donc $SCR(c) \in [0,1]$. Une faible valeur de $SCR(c)$ expliquera une dispersion de la couleur c dans l'image, tandis que pour une couleur dominante homogène $SCR(c)$ sera proche de 1.

4.2. Les descripteurs de textures

La texture est le second attribut visuel largement utilisé dans la recherche d'images par le contenu. Elle permet de combler un vide que la couleur est incapable de faire, notamment lorsque les distributions de couleurs sont très proches. Fondamentalement la texture est définie comme la répétition d'un motif créant une image visuellement homogène. Plus précisément, la texture peut être vue comme un ensemble de pixels (niveaux de gris) spatialement agencés selon un certain nombre de relations spatiales, ainsi créant une région homogène. De ces définitions, les recherches sur la modélisation des textures se sont portées sur la caractérisation de ces relations spatiales [29]. De nombreuses approches et modèles [59] ont été proposées pour la caractérisation de la texture. Parmi les plus connues, on peut citer : les méthodes statistiques, les méthodes fréquentielles et les méthodes géométriques. Nous introduisons dans ce qui suit quelques représentations de la texture qui sont utilisées dans le domaine de la recherche d'images par le contenu.

4.2.1. Les méthodes statistiques

Ce sont les méthodes basées sur des évaluations quantitatives de la distribution de nivaux de gris. Elles étudient les relations entre un pixel et ses voisins. Elles sont utilisées pour caractériser des structures fines, sans régularité apparente. Plus l'ordre de la statistique est élevé et plus le nombre de pixels allant de *1* à *n* mis en jeu est important. Parmi ces méthodes on peut citer la méthode de la dépendance spatiale des niveaux de gris (SGLDM : Spatial Gray Level Dépendance Method) ou matrices de cooccurrences, caractéristiques de Tamura, la matrice de longueur de

plages et la méthode de différence de niveau gris (GLDM : Gray Level Différence Method).

4.2.1.1. Les matrices de longueur de plages (MLDP)

La MLDP est un descripteur statistique d'ordre supérieur de l'image, qui a surtout été utilisé pour la reconnaissance des textures. Il étudie les interactions entre plusieurs pixels. Une plage est un ensemble de pixels connexes de même valeur, orienté selon une direction imposée. La MLDP regroupe, pour une région spécifiée de l'image, le nombre de plages ayant chaque longueur et valeur possibles. Elle regroupe donc le nombre de pixels successifs, sur une direction imposée, ayant le même niveau de gris [60], [61]. D'habitude on ne considère que les directions essentielles (verticale, horizontale et les deux diagonales principales), pour des raisons de simplicité d'implémentation. Pour la description statistique de la texture, les valeurs de la MDLP sont combinées dans des descripteurs généralistes, qui relèvent des attributs statistiques.

Si on note par N_{iz} le nombre total de plages et par N_{reg} le nombre de pixels dans la région d'analyse de l'image, les descripteurs des LDP sont RF_1 à RF_5 (proportion des petites plages (RF_1), proportion des longues plages (RF_2), hétérogénéité de couleurs (RF_3), hétérogénéité des plages (RF_4) et proportion des plages (RF_5)), définis initialement par Galloway [60], par:

$$N_{iz} = \sum_{a=0}^{L-1} \sum_{b=1}^{n_\theta} M_\theta\,(a,b) \tag{9}$$

$$RF_1 = \frac{1}{N_{iz}} \sum_{a=0}^{L-1} \sum_{b=1}^{n_\theta} \frac{M_\theta(a,b)}{b^2} \tag{10}$$

Si la texture comporte des petites plages de niveau de gris ce paramètre RF_1 aura une grande valeur.

$$RF_2 = \frac{1}{N_{iz}} \sum_{a=0}^{L-1} \sum_{b=1}^{n_\theta} b^2\, M_\theta(a,b) \tag{11}$$

Ce paramètre RF_2 met en évidence les longues plages.

$$RF_3 = \frac{1}{N_{iz}} \sum_{a=0}^{L-1} \Big(\sum_{b=1}^{n_\theta} M_\theta(a,b) \Big)^2 \tag{12}$$

Ce paramètre RF_3 mesure la non-uniformité des niveaux de gris. Quand les plages sont uniformément distribuées pour tous les niveaux de gris, ce paramètre a une valeur très faible.

$$RF_4 = \frac{1}{N_{iz}} \sum_{b=1}^{n_\theta} \left(\sum_{a=0}^{L-1} M_\theta(a,b) \right)^2 \quad (13)$$

Ce paramètre RF_4 mesure la non-uniformité des longueurs de plages. Quand les plages sont uniformément distribuées pour toutes les longueurs de plages en niveaux de gris, ce paramètre est minimum.

$$RF_5 = \frac{N_{iz}}{N_{reg}} \quad (14)$$

Ce paramètre RF_5 a une valeur faible pour les textures de structure homogène.

avec $M_\theta(a,b)$: nombre de plages de pixels de niveau de gris a, de longueur b. θ *:* direction de la plage de niveau de gris. L correspond au nombre de niveaux de gris dans l'image et n_θ à la longueur de la plage maximale.

Deux autres paramètres ont été introduits par la suite : LGRE et HGRE de mesure de la proportion des valeurs faibles et, respectivement importantes. Il est évident que ces paramètres sont issus d'un symétrisassions des paramètres correspondants pour les longueurs de plages, *RF1* et *RF2*.

$$LGRE = \frac{1}{N_{iz}} \sum_{b=1}^{n_\theta} \sum_{a=0}^{L-1} \frac{M_\theta(a,b)}{a^2} \quad (15)$$

$$HGRE = \frac{1}{N_{iz}} \sum_{b=1}^{n_\theta} \sum_{a=0}^{L-1} a^2 \, M_\theta(a,b) \quad (16)$$

D'ailleurs, le principe d'introduction des paramètres *RF1*, *RF2*, *HGRE* et *LGRE* est de pondérer le nombre de plages d'une certaine valeur et longueur, $M_\theta(a,b)$, par un poids qui dépend directement (ou par l'inverse) du carré de la mesure d'intérêt.

4.2.1.2. Matrice de cooccurrence

Dans les années 70 Haralick et al [62] ont proposé une des premières méthodes de caractérisation de texture baptisée matrice de cooccurrence. Cette approche consiste à explorer les dépendances spatiales des textures en construisant d'abord une matrice de cooccurrence basée sur l'orientation et

la distance entre les pixels de l'image. De chacune de ces matrices Haralick a défini 14 paramètres caractéristiques de texture, comme le contraste, l'entropie ou la différence inverse des moments [63]. La réussite de cette méthode repose sur le bon choix des paramètres qui sont : la taille de la matrice sur laquelle s'effectue la mesure, et la distance *d* qui sépare les deux pixels du motif. Ce type d'approche a récemment été appliqué à la recherche d'images. Cette méthode est utilisée dans notre travail, sera vu en détail dans le prochain chapitre.

4.2.1.3. La méthode de différence de niveaux de gris

La méthode de différence de niveaux de gris GLDM permet de calculer le nombre d'apparitions d'une différence de niveaux de gris donnée. Cela revient à calculer des paramètres sur une image de différence entre une image initiale et une image translatée de *d*. La GLDM donne un aspect de la texture au sens de la différence de niveaux de gris entre les pixels. Cette différence des niveaux de gris est définie pour chaque pixel d'une région donnée par :

$$\mathrm{g} = |f(x,y) - f(x+dx, y+dy)| \qquad (17)$$

où $f(x,y)$ est le niveau de gris au point de coordonnée *(x,y)*, et les coordonnées du vecteur déplacement sont décrites par *(dx,dy).*

Dans cette technique, on considère que la distribution des valeurs prises par g pour l'ensemble des pixels appartenant à l'objet caractérise la texture. On résume la distribution de g par les paramètres usuels de la statistique. Parmi ces paramètres on peut citer le contraste, la moyenne, l'entropie, le second moment angulaire.

4.2.1.4. Caractéristique de Tamura

L'approche Tamura et al. [64] est intéressante pour la recherche d'images car elle décrit les textures possibles selon des concepts qui correspondent à la perception visuelle humaine. Les auteurs proposent six propriétés visuelles des textures : la grossièreté, le contraste, la direction, présence de lignes (linéarité), régularité et rugosité. Chacun de ces

paramètres est mesuré pour établir un vecteur de texture. La forte liaison de ces descripteurs, notamment les trois premiers, avec la perception humaine fait de cette représentation des textures un champ intéressant et exploitable pour la recherche d'images par le contenu. Il semblerait que l'œil humain soit le plus sensible à la grossièreté de la texture, puis à son contraste et enfin à la direction. Ce type de caractéristiques peut sembler intéressant pour comparer le contenu visuel des images car il correspond directement à la manière dont l'humain les perçoit. Cependant, les méthodes mises en œuvre pour calculer automatiquement ces caractéristiques n'ont donné que des résultats limités pour la recherche d'images dans des données réelles [65].

4.2.2. Les méthodes fréquentielles

L'une des méthodes de description de la texture les plus utilisées concerne les propriétés fréquentielles et s'appuie sur la transformée de Fourier [66], le filtre de Gabor [67,68], les ondelettes [69 ,70 ,71], etc. Elle repose sur l'analyse d'une fonction de densité spectrale dans un domaine fréquentiel [72]. La texture est définie comme un mélange de signaux de fréquences, d'amplitudes et de directions différentes. Ces méthodes consistent à extraire l'énergie portée par le signal dans diverses bandes de fréquence.

4.2.2.1. Les filtres de Gabor

Les filtres de Gabor sont largement utilisés en indexation, pour la description de la texture. Ils permettent une bonne résolution temporelle à haute fréquence et une bonne résolution harmonique sans grande précision temporelle à basse fréquence [73]. Sommairement, les paramètres de texture sont déterminés en calculant la moyenne et l'écart type des niveaux de gris de l'image filtrée par Gabor. En fait, ce n'est pas une seule valeur de moyenne et d'écart type qui sera calculée, mais plutôt un ensemble de valeurs égal au nombre d'échelles multiplié par le nombre d'orientations utilisées. On aura donc ce qui est parfois appelé la banque de filtre de

Gabor. Mathématiquement, toutes les valeurs des moyennes et d'écarts type calculées seront regroupées dans un seul vecteur descripteur.

