

République Algérienne Démocratique Populaire

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université d'Ibn Khaldoun – Tiaret

Faculté des Mathématiques et de l'Informatique

Département Informatique

Thème

Recherche d'Images dans un Contexte Big Data

Pour l'obtention du diplôme de Master II

Spécialité : Informatique

Option : Génie logiciel

Réalisé par : Chadli Rachid

Lakaf Houria

Dirigé par : Mr. Chenine Abdelkader

Année universitaire 2015-2016

Dédicaces

Nous dédions ce mémoire

A nos chers parents

Pour leur soutien, leur patience, leur sacrifice

et leur amour, vous méritez tout éloge,

vous qui avez fait de nous ce que nous sommes maintenant.

Nous espérons être l'image que vous êtes fait de

nous, que dieu vous garde et vous bénisse.

A tous les membres de nos familles.

Nous dédions aussi ce travail à nos chers frères et sœurs,

pour leur affection et leur encouragement qui ont toujours

été pour nous des plus précieux.

Que ce travail soit pour vous le gage de notre profond amour

A tout nos amis

A tous nos enseignants

A tous ceux qui nous ont aidés.

A tous ceux que nous aimons, nous dédions ce travail ...

Remerciements

Nous remercions le Bon Dieu tout puissant, de nous avoir donné le bon sens et la grande volonté pour réaliser ce travail.

Nous remercions du fond de cœur Mr. **CHENINE Abdelkader** pour avoir accepté d'encadrer ce mémoire de fin d'étude et pour nous avoir guidés en toute sincérité.

Nous remercions aussi les jurys d'avoir accepté d'examiner notre travail.

Nous remercions tous nos professeurs de l'université pour leur contribution à notre formation scientifique.

Nous remercions tous ceux qui nous ont aidés à réaliser ce travail.

Liste des abréviations

EM : Electro Magnétique

RGB:Red-Green-Blue.

CBIR:Content Based Image Retrieval

RIC : Recherche d'Images par le Contenu Visuel.

RI : Recherche d'Information.

BD, BDD : Base de données.

3D : Trois Dimensions

CIE : Commission Internationale de l'Eclairage.

HSV(HSI): Hue Saturation Value

RVB: Rouge Vert Bleu.

QBIC: Query by Image Content

MPEG:Moving Picture Experts Group

DWT:Discrete Wavelet Transform.

EMD: Earth Mover Distance.

KL : Kullback Leiber

JD : Divergence de Jeffrey

3V :volume, vitesse et variété

TF :Term Frequency

TF_IDF : Term Frequency, Inverse Document Frequency

TBIR : Text Based Image Retrieval

IP : Internet Protocol

TAL : Traitement Automatique des Langues

SIFT:scale-invariant feature transform

IDE :Integrated Development Environment

BMP:Bitmap

JPG, JPEG:Joint Photographic Experts Group

PNG : Portable Network Graphics

Liste des figures

Figure 1 : Image couleur, son repère et un extrait de pixels.....	1
Figure 2 : Image couleur indexée	2
Figure 3 : Image en niveau de gris et son histogramme.....	4
Figure 4 : L'image (a) contient des informations totalement invisibles pour l'œil humain, l'image (b) est obtenue après un traitement de débruitage.....	6
Figure 5 : (a) Texture aléatoire (b) Texture directionnelle (c) Texture structurale.....	8
Figure 6 : Voisinages (a) 8-connexité (b) 4-connexité.....	9
Figure 7 : Traitements fondamentaux en traitement d'images	10
Figure 8 : Schéma d'un système classique de recherche d'information.....	13
Figure 9 : Aperçu d'un système d'indexation et de recherche d'image.....	15
Figure 10 : Les Différents types de requêtes.....	16
Figure 11 : Approche général concernant CBIR.....	17
Figure 12 : Schéma standard d'une approche d'indexation par le contenu visuel.....	18
Figure 13 : Architecture en module de CBIR.....	19
Figure 14 : Le modèle de couleur RVB.....	21
Figure 15 : Espace de couleur XYZ.....	22
Figure 16 : Espace de couleur $L^*a^*b^*$	23
Figure 17 : Espace de couleur HSV.....	23
Figure 18 : Espaces de couleur YCbCr.....	24
Figure 19 : Description d'un filtre de Gabor.....	33
Figure 20 : recherche d'image basée sur le contenu.....	39