Un filtre de Gabor 2-D est produit d'une gaussienne elliptique dans toute rotation et un exponentiel complexe représentant une onde plane sinusoïdale [74]. On rappelle que, dans le domaine spatial, la fonction de Gabor bidimensionnelle est une somme de deux fonctions sinusoïdales, l'une paire et réelle, l'autre impaire et imaginaire, modulée par une enveloppe gaussienne. Le filtrage correspond à une convolution par des filtres de réponse impulsionnelle de la forme suivante:

$$h(x,y) = \mathrm{g}(x',y')exp\left[2\pi j\left(u_0(x-x_0)+v_0(y-y_0)\right)\right] \quad (18)$$

avec :

$$\mathrm{g}(x',y') = exp\left[-\frac{1}{2}\left(\frac{x'^2}{\sigma_x^2}+\frac{y'^2}{\sigma_y^2}\right)\right] \quad (19)$$

où σ_x et σ_y sont des constantes d'espace de l'enveloppe gaussienne qui déterminent l'étendue de l'onde suivant les axes *x* et *y* respectivement et (x_0, y_0), le point d'origine où s'applique la fonction *h(x,y)*. En ce point, la fonction est maximale. Les coordonnées *(x', y')* se déduisent de *(x, y)* par :

$$(x',y') = (xcos\gamma + ysin\gamma, -xsin\gamma + ycos\gamma) \quad (20)$$

où l'angle de rotation γ de *(x', y')* par rapport à *(x,y)* donne l'orientation de l'enveloppe gaussienne dans le domaine spatial. Dans ce domaine, la réponse impulsionnelle d'un filtre de Gabor est donnée par l'expression :

$$H(u,v) = exp[-2\pi^2(\sigma_u^2(u-u_0)^2+\sigma_v^2(v-v_0)^2)] \quad (21)$$

Ainsi, un filtre de Gabor est complètement défini par la connaissance de six paramètres à savoir l'orientation de l'enveloppe gaussienne γ, l'orientation de l'onde sinusoïdale θ, les deux paramètres de positions (u_0,v_0) ou (F_0, θ) et les deux paramètres d'étalement (σ_u^2,σ_v^2) ou bandes radiale et transerverse *B* et Ω tels que :

$$\theta = \mathrm{Arctg}\frac{u_0}{v_0} \quad (22)$$

$$F_0 = \sqrt{{u_0}^2+{v_0}^2}, \quad B = \log_2\frac{\pi\sigma_u F_0+\alpha}{\pi\sigma_u F_0-\alpha} \quad (23)$$

$$\Omega = 2Arctg\frac{\alpha}{\pi\sigma_v F_0},\ \ avec: \alpha = \sqrt{\frac{\log(1/s)}{2}} \tag{24}$$

où s est un seuil de troncature [75].

4.2.2.2. Les ondelettes

La transformée en ondelettes consiste à décomposer un signal en utilisant une famille de fonctions $\psi_{m,n}(x)$ obtenues par la translation et la dilatation d'une fonction mère $\psi(x)$. Les ondelettes (équation 25) sont générées à partir de cette fonction $\psi(x)$.

$$\psi_{m,n}(t) = \frac{1}{\sqrt{m}}\psi\left(\frac{t-n}{m}\right) \quad m > 0, n \in \mathcal{R} \tag{25}$$

Où m est le facteur d'échelle (dilatation), n est le facteur de translation et $\frac{1}{\sqrt{m}}$ est un facteur de normalisation à travers les différentes échelles. On définit alors les coefficients de la transformée en ondelettes d'un signal *s(x)*, comme étant les produits scalaires :

$$C_s(m,n) = \frac{1}{\sqrt{m}}\int_{-\infty}^{+\infty} s(x)\,\overline{\psi\left(\frac{x-n}{m}\right)}\,dx \tag{26}$$

À partir de la transformée en ondelettes on peut extraire des attributs de différents types et à différents niveaux de résolution. L'image d'approximation donne des informations sur les régions qui composent l'image, d'une résolution fine à une résolution grossière. Les images de détails donnent des informations horizontales, verticales et diagonales sur l'image. L'énergie des coefficients d'ondelettes est directement disponible, on la calcule en prenant la somme des carrés des coefficients d'ondelettes. On a ainsi des mesures d'énergie à différents niveaux de résolution.

L'extraction d'attributs caractéristiques de texture a donné aussi lieu à un certain nombre de méthodes d'analyse géométrique. Ces méthodes sont dédiées à l'analyse structurelle basée sur l'identification d'un ou plusieurs motifs et de leur répartition spatiale. Elles permettent de décrire la texture en définissant les primitives et les règles d'arrangement qui les relient. En effet, les textures ordonnées possèdent des primitives qui se répètent dans les images en des positions suivant une certaine loi [76]. A la différence des méthodes statistiques, les méthodes structurelles permettent de synthétiser

des textures en modifiant ces règles d'arrangement. De telles méthodes semblent plus adaptées à l'étude de textures périodiques ou régulières.

4.3. Les descripteur de formes

Au même titre que pour la texture, l'information de forme est complémentaire de celle de la couleur. La forme est généralement une description très riche d'un objet. L'extraction d'attribut géométrique a été le fer de lance de la recherche d'image par le contenu ces dernières années. De nombreuses solutions ont été proposées pour représenter une forme, nous distinguons deux catégories de descripteurs de formes : les descripteurs basés sur les régions et les descripteurs basés sur les frontières.

Les premiers font classiquement référence aux moments invariants et sont utilisés pour caractériser l'intégralité de la forme d'une région. Ces attributs sont robustes aux transformations géométriques comme la translation, la rotation et le changement d'échelle. La seconde approche fait classiquement référence aux descripteurs de Fourier et porte sur une caractérisation des contours de la forme. Nous présentons dans ce qui suit quelques méthodes de description de la forme

4.3.1. Les moments géométriques

Les moments géométriques [77] permettent de décrire une forme à l'aide de propriétés statistiques. Ils représentent les propriétés spatiales de la distribution des pixels dans l'image. Ils sont facilement calculés et implémentés. Par contre, cette approche est très sensible au bruit et aux déformations et le temps de calcul de ces moments est très long.

La formule générale des moments géométriques est donnée par la relation suivante:

$$m_{p,q} = \sum_{p=0}^{m} \sum_{q=0}^{n} x^p y^q f(x,y) \qquad (27)$$

$p+q$ est l'ordre du moment. Le moment d'ordre 0 $m_{0,0}$ représente l'aire de la forme de l'objet.

Les deux moments d'ordre 1 $m_{0,1}$ et $m_{1,0}$, associés au moment d'ordre 0 $m_{0,0}$ permettent de calculer le centre de gravité de l'objet. Les coordonnées de ce centre sont :

$$x_c = \frac{m_{1,0}}{m_{0,0}} \qquad et \qquad y_c = \frac{m_{0,1}}{m_{0,0}} \tag{28}$$

Il est possible de calculer à partir de ces moments l'ellipse équivalente à l'objet. Afin de calculer les axes de l'ellipse, il faut ramener les moments d'ordre 2 au centre de gravité :

$$\begin{aligned} m_{2,0}^g &= m_{2,0} - m_{0,0}x_c^2 \\ m_{1,1}^g &= m_{1,1} - m_{0,0}x_c y_c \\ m_{0,2}^g &= m_{0,2} - m_{0,0}y_c^2 \end{aligned} \tag{29}$$

Puis on détermine l'angle d'inclinaison de l'ellipse α :

$$\alpha = \frac{1}{2} arctan \frac{2m_{1,1}^g}{m_{2,0}^g - m_{0,2}^g} \tag{30}$$

- **Les invariants de Hu**

A partir des moments géométriques, Hu [11] a proposé un ensemble de sept moments invariants aux translations, rotations et changement d'échelle. Ils sont très utilisés dans la littérature pour la description de formes en vue d'une classification ou d'une indexation, mais sont assez sensibles aux bruits. Par ailleurs cette famille de descripteurs n'est ni orthogonale, ni complète (cette méthode sera détaillée dans le chapitre 3).

4.3.2. Les moments orthogonaux

Par opposition aux moments géométriques qui sont définis par rapport à une base quelconque $(x^p y^q)$, les moments orthogonaux, comme leur nom l'indique, sont définis dans une base orthogonale, ce qui évite la redondance des informations portées par chacun des moments.

Les deux types de moments orthogonaux les plus utilisés sont : les moments de Legendre et les moments de Zernike, dont nous donnons les définitions ci-dessous.

❖ Moments de Legendre

Les moments de Legendre sont définis à partir des polynômes du même nom. Ils sont définis dans le carré unité $[-1,1] \times [-1,1]$, ce qui oblige à normaliser l'objet dont on veut calculer ces moments.

Le polynôme de Legendre d'ordre n est donné par :

$$\forall x \in [-1,1], \forall n \in \mathbb{N}, \quad P_n(x) = \frac{1}{2^n n!} \frac{d^n (x^2-1)^n}{dx^n} \tag{31}$$

Les polynômes de Legendre $\{P_n(x)\}$ forment une base complète et orthogonale sur le domaine de définition $[-1,1]$:

$$\forall (x,y) \in [-1,1]^2, \forall (m,n) \in \mathbb{N}^2, \iint_{-1}^{1} P_m(x) P_n(y)\, dx\, dy = \frac{2}{2m+1} \delta_{mn}$$

δ_{mn} représente la fonction de *Kronecker*.

Les moments de Legendre d'ordre N sont donc donnés par :

$$\forall (x,y) \in [-1,1]^2, \forall (p,q) \in \mathbb{N}^2, N = p + q,$$

$$L_{pq} = \frac{(2p+1)\times(2q+1)}{4} \iint_{-1}^{1} P_p(x) P_q(y)\, f(x,y) dx\, dy \tag{32}$$

Où f(x, y) le niveau de gris d'un pixel de l'image I sur laquelle on calcule le moment. A partir de cette équation, on peut générer une infinité de moments de Legendre. Plusieurs études sur la reconnaissance des formes ont démontré que l'utilisation des moments de Legendre de bas ordre (jusqu'à l'ordre 3) est suffisante pour représenter la forme globale de l'entité donnée.

❖ Les moments de Zernike

Ce type des moments a été initialement introduit par Teague [78] et qui sont construits à partir de polynômes complexes et forment un ensemble orthogonal complet définie sur le disque unité. Ils sont invariants par rotation et changements d'échelles et présentent des propriétés intéressantes en termes de résistance aux bruits, efficacité informative et possibilité de reconstruction.

Les moments orthogonaux de Zernike d'ordre p sont définis de la manière suivante :

$$A_{mn} = \frac{m+1}{\pi} \iint I(x,y) \left[V_{m,n}(x,y)\right] dx\, dy \tag{33}$$

Où m et n définissent l'ordre du moment et $I(x, y)$ le niveau de gris d'un pixel de l'image I sur laquelle on calcule le moment.

Les polynômes de Zernike $V_{m,n}(x, y)$ sont exprimés en coordonnées polaires :

$$V_{m,n}(r, \theta) = R_{m,n}(r)e^{-jn\theta} \quad (34)$$

Où $R_{m,n}(r)$ est le polynôme radial orthogonal :

$$R_{m,n}(r) = \sum_{s=0}^{\frac{m-|n|}{2}} (-1)^s \frac{(m-s)!}{s!\left(\frac{m+|n|}{2}-s\right)!\left(\frac{m-|n|}{2}-s\right)!} r^{m-2s} \quad (35)$$

avec $n = 0,1,2 \ldots \infty$; $0 \leq |m| \leq n$ *et* $n - |m|$un entier pair.

Les polynômes de Zernike sont orthogonaux, et donc les moments correspondants le sont également. Cette propriété d'orthogonalité annule l'effet de redondance de l'information portée par chaque moment.

4.3.3. Descripteurs de Fourier

Les Descripteurs de Fourier DFs font partie des descripteurs les plus populaires pour les applications de reconnaissance de formes et de recherche d'images. Ils ont souvent été utilisés par leur simplicité et leurs bonnes performances en terme de reconnaissance [79] et facilitent l'´etape d'appariement. De plus, ils permettent de décrire la forme de l'objet à différents niveaux de détails. Les descripteurs de Fourier sont calculés à partir du contour des objets. Leur principe est de représenter le contour de l'objet par un signal 1D, puis de le décomposer en séries de Fourier. Les DFs sont généralement connus comme une famille de descripteurs car ils dépendent de la façon dont sont représentés les objets sous forme de signaux.

5. Conclusion

Nous avons détaillé dans ce chapitre les points clés de la recherche d'images par le contenu, à savoir l'extraction et l'indexation de caractéristiques et la construction de signatures. Nous avons indiqué que la recherche débute le plus souvent par une requête sous plusieurs formes. Nous avons ensuite vu qu'il existe une multitude de descripteurs utilisés pour la représentation du contenu visuel de l'image qui sont regroupés en

général suivant les catégories couleur, texture ou forme. Le choix de ces attributs constitue la première étape de la recherche d'images par le contenu et est déterminant pour la qualité des résultats. Cependant, d'une part il n'y a pas d'attributs universels, et d'autre part le choix des descripteurs dépend fortement de la base d'images à utiliser et des connaissances à priori qu'on peut avoir sur la base. En effet, quelle que soit la mesure de distance utilisée ensuite pour rechercher des images, cette distance est en fonction des caractéristiques choisies et les résultats vont dépendre de celles-ci.

Chapitre 3

Implémentation et évaluation expérimentale

1. Introduction

L'objectif de ce chapitre est d'étudier, de définir et de mettre en place un système d'indexation et de recherche d'images par le contenu. Ce chapitre donne une vue plus détaillée sur les techniques utilisées dans notre travail. Nous abordons plus précisément trois aspects du problème. Tout d'abord, nous verrons la technique utilisée pour mesurer la similarité entre les images à partir des vecteurs descripteurs résultants de l'étape d'extraction des caractéristiques. Ensuite nous nous attachons à capitaliser de l'information sur une image avant même de rechercher une similarité. Dans ce contexte, nous allons présenter les méthodes d'analyse d'images permettant de décrire le contenu visuel que nous avons retenues pour cette application. Ce contenu est généralement représenté par des descripteurs de bas niveau. Enfin, nous appliquons l'ensemble de ces techniques à des bases de données d'images numériques afin de pouvoir évaluer notre système de recherche d'images.

2. Fonction de similarité

Un système de recherche d'images par le contenu visuel calcule la similarité visuelle entre le descripteur de l'image requête et les descripteurs des images de la base. Différent types de mesure de similarité sont présentées dans la littérature. Le premier type correspond aux distances géométriques entre vecteurs. Dans ce cas, on parle de distances car ces mesures ont la propriété de respecter les axiomes des espaces métriques. Un espace métrique E se définit par Maurice Fréchet (1878-1973) comme un ensemble non vide doté d'une application d, appelée distance, de $E \times E$ dans R^+ vérifiant les axiomes suivants :

$\forall x, y, z \in E$

1. $d(x,y) = 0 \Leftrightarrow x = y (Identité)$
2. $d(x,y) = d(x,y) (symétrie)$
3. $d(x,y) + d(y,z) \geq d(x,z) (Inégalité\ triangulaire)$

Notre choix est porté sur la distance Euclidienne. L'idée de cette distance, c'est que chaque image qui a *N* caractéristiques est un point dans l'espace de *N* dimension. Chaque caractéristique est un vecteur de *V(k)* de cet espace.
La distance entre deux images, c'est la distance entre deux points de cet espace, elle est donnée par la formule suivante :

$$Dist(Im_{req}, Im_{cour}) = \left(\sum_{k=1}^{K} \left|V(k)_{Imreq} - V(k)_{Imcour}\right|^2\right)^{1/2} \qquad (1)$$

Où : Im_{req} est l'image requête et Im_{cour} est l'image courante

$V(k)_{Imreq}$ le vecteur descripteur de l'image requête

$V(k)_{Imcour}$ le vecteur descripteur de l'image courante.

3. les descripteurs utilisés

Comme on l'a déjà vu au chapitre précédent, la comparaison directe des images entre elles n'est pas envisageable. Il est donc nécessaire d'en extraire au préalable des informations représentatives : les descripteurs. Ces derniers sont des mesures de caractéristiques de l'image qui doivent être invariantes en rotation, translation et changement d'échelle. Les caractéristiques utilisées sont la couleur, la texture et la forme, nous allons les détaillés précisément:

Pour la couleur, nous calculons l'histogramme de couleurs et les moments statistiques, dans l'espace de couleurs RGB.
Pour la texture, nous calculons la matrice de cooccurrences et nous créons un vecteur descripteur en extrayant les quatre paramètres les plus appropriées des matrices : l'énergie, le contraste, l'entropie et le moment inverse de différence.
Pour la forme, nous calculons les moments de Hu. Nous combinons les trois caractéristiques (couleur, texture et forme).
Ces descripteurs choisis sont adaptés à la nature des images contenus dans les bases.

3.1. L'histogramme

Comme on l'a vu dans le chapitre 2, l'information couleur est la plus importante caractéristique utilisée pour la description du contenu visuel. Chaque pixel d'image peut être présenté comme un point dans un espace de couleurs 3D. Il est donc nécessaire d'utiliser des espaces couleurs bien adaptés au calcul de distance entre couleurs. Le modèle adapté pour ce travail est l'espace RGB. Pour extraire le premier vecteur descripteur de la couleur, nous avons choisi l'histogramme de couleurs. Ce choix est justifié par le fait que ce dernier est simple à calculer, invariable selon la translation ou la rotation de l'axe de vue. Ce qui fait que l'histogramme de couleurs est un outil particulièrement intéressant pour la recherche d'objets ou d'images ayant une position et une rotation inconnue par rapport à la scène.
Un histogramme de couleurs (ou distributions de couleurs) est obtenu en comptant, pour chaque couleur, le nombre des pixels de cette couleur contenus dans l'image. Il a été introduit en tant que descripteur par Swain et Ballard [22]. Etant donnée une image I, de taille $M \times N$ pixels, caractérisée pour chaque pixel (i,j) par une couleur c appartenant à l'espace de couleurs C (c'est à dire $c = I(i,j)$), alors l'histogramme h est un vecteur à n composantes : $(h_{c1}, h_{c2}, \ldots\ldots, h_{cn})$ pour lequel h_{cj} représente le nombre de pixels de couleur c_j dans l'image I. On a :

$$\sum_{i=1}^{n} h_{ci} = N \qquad (2)$$

où N est le nombre de pixels de l'image.

Ainsi, un histogramme contient, pour chaque couleur de l'espace, le nombre de pixels de l'image qui sont de cette couleur. En divisant par la surface de l'image (MN), on obtient la probabilité de chaque couleur d'être associée à un pixel donné. Cette normalisation permet d'avoir une invariance par changement d'échelle.

- *Calcul de la distance entre les couleurs de deux images différentes*

Une technique assez classique consiste à utiliser la distance d'Euclidienne pour la mesure de la similarité entre histogrammes. Elle est donnée par la formule suivante :

$$Dist(H_{req}, H_{cour}) = \left(\sum_{k=1}^{K} \left|h(k)_{req} - h(k)_{cour}\right|^2\right)^{1/2} \quad (3)$$

Où H_{req} est l'histogramme de l'image requête et H_{cour} est l'histogramme de l'image courante.

3.2. Les moments statistiques

Afin d'échapper aux effets de la discrétisation de l'espace de couleurs qui est intrinsèque à l'utilisation des histogrammes.Une autre approche a été utilisée avec succès dans plusieurs systèmes de recherche d'images est les moments statistiques de couleurs. Cette approche consiste à calculer une somme pondérée de la moyenne, de la variance et du moment du troisième ordre pour chaque canal de couleur, pour fournir un nombre unique utilisé pour indexer .

Si p_{ij} est la valeur du pixel j pour le canal i *(i :RGB),* N est le nombre des pixels dans l'image, ces moments sont definis par :

$$\mu_i = \frac{1}{N}\sum_{j=1}^{N} p_{ij} \quad (4)$$

$$\sigma_i = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(p_{ij} - \mu_i\right)^2} \quad (5)$$

$$s_i = \left(\frac{1}{N}\sum_{j=1}^{N}\left(p_{ij} - \mu_i\right)^3\right)^{\frac{1}{3}} \quad (6)$$

Le moment d'ordre un μ_i représente la couleur moyenne de l'image. Le moment d'ordre deux σ_i caractérise le contraste d'une image. Plus la variance des couleurs est grande, plus l'image est contrastée. Le moment d'ordre trois s_i caractérise la quantité de lumière dans une image. Une image avec un coefficient de dissymétrie positif a tendance à apparaître plus sombre et plus brillante qu'une image semblable avec un coefficient de dissymétrie inférieur.

- *Calcul de la distance entre deux images différentes*

La distance entre l'image requête I et une image H de la base en utilisant les moments statistiques est donnée par la relation suivante :

$$Dist_{Mom} = \left(\sum \left|\left(\mu_i(I) - \mu_i(H)\right) + \left(\sigma_i(I) - \sigma_i(H)\right) + \left(s_i(I) - s_i(H)\right)\right|^2\right)^{1/2} \quad (7)$$

3.3. La matrice de cooccurrence

La texture, autre caractéristique pour la description du contenu visuel, donne une description de la structure locale ainsi qu'une information sur le champ aléatoire qui contrôle les détails à petite échelle. Elle décrit habituellement ce qui est laissé après élimination de l'information de couleur et de forme. Bien que plusieurs techniques aient été proposées pour caractériser la texture aucune ne ressort comme un descripteur universel : chacune présente des avantages et des inconvénients qui dépendent notamment du domaine d'application. Nous avons implémenté la méthode de la matrice de cooccurrence à niveau de gris pour extraire les indices de textures, car elle est très utilisée en recherche d'images, et en général donne de bons résultats.

On rappels que la matrice de cooccurrence à niveau de gris (GLCM) a été proposée par Haralick et al. [63] en 1973. Cette approche est basée sur la probabilité jointe de la distribution des pixels dans l'image [80] [81]. L'élément $p_{d,\theta}(i,j)$ de la matrice de cooccurrence définit la fréquence d'apparition des couples de niveaux de gris i et j pour les couples de pixels séparés par une distance d selon la direction θ. Cette matrice décrit les régularités observables dans les niveaux de gris des pixels d'une région. Le

calcul de la matrice de cooccurrence nécessite le choix d'une distance et d'un angle de déplacement. Il a été noté par plusieurs chercheurs [82, 83] que la distance d'un pixel combinée avec des angles respectifs de 0°, 45°, 90° et 135° donne de bons résultats. C'est la solution que nous avons adoptée, ce qui nous donne à la fin quatre matrices de cooccurrence pour chaque image. Les matrices sont ensuite normalisées par la formule suivante :

$$P_{d,\theta}(i,j) = \frac{p_{d,\theta}(i,j)}{M \times N} \quad (8)$$

Pour une image de taille $M \times N$.

Afin d'estimer la similarité entre les matrices de cooccurrences, Haralick a proposé 14 caractéristiques statistiques extraites à partir de cette matrice. Actuellement, seulement les quatre caractéristiques les plus appropriées sont largement utilisées, et que nous avons calculés pour notre application sont : l'énergie, l'entropie, le contraste et le moment inverse de différence.

- **L'énergie**

$$ENE = \sum_i \sum_j \left(P_{d,\theta}(i,j)^2\right) \quad (9)$$

Ce paramètre mesure l'uniformité de la texture. Il atteint de fortes valeurs lorsque la distribution des niveaux de gris est constante ou de forme périodique. Dans ce dernier cas, les valeurs élevées d'énergie sont obtenues pour les matrices $P_{d,\theta}(i,j)$ lorsque d, θ correspond à la période.

- **L'entropie**

$$ENT = -\sum_i \sum_j \left(\log P_{d,\theta}(i,j) P_{d,\theta}(i,j)\right) \quad (10)$$

Ce paramètre mesure le désordre dans l'image. Contrairement à l'énergie, l'entropie atteint de fortes valeurs lorsque la texture est complètement aléatoire (sans structure apparente). Elle est fortement corrélée (par l'inverse) à l'énergie.