Figure 21 : Le rappel et la précision pour une requête (Yates, 1999).....	45
Figure 22 : Caractéristique du Big Data (Les 3 V).....	48
Figure 23 : Architecture de l'application.....	55
Figure 24 : Interface Utilisateur.....	56
Figure 25 : Présentation des résultats.....	57
Figure 26 : Un échantillon de la base.....	58

Résumé

Les méthodes de recherche d'images fournissent un moyen d'aide à la décision dans de nombreux domaines d'application. Un cas typique d'utilisation est la recherche d'images similaires à une image requête fournie par l'utilisateur. La recherche par le contenu (CBIR) se base généralement sur des descripteurs bas niveau (couleur, texture et forme). Ce travail constitue une tentative de construction d'un système de recherche combinant les requêtes texte et/ou image. Pour la recherche à base de requête texte, le modèle vectoriel a été utilisé ; alors que les descripteurs visuels de couleur (modèles RGB et HSV), de forme ainsi que de texture ont été pour la recherche à base de requête image.

Mots-clés : image, recherche d'images, modèle vectoriel, descripteurs visuels.

Abstract

The image search methods provide a means for decision support in many areas of application. A typical use case is looking for images similar to a query image provided by the user. Content based image retrieval (CBIR) is generally based on low-level descriptors (color, texture and shape). This work is an attempt to construct a search system combining text query and/or image query. For text-based search application, the vector model was used; while the visual descriptors of color (RGB and HSV models), shape and texture were for the imagequery-based search.

Keywords: image, search, vector model, visual descriptors

Sommaire

Liste des abréviations

Liste des figures

Liste des tableaux

Résumé

Sommaire

Introduction Générale

Chapitre1 : Généralités sur le Traitement d'images 3

Introduction 3

I.1. L'image 3

I.2. Types d'images 4

I.3. Caractéristiques de l'image 5

I.3.1. Pixel 5

I.3.2. Poids de l'image..... 5

I.3.3. Transparence..... 5

I.3.4. Luminance 5

I.3.5. Contraste 6

I.3.6. Histogramme 6

I.3.7. Bruit 7

I.3.8. Texture 9

I.3.9. Homogénéité..... 10

I.3.10. Connexité..... 10

I.3.11. Région..... 11

I.4. Etapes fondamentales en traitement d'images 12

Conclusion..... 14

Chapitre II : Recherche d'images 16

Introduction 16

II.1. Systèmes de recherche d'images	17
II.2. Architecture d'un système d'indexation et de recherche d'images	19
II.3. Recherche d'Image par le Contenu (CBIR)	21
1. L'indexation	22
2. La Segmentation.....	22
3. La recherche où l'appariement requête-image	23
II.4. Descripteurs d'images	24
II.4.1 Descripteurs de couleurs	24
II.4.2. Descripteurs de textures	33
II.4.3. Descripteurs texturaux	37
II.4.4. Les descripteurs des points d'intérêts.....	39
II.4.5. Descripteurs de formes	41
II.5. Mesures de similarité	45
II.5.1. Les méthodes de calcul	46
II.6. Domaines D'application de CBIR	50
II.7. Méthodes d'évaluation d'exécution de CBIR (Précision, Rappel)	50
II.7.1. Le rappel (en anglais : Recall)	50
II.7.2. La Précision (en anglais : Precision).....	50
II.7.3. La courbe de rappel et précision	52
II.8. Big Data	53
II.8.1. Dimensions	53
Conclusion.....	57
Chapitre III : Expérimentation et résultats	59
Introduction	59
III.1. Les outils de développement	59
III.1.1. Matlab	59
III.2. Conception	59

III.2.1. Les caractéristiques visuelles choisies.....	59
III.2.2. Les caractéristiques textuelles choisies.....	60
III.2.3. Les requêtes	61
III.2.4. Mesures de distance utilisées	62
III.2.5. Architecture de l'application.....	62
III.3. Réalisation	63
III.3.1. L'interface utilisateur	63
III.3.2. La présentation des résultats	64
III.4. Evaluation de l'application	64
III.4.1. Présentation des résultats.....	64
III.4.2. Processus suivi.....	66
III.4.3. Problèmes rencontrés.....	73
III.5. Perspective	73
Conclusion	73

Conclusion Générale

Bibliographie

Webographie

Introduction Générale

Cette thèse s'est déroulée dans le cadre des systèmes de recherche d'image (RIC). L'approche que nous avons suivie au cours de ce travail a été d'implémenter une application (Moteur de recherche) qui permet de retrouver et afficher des images similaires à une image requête (données par l'utilisateur).