- **Le contraste**

$$CONT = \sum_i \sum_j \left((i-j)^2 P_{d,\theta}(i,j)\right) \quad (11)$$

Il mesure les variations locales des niveaux de gris. Il permet de caractériser la dispersion des valeurs de la matrice par rapport à sa diagonale principale. Ce paramètre est fortement non corrélé à l'énergie.

- **Le moment inverse de différence**

$$MID = \sum_i \sum_j \left(\frac{p_{d,\theta}(i,j)}{1+(i-j)^2} \right) \quad (12)$$

Ce paramètre mesure l'homogénéité de l'image. Il est corrélé à une combinaison linéaire des variables énergie et contraste.

➢ *Calcul de la distance entre les textures de deux images différentes*

Pour mesurer la similarité entre l'image requête et les images de la base, la formule suivante a été utilisée:

$$Dist(im_{Req}, im_{Cour}) = \frac{1}{m}\left(\sum_{i=1}^{m} \left| param_i(im_{Req}) - param_i(im_{Cour}) \right|^2 \right)^{1/2} \quad (13)$$

Où im_{Req} est une image requête, im_{Cour} est une image de la base, m représente le nombre total de paramètres (l'énergie, l'entropie, le contraste et le moment inverse de différence).

3.4. Les moments invariants de Hu

La forme est une caractéristique visuelle importante. Elle présente un des attributs de base pour décrire le contenu visuel d'une image. Pour extraire l'information pertinente de la forme, nous avons utilisé les moments invariants de Hu. Ces moments permettent de décrire la forme à l'aide de propriétés statistiques. Ils sont simples à manipuler, et sont robustes aux transformations géométriques comme la translation, la rotation et le changement d'échelle. Plusieurs techniques ont été développées pour la caractérisation et la représentation d'objets par ces moments. Hu [11] a défini sept moments invariants. La formulation générale de ces moments géométriques dans le domaine continu est donnée par l'équation suivante :

$$M_{pq} = \int \int x^p y^q f(x,y)\, dx\, dy \quad (15)$$

où x et y sont des variables indépendantes d'une fonction f quelconque.
Pour des images numérisées le moment d'ordre $(p+q)$ est donné par :

$$m_{pq} = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} i^p j^q f(i,j); p,q = 0,1,\dots,\infty \quad (16)$$

où M et N sont respectivement la dimension horizontale et verticale de l'image et $f(i,j)$ l'intensité du point (i, j) dans l'image.

Le moment d'ordre *0*, noté m_{00} est une valeur représentant la surface de l'objet et le rapport des moments d'ordre *1*, notés m_{01}, m_{10}, et m_{00} définissent le centre de gravité de la surface de l'objet. Il est calculé par l'expression suivante :

$$x_g = \frac{m_{10}}{m_{00}};\ y_g = \frac{m_{01}}{m_{00}} \tag{17}$$

Ces moments de base sont d'utilité limitée puisqu'ils varient à chaque changement de l'origine, échelle et orientation de l'objet. Un ensemble de moments invariants serait plus utile. Ceci peut être dérivé en calculant d'abord les moments centrés par l'équation suivante :

$$\mu_{pq} = \sum_i \sum_j (i - x_g)^p \, (j - y_g)^q f(i,j) \tag{18}$$

Les moments centrés sont utilisés pour le calcul des moments centrés normalisés : Ils sont calculés par l'expression suivante :

$$\eta_{pq} = \frac{\mu_{pq}}{s^{\frac{p+q}{2}+1}}; p + q \geq 2 \tag{19}$$

où *s* représente la surface de l'objet.

A partir des moments centrés normalisés, nous avons calculé un ensemble de sept paramètres invariants. Ces 7 moments invariants sont :

- **Invariants du second ordre**

$\Phi_1 = \eta_{20} + \eta_{02}$

$\Phi_2 = (\eta_{20} - \eta_{02})^2 + 4{\eta_{11}}^2$

- **Invariants du troisième ordre**

$\Phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$

$\Phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$

$\Phi_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2$
$+ (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$

$\Phi_6 = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})$

$\Phi_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2]$
$+ (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$

Dans notre travail, ces invariants sont extraits à partir des images binaires. Le vecteur $V = (\Phi_1, \Phi_2, \Phi_3, \Phi_4, \Phi_5, \Phi_6, \Phi_7)$ constitue un vecteur d'index final.

- *Le Calcul de la distance entre la forme de deux images différentes*

$$Dist(V_{req}, V_{cour}) = \left(\sum_{i=1}^{7} |V(i)_{req} - V(i)_{cour}|^2\right)^{1/2} \quad (20)$$

4. Combinaison des descripteurs

Les attributs de bas niveau (couleur, texture et forme) décrivent les images par leur contenu visuel. Ces attributs peuvent être combinés pour caractériser de façon encore plus efficace le contenu. Pour fusionner les différents descripteurs une méthode utilisée dans différents travaux consiste à pondérer les distances de similarité. Ainsi les descripteurs pertinents pour la recherche se verront attribuer des poids les plus importants, et auront donc plus d'influence dans le résultat final. La combinaison des descripteurs est donnée par la relation suivante :

$$Dist_{glb} = w_1 Dist_{Histo} + w_2 Dist_{M.statistique} + w_3 Dist_{Texture} + w_4 Dist_{Forme}$$

Où w_i un poids qui prend une valeur entre *0* et *1*.

5. Protocole d'évaluation

La recherche d'images par le contenu est un sous-domaine de la recherche d'informations (RI). Les mesures d'évaluation adoptées ont donc logiquement été reprises de la RI classique. Les deux informations largement utilisées sont le rappel et la précision. Ces paramètres mesurent la pertinence d'un système RI. Autrement dit, ils mesurent la concordance des informations retournées relatives à la requête. La figure.8 montre une illustration de la précision et du rappel.

5.1. Le rappel (Recall)

Le rappel est le rapport entre le nombre d'images pertinentes dans l'ensemble des images trouvées et le nombre d'images pertinentes dans la base d'images.

$$Rappel = \frac{|R_a|}{|R|} \in [0,1]$$

5.2. La précision (precision)

La précision est le rapport entre le nombre d'images pertinentes dans l'ensemble des images trouvées et le nombre d'images trouvées.

$$Precision = \frac{|R_a|}{|A|} \in [0,1]$$

Où R est l'ensemble d'images pertinentes dans la base d'images utilisée pour évaluer (R est l'ensemble des images similaires à la requête), $|R|$est le nombre d'images pertinentes dans la base d'images, A est l'ensemble des réponses et $|R_a|$est le nombre d'images pertinentes dans l'ensemble des réponses.

Par exemple, si la base contient 50 images d'une catégorie jardin, et si le système renvoie 5 bonnes réponses parmi les 10 images qu'il peut afficher, le rappel vaut *(5/50)=0,1* et la précision *(5/10)=0,5*.

Les deux métriques *rappel* et *précision* s'utilisent conjointement pour l'évaluation des performances des systèmes de recherche d'information et varient inversement: lorsque la précision diminue, le rappel augmente et réciproquement. Les valeurs de ces deux métriques reflètent le point de vue de l'utilisateur:

- si le rappel est faible, une partie de l'information pertinente ne lui sera pas accessible;
- si la précision est faible, l'utilisateur ne sera pas satisfait à cause de la forte concentration des informations non-pertinentes fournies dans les résultats.

Dans les deux cas, le système ne répond pas aux attentes des utilisateurs à retourner l'information utile et pertinente et par la suite, il est non-performant. Le cas idéal est d'avoir la valeur de précision et rappel respectivement égale à un, chose qui n'est pas atteinte en réalité.

Une des manières de tenir compte à la fois du rappel et de la précision d'un système est d'exprimer les valeurs de précision en fonction des différents niveaux de rappel selon la courbe rappel/précision.

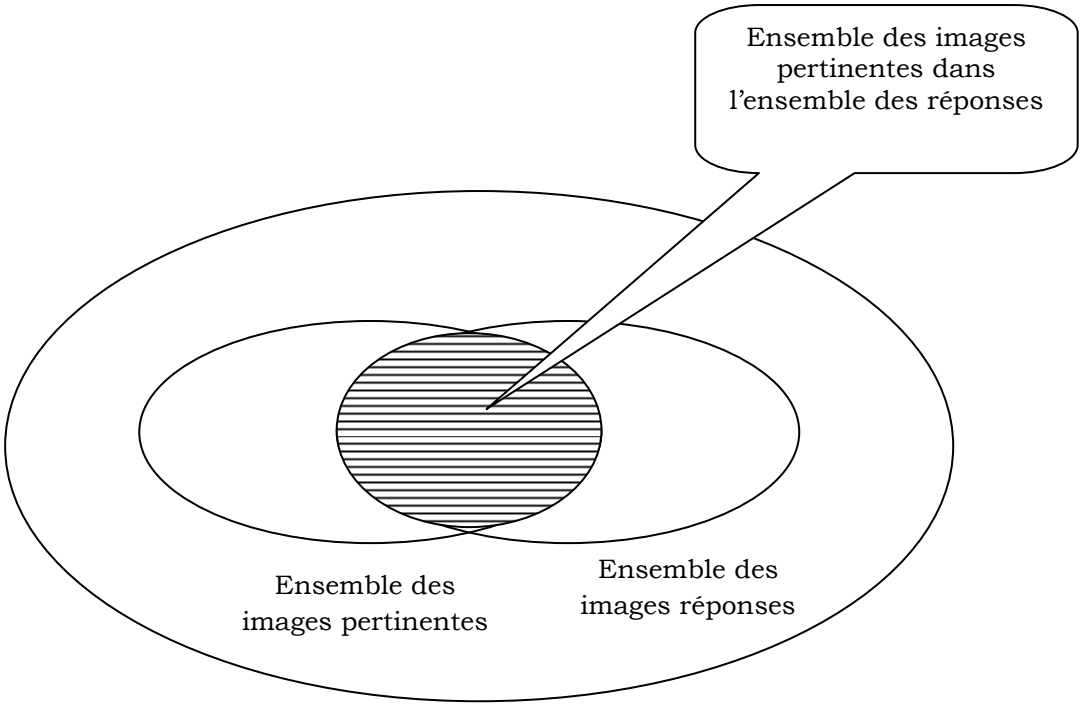


Fig.8. La précision et le rappel pour la recherche.

5.3. Courbe rappel/précision

L'évaluation des performances d'un système CBIR est une tâche ardue car elle dépend étroitement de la perception visuelle de l'humain, du domaine d'application et du contenu de la base elle-même. La courbe de *précision/ rappel* [84, 85, 86, 87] a été utilisée pour évaluer la performance notre système. Il s'agit d'un calcul statistique sur la base de données mettant aussi en évidence la capacité de la méthode étudiée à retrouver les classes d'images. Il ne faut prendre en compte que les images classables. Cependant, cet outil nécessite un grand nombre de requêtes pour évaluer un système. La courbe rappel/précision permet de suivre la qualité du résultat en fonction du nombre d'images retournées par le système en réponse d'une requête. En pratique, cette courbe a l'allure générale de la figure.9.

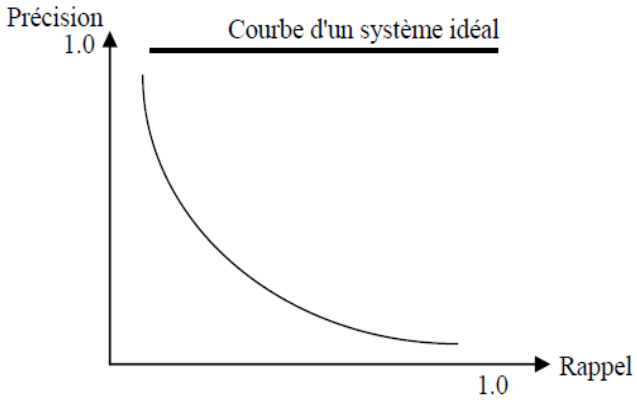


Fig.9. Allure d'une courbe de rappel-précision

6. Description des données

Nous travaillons principalement sur quatre bases d'images dans le but d'évaluer et de valider notre système d'indexation et de recherche d'images par le contenu. Ces bases sont des bases standard utilisées par la communauté scientifique lors des tests et validations de leurs approches et elles sont disponibles sur Internet librement. Chaque base d'images possède des classes définies où chaque image n'appartient qu'à une seule classe.

Nous avons commencé par la base **Coil (Columbia Object Image Library).** Il y a deux bases d'images COIL [88]: COIL-20 qui se compose des images en niveaux de gris représentant 20 objets différents et COIL-100 qui contient des images en couleurs représentant 100 objets différents. La base COIL-100 est formée de 7200 images couleurs (100 objets x 72 images/objet), chaque image est de taille 128×128 pixels. La base COIL-20 est formée de 1440 images en niveaux de gris (20 objets x 72images/objet). Chaque image a une taille de 128×128 pixels. Cette base possède la caractéristique d'être très utilisée en indexation d'images. Un aperçu de cette base est présenté par la figure suivante :

Fig.10. Les objets de la base de Coil 100.

Nous avons également utilisé une base d'images médicale **CEREBRAL STANDARD MR**. Cette base se compose de 400 images classées en plusieurs classes. La figure.11 présente un extrait de cette base.

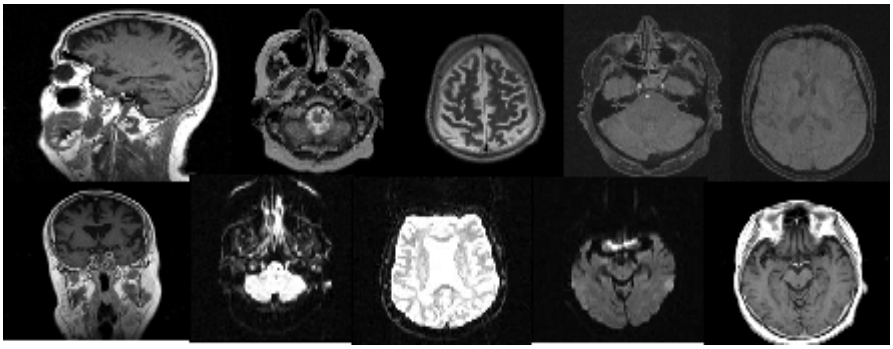
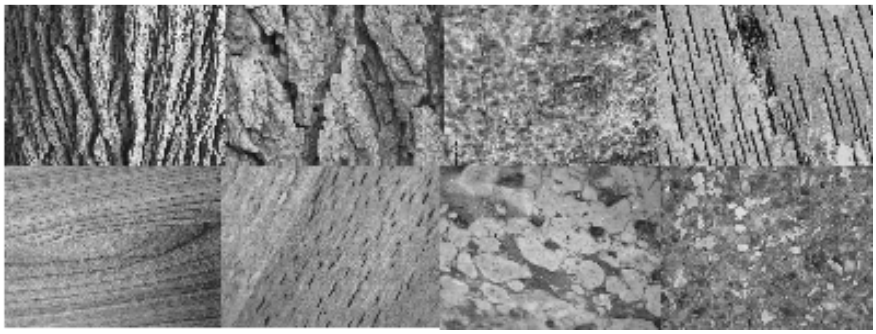


Fig.11. Quelques exemples des images de la base Cérébral standard MR.

Pour évaluer les performances de notre système, des tests de simulations ont été mené sur la base de visages internationale **ORL** [89] (Olivetti Research Laboratory) à Cambridge. La base ORL comprend 40 personnes ayant chacune 10 vues différentes. Ces images sont de dimensions (112*92) pixels. Un extrait de cette base est représenté à la figure.12.

Fig.12. Les 40 personnes de la base ORL.

Nous avons aussi testé une autre base **d'images texturées** [90] de 400 images. La figure suivante présente quelques exemples d'images de cette base.



Fi.13. Quelques exemples des images de la base d'images texturées.

Notons que toutes les méthodes présentées dans notre travail ont été implémentées sous l'environnement visuel C++ 2008 sur micro ordinateur ayant une fréquence de 1,80GHZ, mémoire vive de 1Go et disque dur de 150Go.

7. Quelques résultats

Nous présentons dans cette partie quelques exemples de résultats expérimentaux obtenus à partir des méthodes étudiées. Les résultats seront présentés selon deux formes :

-L'affichage des douze premières images retrouvées pour quelques requêtes.

-Les courbes de rappel/précision, afin d'avoir une représentation visuelle global de la qualité de notre système de recherche d'images.

Les figures14 et 15 présentent les résultats de recherche par similarité dans la base d'images médicale. Les images résultats correspond aux 12 images les plus similaires à la requête, sont triées et affichées, selon le score, de gauche à droite et de haut en bas.

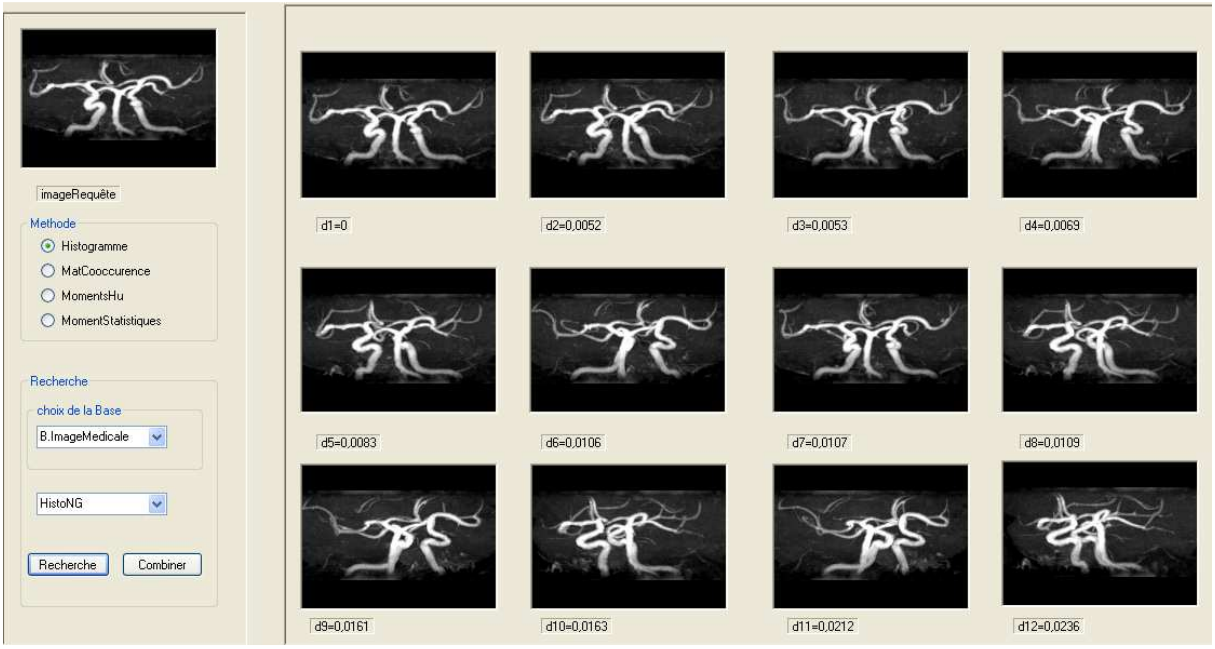


Fig.14. Résultats de la recherche en utilisant l'histogramme

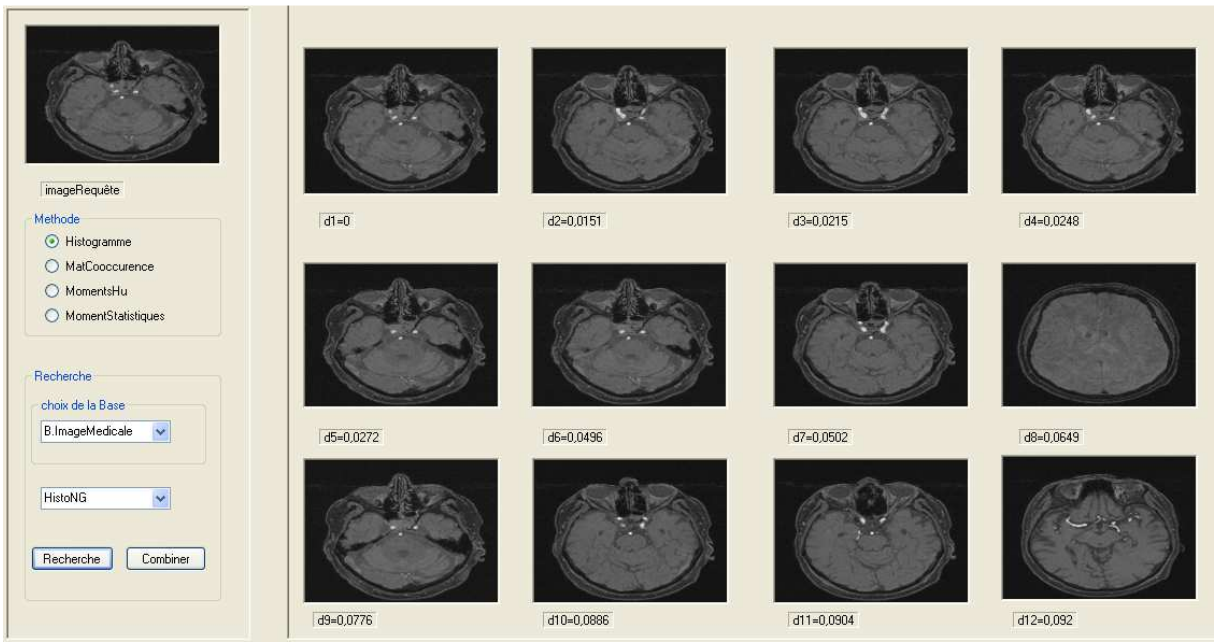


Fig.15. Résultats de la recherche en utilisant l'histogramme

A partir de ces figures, nous constatons que ces exemples sont une bonne illustration du bon fonctionnement de notre système sur cette base. Cela peut être interprété par la spécificité du domaine.

Les figures 16, 17, 18 et 19 illustrent les résultats d'une même image requête dans la base Columbia avec des options de caractéristiques différentes.