Les images exigent une grande capacité de stockage, une importance de traitement et de recherche. Dans notre travail, nous nous sommes intéressés à des problèmes à savoir l'indexation (description) et la recherche d'images par des requêtes sous forme d'image et/ou texte. L'indexation de données fait l'objet de nombreux travaux de recherche [23]. On peut distinguer deux phases complémentaires dans le processus d'indexation : la structuration et l'annotation.

La structuration, pour les images, elle permet de mettre en évidence des régions caractéristiques de l'image. Ces informations sont traduites par une structure, respectivement spatiale et temporelle, ainsi que diverses métadonnées de bas niveau décrivant le contenu des images. La phase d'annotation consiste alors à associer des métadonnées à différentes images. L'annotation manuelle est alors l'approche généralement suivie pour l'association d'information sémantique pertinente aux images. Mais, elle est longue pour une grande base d'image. Il est donc essentiel d'avoir des techniques permettant une annotation automatique tout du moins semi-automatique pour faciliter la description d'une information pertinente dans une base de données massives.

Le mémoire décrit les différents aspects de nos travaux concernant la recherche d'image dans le contexte BIG Data. Il est divisé en trois chapitres essentiels :

Chapitre 1 : décrit les généralités sur le [traitement d'images](#).

Chapitre 2 : présente la recherche d'images, les différents descripteurs d'images, les mesures de similarité, et les Big Data.

Chapitre 3 : illustre les expérimentations et les résultats obtenus à partir de notre travail.

Chapitre 1

Généralités sur le Traitement d'images

Chapitre1 : Généralités sur le Traitement d'Images

Introduction

Le domaine du traitement d'images numériques est un ensemble de traitements appliqués à l'image numérique par le biais de l'ordinateur. La vision joue un rôle important dans notre vie, d'où une intense utilisation d'images pour la perception humaine. Alors que nous sommes limités à la bande visible du spectre électromagnétique (EM), les ordinateurs couvrent le spectre tout entier de la gamme gamma aux ondes radio. Nous allons introduire quelques techniques relevant de ce traitement en passant de la définition de l'image jusqu'à sa segmentation.

1.1. L'image [10]

L'image est définie comme étant une fonction $f(x, y)$ à deux dimensions, où x et y sont les coordonnées spatiales, et f l'amplitude à tous points (x, y) correspondant à l'intensité ou au niveau de gris. Lorsque les points (x, y) et l'amplitude sont discrétisés, on parle d'image numérique ou digitale. Dans ce dernier cas la fonction f est remplacée par la lettre I et le couple (x, y) par le couple (i, j) .