Fig.16. Résultats de la recherche en utilisant l'histogramme RGB

Fig.17. Résultats de la recherche en utilisant l'histogramme les moments statistiques

Fig.18. Résultats de la recherche en utilisant les moments de Hu

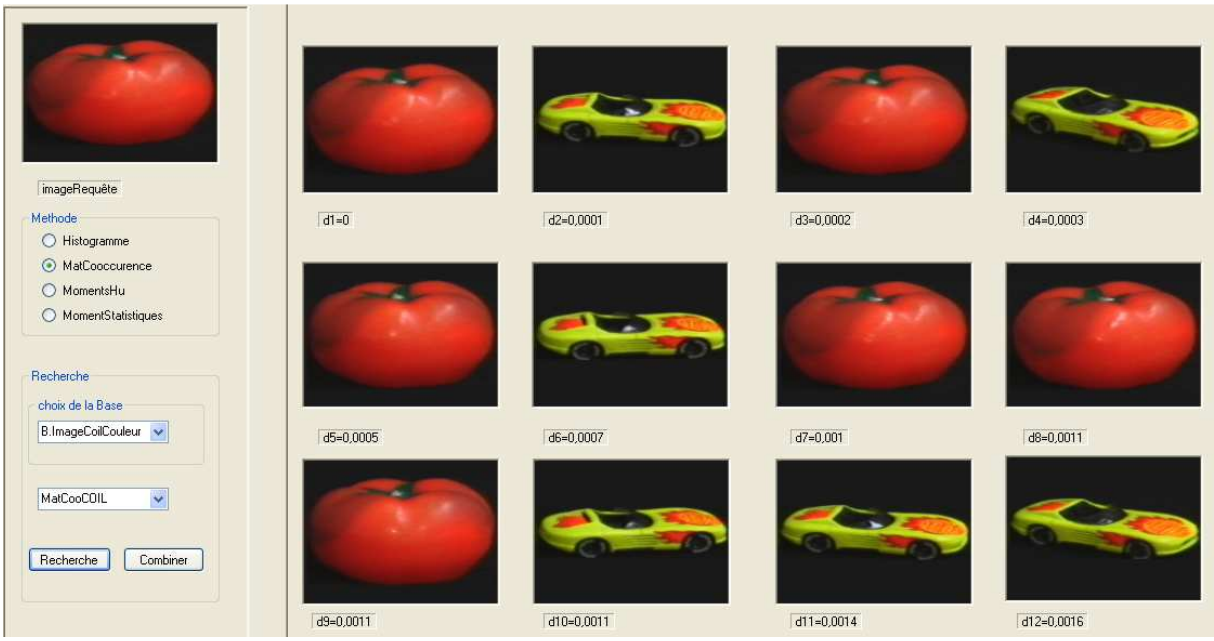


Fig.19. Résultats de la recherche en utilisant la matrice de cooccurrence

Sur les figures.18 et 19, nous constatons que quelques images non pertinentes se sont affichées (4 images pertinentes sont retrouvées avec les moments de Hu et 6 images avec la matrice de cooccurrence). Par contre, sur les figures.16 et 17 toutes les images retrouvées sont pertinentes. On peut conclure que l'histogramme RGB et les moments de couleurs donnent des résultats cohérents et perceptiblement satisfaisants, pour cette image requête, par rapport au descripteur de Hu qui ne semble pas adapté.

La figure.20 illustre un exemple de recherche avec le descripteur histogramme en niveau de gris testé sur la base ORL. Nous constatons que le résultat est satisfaisant. Huit images pertinentes sont retrouvées sur dix.

Fig.20. Résultats de la recherche en utilisant l'histogramme en niveau de gris

La figure.21 présente les résultats obtenus avec la matrice de cooccurrence testée sur une base d'images texturées. Pour cette image requête le résultat est satisfaisant.

Fig.21. Résultats de la recherche en utilisant la matrice de cooccurrence

8. Mesure de la qualité des réponses

A partir des résultats présentés dans la section précédente, nous pouvons calculer la table de rappel et précision. Par exemple, considérons la requête 44.png pour laquelle 12 images sont pertinentes dans la base. Soit la liste des réponses du système {im1, im2,..,im12}. Les images pertinentes sont marquées par la lettre P comme indiquer le tableau suivant :

Image	Pertinent	Précision	Rappel
Im1	P	1	0,08
Im2	P	1	0,16
Im3	P	1	0,25
Im4	P	1	0,33
Im5	P	1	0,41
Im6	P	1	0,5
Im7	P	1	0,58
Im8	P	1	0,66
Im9	P	1	0,75
Im10	P	1	0,83
Im11	P	1	0,91
Im12	P	1	1

Tab.1. Exemple de valeur de rappel-précision

On considère d'abord la première image Im1 restituée par le système. A ce point, on a retrouvé une image pertinente parmi les 12 existantes. Donc on a un taux de rappel de 0,08. La précision est 1/1. Le point de la courbe est donc (0.08, 1).

Image.44.png

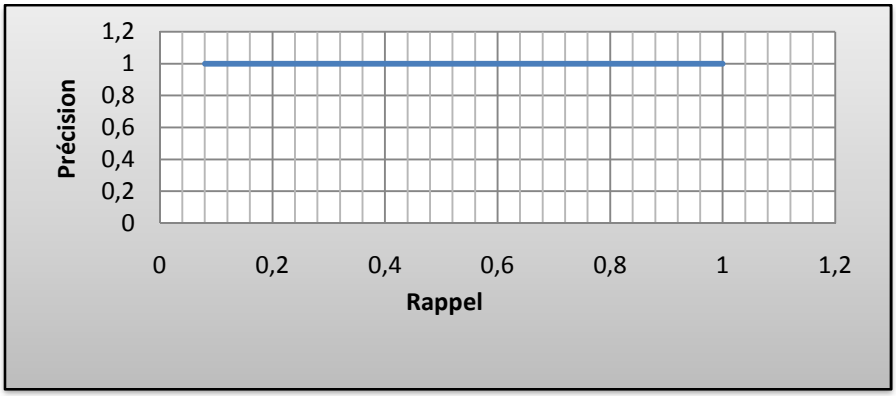


Fig.22. Courbe de rappel/précision pour l'image 44.png avec l'histogramme RGB

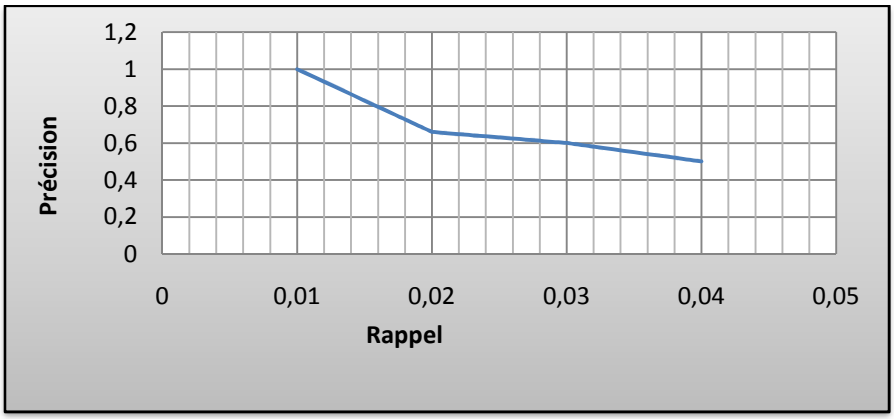


Fig.23. Courbe de rappel/précision pour l'image 44.png avec les moments de Hu

Les courbes précédentes montrent que le descripteur Histogramme RGB donne de meilleures performances par rapport au descripteur de forme (moments de Hu).

A partir des résultats présentés précédemment de la base de Columbia, nous pouvons calculer la table de rappel et précision. Avec cette image requête, nous avons obtenu le résultat consigné dans le tableau suivant :

descripteur	Histogramme RGB	Moments statistiques	Matrice de cooccurrence	Moments de Hu	combinaison
Nombre d'images pertinentes dans 12 images	11	12	5	3	12
Précision	91.66%	100%	41.66%	25%	100%
Rappel	11%	12%	5%	3%	12%

Tab.2. les valeurs de rappel et précision de la base de Coil

On trouve qu'avec cette image requête, la combinaison et les moments statistiques sont les meilleurs descripteurs. L'histogramme RGB est un descripteur efficace. La matrice de cooccurrence et les moments de Hu ne sont pas des descripteurs très forts dans ce cas.

Les figures suivantes donnent une vue global de la performance de notre système. Les courbes de précision/rappel sont présentées pour chaque base avec différentes caractéristiques. Donc, nous pouvons voir le rôle de chaque caractéristique. Les résultats changent avec les différentes caractéristiques.

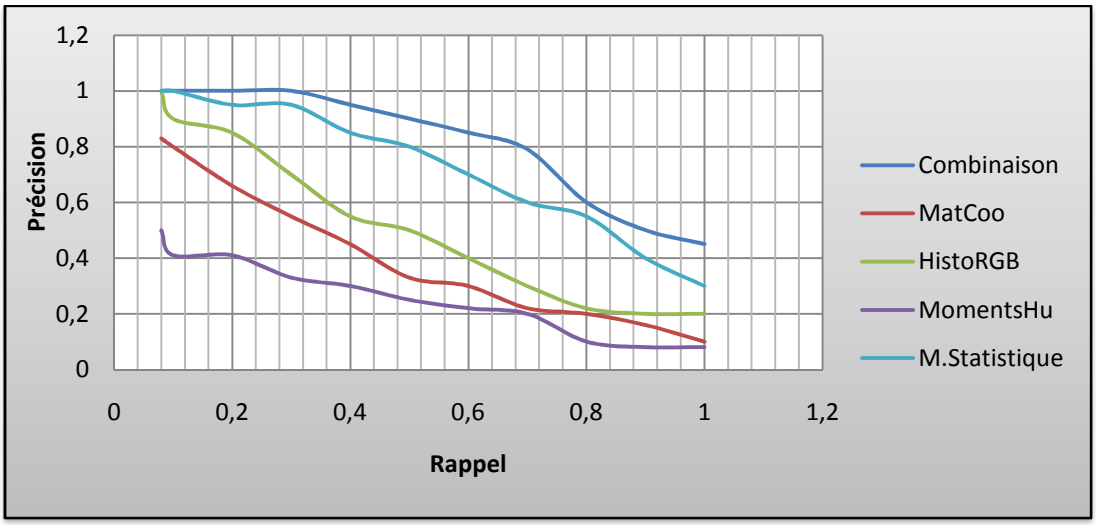


Fig.24. La courbe de rappel/précision pour la base Coil-100

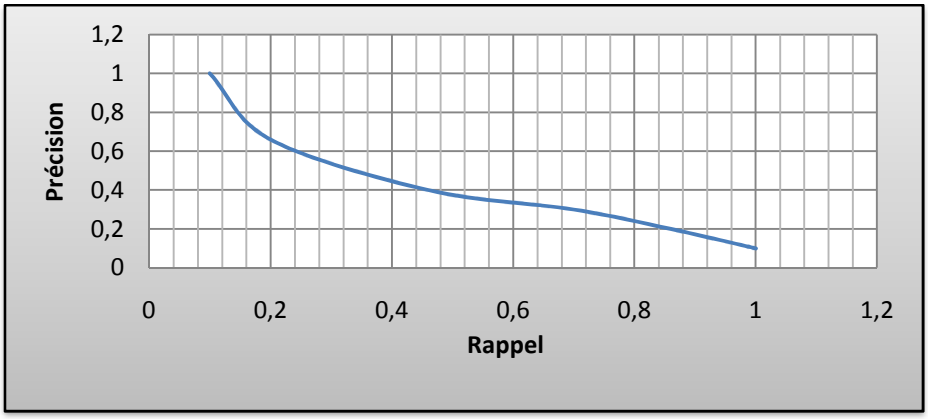


Fig.25. La courbe de rappel/précision pour la base ORL

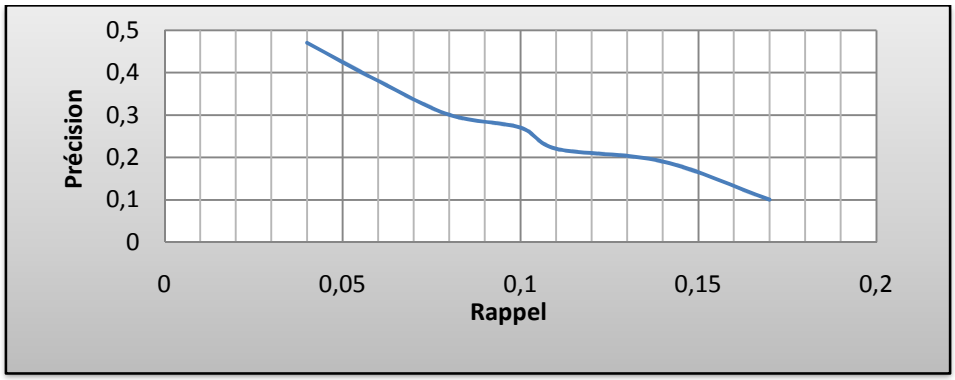


Fig.26. La courbe de rappel/précision pour la base d'images texturées

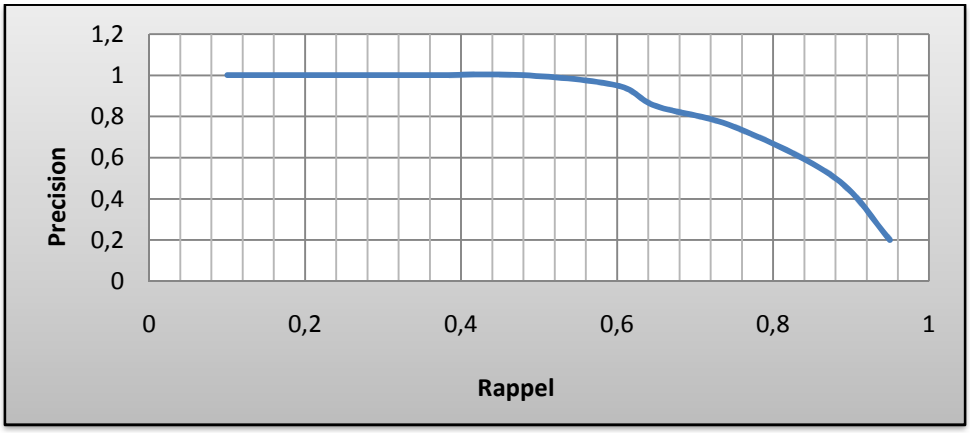


Fig.27. La courbe de rappel/précision pour la base d'images médicales

9. Discussion

Les résultats expérimentaux obtenus sont prometteurs mais ils doivent être considérés comme préliminaires. Nous constatons que toutes les courbes de rappel/précision des quatre bases sont globalement décroissantes, la précision décroit au fur et à mesure que des images non pertinentes sont retrouvées. Cela peut s'expliquer par le fait que nous avons

utilisé seulement quatre caractéristiques (histogramme RGB, les moments de couleurs, la matrice de cooccurrence, les moments de Hu et la combinaison des quatre descripteurs), un seul espace de couleurs (RGB) et que la mesure de similarité utilisé dans notre système, c'est une mesure classique. Néanmoins, il est significatif si on applique d'autres mesures pour vérifier si la mesure de similarité par la distance euclidienne est vraiment meilleure ou non, et utiliser d'autres caractéristiques comme les points d'intérêts, le filtre de Gabor, les ondelettes, etc.

10. Conclusion

Nous avons présenté dans ce chapitre notre contribution en essayant de répondre aux besoins des utilisateurs liés au domaine de la recherche d'images par le contenu. Nous y avons développé les principales étapes d'un processus de recherche d'images, à savoir, la représentation ou l'indexation et la comparaison des images. La première étape a été de définir des descripteurs adaptés aux images testées. Pour ce faire, nous avons proposé une adaptation du contexte de couleur, de texture et de forme.

L'histogramme et les moments de couleurs ont été utilisés pour caractériser la couleur. L'approche histogramme fournit une bonne représentation de la configuration correspondant à chaque image. Cependant, l'utilisation d'histogramme comme vecteur descripteur pose quatre inconvenients : la difficulté de créer une indexation rapide et efficace à cause de sa grande taille, le manque d'informations spatiales, la sensibilité à des petits changements de luminosité et l'inutilité pour la comparaison partielle des images. Malgré les bons résultats obtenus et les avantages de l'utilisation des moments statistiques de couleurs qui sont leurs tailles réduites, leur vitesse d'extraction, et leurs invariances à la rotation, à l'échelle et à la translation. Puisque seulement trois moments pour chaque trois couleurs composantes sont utilisés pour représenter le contenu en couleur pour chaque image, cette technique est une représentation plus compact. A cause de cette compacité les moments de couleurs peuvent diminuer la discrimination entre les images. Ainsi, pour réaliser ce système, deux autres informations essentielles ont été prises en compte : l'information texture et

l'information forme. La texture est représentée par la matrice de cooccurrence, malgré que la texture soit une caractéristique discriminante de l'image, mais son seule utilisation dans CBIR ne suffit pas et quatre paramètres calculés des matrices de cooccurrences ne sont pas suffisants pour identifier les images similaires. Pour décrire la forme nous avons utilisé les invariants de Hu. Cependant leur calcul est relativement long et ils sont très sensibles au bruit, ce qui peut s'avérer être un gros inconvénient dans un système de recherche d'images.

Conclusion générale

Dans ce mémoire, nous avons abordé le problème de la recherche d'images. Plus précisément, nous nous sommes focalisés à la recherche d'images basée sur le contenu visuel. Notre choix a été motivé par la quantité phénoménale d'images disponible aujourd'hui, qui ne cesse de croître.

Le but de ce travail n'a pas été de proposer une nouvelle définition, néanmoins l'idée a été d'extraire, dans les images, une certaine information pertinente de la couleur, la texture et la forme. Ces informations permettent une recherche efficace dans des bases d'images. Pour atteindre cet objectif, nous avons commencé par un état de l'art sur la recherche d'images par le contenu. Ensuite, nous avons étudié les différentes méthodes d'extractions des caractéristiques visuelles. Enfin nous nous sommes intéressés à l'intégration de nos méthodes proposées à un système d'indexation et de recherche d'images par le contenu. Ce système fonctionne avec des descripteurs visuels qui sont extraites d'une image telle que l'histogramme RGB, les moments statistiques de couleurs, la matrice de cooccurrence et les moments de Hu. Nous avons aussi développé une interface qui permet à l'utilisateur de proposer facilement une requête et d'afficher les résultats. Notre système a été testé avec quelques bases d'images différentes. Pour l'évaluer, nous avons calculé les courbes de rappel et précision pour chaque base d'images avec les différents choix de caractéristiques. Nos résultats ne sont pas toujours excellents avec toutes les bases mais ils sont cependant très prometteurs. A ce stade, plusieurs problèmes d'actualité complexes et

incontournables ont été soulevés : le premier se pose sur l'extraction et le choix des descripteurs visuels. Le deuxième se porte sur le fossé sémantique, qui sépare les descripteurs de bas niveau extraites du contenu des images et la sémantique que l'utilisateur place dans sa requête, qui conduit à des résultats souvent erronés. Le troisième se traduit par la malédiction de la dimension de l'espace de vecteurs de caractéristiques souvent de grande taille.

A l'issue des travaux menés dans le cadre de ce mémoire, nous dégageons quelques perspectives qui peuvent être réalisées à court et à moyen terme. En premier lieu, nous envisageons d'enrichir notre système en utilisant d'autres caractéristiques visuelles et d'autres espaces de couleurs afin d'étendre la collection des descripteurs utilisables. La seconde amélioration intervient dans la phase de recherche où nous prévoyons de combiner les approches textuelles et celles proposées pour améliorer l'interprétation sémantique des images. De plus, un système de bouclage de pertinence est une solution intéressante pour réduire le fossé sémantique. Nous pouvons également envisager d'utiliser des techniques de pondération d'attributs qui peuvent conduire à de meilleurs résultats sans nuire à la rapidité du système. Nous pensons également d'étudier et de tester les différentes mesures de similarité afin de les comparer et de pouvoir sélectionner celles qui ont la meilleure performance.

Bibliographie

[1] H. V. Dai. Association texte + images pour l'indexation et la recherche d'images. Rapport final, Juillet 2009.

[2] Y. Bouaziz. Recherche d'information spatiale : contribution à l'interrogation par croquis (sketch). Université Paul Sabatier Toulouse. Master 2. 2008-2009.

[3] N. Idrissi. La naviguation dans les bases d'images : prise en compte des attributs de texture. Université de Nantes. Thèse de Doctorat, Octobre 2008.

[4] Gwénélé Quellec. Indexation et fusion multimodale pour la recherche d'information par le contenu. Application aux bases de données d'images médicales. Université européenne Bretagne. Thèse de Doctorat, Septembre 2008.

[5] T. O. Nguyen. Localisation de symboles dans les documents graphiques. Université Nancy 2. Thèse de Doctorat, Décembre 2009.

[6] T. Kato, K. Hirata. Query by visual example in content-based image retrieval, Proc. EDB192. Lecture Notes in computer Science, 1992, p.56-71.

[7] A. Jain, A. Vailaya. Image retrieval using color and shape. Pattern Recognition, vol.29, n°8, 1996.

[8] J. Canny. A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 8, n°6 :679 – 698, Novembre 1986.

[9] R. Deriche. Using canny's criteria to derive an optimal edge detector recursively implemented. International Journal Computer Vision, Volume 1, n°2,167-187, 1987.

[10] M. Ferecatu. Image retrieval with active relevance feedback using both visual and keyword-based descriptors. Université de Versailles Saint-Quentin-En-Yvelines. Thèse de Doctorat, 2005.

[11] M.K. Hu. Visual Pattern Recognition by moment invariants. IRE Transaction on Information Theory, Volume 8, n°2 :179–187, 1962.

[12] I. Daoudi. Recherche par similarité dans les grandes bases de données multimédia Application à la recherche par le contenu d'images. INSA Lyon. Thèse de Doctorat, 2009.

[13] O. Aude, A. Torralba. Modeling the shape of the scene : a holistic representation of the spatial envelope. International Journal of Computer Vision, 42(3):145_175, 2001.

[14] B. Manjunath, W.Y.Ma. Texture features for browsing an retrieval of image data. IEEE Transaction on Pattern analysis and machine Intelligence, vol 18 numéro 8, 1996.

[15] S. Marcelaje. Mathematical description of the response of simple cortical cells. Journal of Optical Society of America. Vol 70 No 11 : 1297-1300. Novembre 1980.

[16] B.S.Manjunath, W.Ma. Texture features for browsing and retrieval of image data. IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI-Special issue on Digital Libraries), 18(8): 837-42.1996. pages 18.

[17] J. Landré. Analyse multi-résolution pour la recherche et l'indexation d'images par le contenu dans les bases de données application à la base d'images paléontologique Trans'Tyfipal. Université de Bourgogne. Thèse de Doctorat, Décembre 2005.

[18] J-Y Chen, C. A. Bouman, John C. Dalton. Hierarchical Browsing and Search of Large Image Databases. IEEETIP : IEEE Transactions on Image Processing, 9(3) : 442-455, 2000.

[19] C. Jacobs, A. Finkelstein, D. Salesin. Fast multiresolution image querying. Proceedings of SIGGRAPH95, Los Angeles, California, 1995.

[20] M. K. Mandal, T. Aboulnasr, S. Panchanathan. Illumination invariant image indexing using moments and wavelets. Journal of electronic Imaging, 7(2) :282– 293, 1998.

[21] M. Tuceryan, A. K. Jain. Texture analysis. The Handbook of Pattern Recognition and Computer Vision (Edition 2), pages 207-248, 1998.

[22] M. Swain, D. H. Ballard. Color indexing. International Journal of computer vision, 32(11) : 11-32.1991.

[23] M. A. Stricker, M. Orengo. Similarity of color images. SPIE. San Jose. 1995.

[24] S. Wang, *A* Robust CBIR Approach Using Local Color Histograms. Département d'informatique, Université Alberta. Alberta : s.n, 2001. Rapport technique.

[25] G. Lu, J.Phillips, Using perceptually weighted histograms for color-based image retrieval. Proceedings of the 4th International Conference on Signal Processing. pp.1150–1153.

[26] G. Pass, R. Zabih, Comparing images using joint histograms. Multimedia Systems, Vol. 7, pp. 234–240.

[27] M. Stricker, M. Swain, The capacity of color histogram indexing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 704–708. 1994.

[28] G. Pass, R. Zabih, J.Miller. Comparing images using color coherence vectors. ACM International Multimedia Conference, pp.6573.1996.

[29] L. Gueguen. Extraction d'information et compression conjointes des séries temporelles d'images satellitaires. Ecole Nationale Supérieure des Télécommunications Paris. Thèse de Doctorat, Octobre 2007.

[30] C. Vertan, N. Boujemaa. Upgrading color distributions for image retrieval : can we do better ? International Conference on Visual Information Systems, 2000.

[31] G. Finlayson, S. Chatterjee, B. Funt. Color angular indexing. In Proceedings of the 4 th European conference on computer vision, Combridge, Angleterre. 1996.

[32] C. Schmid. Appariement d'images par invariants locaux de niveaux de gris application à l'indexation d'une base d'objets. Institut National Polytechnique Grenoble. Thèse de Doctorat, Juillet 1996.

[33] P. Gros. De l'appariement à l'indexation des couleurs. Institut National Polytechnique Bretagne. Thèse de Doctorat 2008.

[34] K. Mikolajczyk. Detection of local features invariant to affine transformations application to matching and recognition. Institut National Polytechnique Bretagne. Thèse de Doctorat, Juillet 2002.

[35] S. Matusiak. Description invariante et locale des formes planes, application à l'indexation d'une base d'images. Université Valencienne et du Hainaut Combris. Thèse de Doctorat 1999.

[36] P. Lambert, H. Grecu. Indexation par descripteurs flous : application à la recherche d'images. GRETSI01. 2001.

[37] J. Fauqueur. Contributions pour la Recherche d'Images par Composantes Visuelles. Université de Versailles Saint-Quentin en Yvelines. Thèse de Doctorat, Novembre 2003.

[38] S. Cohen. Finding color and shape patterns in images. Université Stanford. Thèse de Doctorat, Juin 1999.

[39] S. A. Berrani. Recherche approximative de plus proches voisins avec contrôle probabiliste de la précision ; application à la recherche d'images par le contenu. Université Rennes 1. Thèse de Doctorat, Février 2004.

[40] M. N. Do, M. Vetterli. Wavelet-Based Texture Retrieval Using Generalized Gaussian Density and Kullback-Leibler Distance. In IEEE Transactions on Image Processing, vol. 11, 146-157, 2002.

[41] C. Liu, M. Mandal. Image Indexing in the JPEG2000 Framework. SPIE Internet multimedia management systems, 272-280, Boston, 2000.

[42] Z. Xiong and T. S. Huang," Subband-based, memoryefficient JPEG2000 images indexing uncompresseddomain", IEEE Symp. on IAI, 2002.

[43] R. Bensalma, M. C. Larabi. Indexation d'images multi-spectrales par une approche conjointe dans les domaines spatial et compressé. Laboratoire SIC Université Poitiers, Novembre 2007.

[44] P. Gros, G. Mclean, R. Delon, R. Mohr, C. Schmid et G. Mistler. Utilisation de la couleur pour l'appariement et l'indexation d'images. Technical report n°RR-3269, INRIA, Septembre 1997.

[45] J.M. Jolion. Principles of Visual Information Retrieval, chap. Feature Similarity. Springer, 2000, M. Lew édition.

[46] W. Niblack, R. Barber, W. Equitz, M. Flickner, E.H. Glasman, D. Petkovic, P. Yanker, C. Faloutsos et G. Taubin. The QBIC project: Querying images by content, using color, texture, and shape. Storage and Retrieval for Image and Video Databases (SPIE), pp. 173-187. February 1993.

[47] K. Houari. Recherche d'images par le contenu. Université Mentouri Constantine, Thèse de Doctorat, Juin 2010.

[48] N. A. Thacker, F.J. Aherne, P. I. Rockett. The bhattacharyya metric as an absolute similarity measure for frequency coded data. 1998.

[49] T.Quach, U. Monich, B. S. Manjunath. A system of large scale, content based web image retrieval. Université California Santa Barbara. 2004.

[50] S. Ardizzoni, I. Bartolini, and M. Patella. Windsurf : Region-based image retrieval using wavelets. *DEXA Workshop*, pages 167–173, 1999.

[51] J. Fournier. Indexation d'images par le contenu et recherche interactive dans les bases généralistes. Université Cergy Pontoise. Thèse de Doctorat, Octobre 2002.

[52] Lê Thi Lan. Indexation et recherche d'images par le contenu. IP de HONOI. Mémoire de Master 2005.

[53] A. Brilhault. C. Garbay. G. Quénot. Indexation et recherche par le contenu de documents vidéo.

[54] M. A. Bourenane. Un outil pour l'indexation des vidéos personnelles par le contenu. Université de Québec à trois- rivières. Thèse de Doctorat, 2009.

[55] A. Hafiane. Caractérisation de textures et segmentation pour la recherche d'images par le contenu. Université de Pris-Sud XI. Thèse de Doctorat, Décembre 2005.

[56] O. Adjemout, Reconnaissance automatique de formes à partir des paramètres morphologiques, de couleur et de texture : application au tri des graines de semences, thèse de magister, UMMTO, 2005

[57] G. Pass, R. Zabith, Histogramme refinement for content-based image retrieval. IEEE Workshop on Applications of Computer Vision pp. 96-102, 1996.

[58] A. Znaidia, T. Zaharia, F. Preteux, Une évaluation des descripteurs visuels MPEG-7 pour la recherche d'images par le contenu. Institut Télécom ; Télécom SudParis ; Département ARTEMIS.

[59] J. Zhang, T. Tan. brief review of invariant texture analysis methods. Pattern Recognition, vol 35, pp. 735-747, 2002.

[60] M. M. Galloway, "Texture analysis using gray level run lengths", In graphical Models and image Processing. Vol. 4. Pp.172-179, 1975.

[61] J. P. Cocquerez, Philipp S. (coord.), Analyse d'images : filtrage et segmentation, Masson, Paris, 1995.

[62] R.M. Haralick, K. Shanmugan, I. Dinstein, Textural features for image classification. IEEE Trans. Syst. Man. Cybern. SMC-3(6): 610–621. 1973.

[63] R. M. Haralick, "Statistical and structural approaches to texture", Proceedings of the IEEE, Mai 1979, number 5, vol.67.

[64] H. Tamura, S. Mori, and T. Yamawaki, Texture features corresponding to visual perception, IEEE Trans. On Systems, Man, and Cybernetics, vol. Smc-8, No. 6, Juin 1978.

[65] P. Howarth, S. Rüger. Evaluation of texture features for content based image retrieval. In Proceedings of the Conference on Image and Video Retrieval (CIVR), pages 326_334, Dublin, Irlande, juillet 2004.

[66] A. Rosenfeld, J. Weszka, Picture Recognition. In *Digital Pattern Recognition*, K. Fu (Ed.), Springer-Verlag, 135-166, 1980. 1980.

[67] J. Daugman, Uncertainty Relation for Resolution in Space, Spatial Frequency and Orientation Optimised by Two-Dimensional Visual Cortical Filters. Journal of the Optical Society of America, 2, 1160-1169. 1985.

[68] A. Bovik, M. Clark, W. Giesler, Multichannel Texture Analysis Using Localised Spatial Filters. IEEE Trans. Pattern Analysis and Machine Intelligence, 12, 55-73. 1990.

[69] S. Mallat, Multifrequency Channel Decomposition of Images and Wavelet Models. IEEE Trans. Acoustic, Speech and Signal Processing, 37, 12, p. 2091- 2110. 1989.

[70] A. Laine, J. Fan, Texture Classification by Wavelet Packet Signatures. IEEE Trans. Pattern Analysis and Machine Intelligence, 15, 11, p. 1186-1191. 1993.

[71] C. Lu, P. Chung, C. Chen, Unsupervised Texture Segmentation via Wavelet Transform. Pattern Recognition, 30, 5, p. 729-742. 1997.

[72] R. Bajscy. Computer Identification of Visual Surface. Computer Graphics and Image Processing, 2, 118-130, 1973.

[73] D. Mercier, R. Séguier, Utilisation des STANN en audio : illustrationen reconnaissance de chiffre, Journée Valgo 2001.

[74] Laboratoire L3I- Université de La Rochelle, Segmentation d'Images de Documents Anciens par Approche Texture, 2006.

[75] G. A. Wright, ME. Jenigan, Texture Discrimination Using Spatial Frequancy Channels", IEEE, pp. 519-524, 1986.

[76] K. Albeau. Analyse à grande échelle des textures des séquences protéiques via l'approche Hydrophobic Cluster Analysis (HCA).Université Versailles Saint-Quentin-En-Yvelines, Octobre 2005.

[77] M. Sonka, V. Hlavac, R. Boyle. Image Processing, Analysis and Machine Vision. PWS Publishing, seconde edition, 1999.

[78] M.R. Teague : Image analysis via the General Theory of moments, Applied optics, vol. 19, n° 8 (1980), pp. 1353-1356, 1980.

[79] D. ZHANG, G. LU, Study and evaluation of different Fourier methods for image retrieval, Image Vision Computing, vol. 23, 2005.

[80] J. F. HADDON et J. F. BOYCE. Co-occurrence matrices for image analysis. IEEE Electronics and Communications Engineering Journal, 5(2):71–83, 1993.

[81] R. M. HARALICK, K. SHANMUGAM et I. DINSTEIN. Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics, SMC-3(6):610–621, Novembre 1973.

[82] M. Sharma, S.Singh. Evaluation of texture methods for image analysis. Proceedings of the 7th Australian and New Zealand Intelligent Information Systems Conference 2001.

[83] J. Zhang, T. Tan. Brief review of invariant texture analysis methods. Pattern Recognition. Vol. 35, iss. 3, March, pp. 735-747, 2001.

[84] R. Baeza-Yates, B. Ribeiro-Neto, Modern Information Retrieval, ACM Press / Addison-Wesley, New York, 1999.

[85] J. R. Smith, Image Retrieval Evaluation, IEEE Workshop on Content-based Access to Image and Video Databases, IEEE Computer Society Washington, DC, USA, June 1998, page 112.

[86] T. Gevers, A. Smeulders, Emerging Topics in Computer Vision, Chapitre Content-Based Image Retrieval : An Overview, Addison-Wesley / Prentice Hall, 2004.

[87] K. Mikolajczyk, C. Schmid, A performance evaluation of local descriptors, *PAMI*, vol. 27, no 10, 2005, pp. 1615-1630.

[88] http://www1.cs.columbia.edu/CAVE/research/softlib/

[89] http://www.cl.cam.ac.uk/Research/DTG/attarchive:pub/data/att_faces.tar.Z

[90] http://www-cvr.ai.uiuc.edu/ponce_grp/data/.