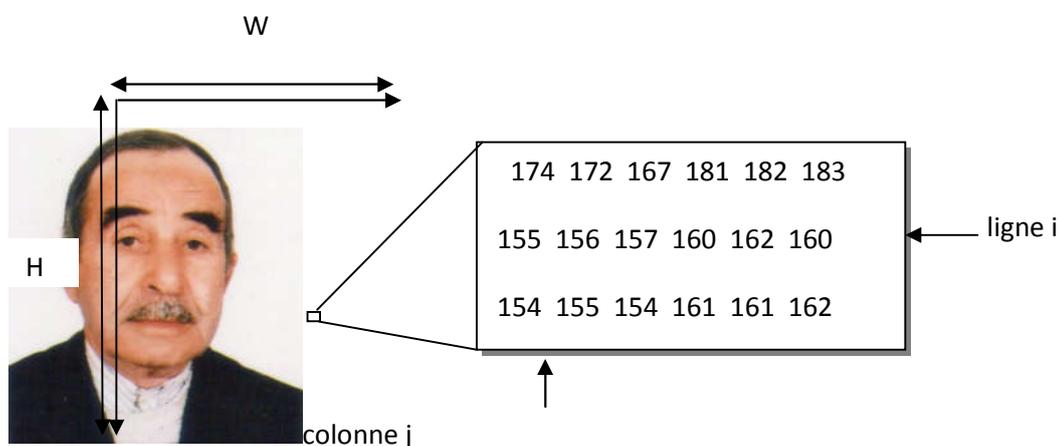


Figure 1 :Image couleur, son repère et un extrait de pixels

L'image numérique, désignée aussi par le terme scène, possède un repère comme indiqué en **Figure 1**, il est différent de celui d'une fonction mathématique. Elle a une hauteur (H) et une largeur (W).

1.2. Types d'images [10]

En traitement d'images nous aurons à faire à quatre types d'images :

Image binaire : chaque pixel est soit blanc soit noir. Puisqu'il y a uniquement deux valeurs pour chaque pixel, un seul bit est utilisé pour le coder.

Image en niveau de gris : chaque pixel est un niveau de gris, allant de 0 (noir) à 255 (blanc). Cet intervalle de valeur signifie que chaque pixel est codé sur huit bits (un octet). 256 niveaux de gris suffisent pour la reconnaissance de la plus part des objets d'une scène.

Image couleur (RGB) : chaque pixel possède une couleur décrite par la quantité de rouge (R), vert (G) et bleu (B). Chacune de ces trois composantes est codée sur l'intervalle [0, 255], ce qui donne $255^3 = 16\ 777\ 216$ couleurs possibles. Il faut 24 bits pour coder un pixel.

Image indexée : la plupart des images n'utilisent pas les 16 millions de couleurs possibles, pour un besoin de stockage et de gestion de fichier, l'image à une palette de couleur qui représente la liste des couleurs qu'elle utilise. La valeur du pixel est un index dans cette palette.

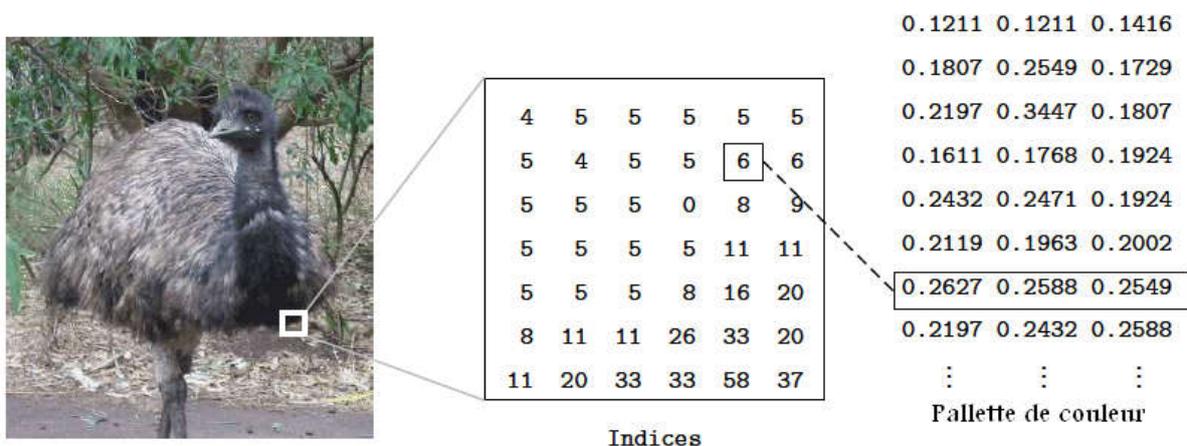


Figure 2 : Image couleur indexée

Comme l'indique la **Figure 2**, l'indice 6 pointe vers la palette de couleur qui fournit la proportion de chaque couleur de base (R, G, B).

I.3. Caractéristiques de l'image [10]

I.3.1. Pixel

Une image est constituée d'un ensemble de points appelés *pixels* (voir l'extrait de la **Figure 1**). Le pixel (pictureelement) représente ainsi le plus petit élément constitutif d'une image numérique. La quantité d'information que véhicule chaque pixel donne des nuances entre images monochromes et images couleurs. Pour les images 3D le «pixel» est alors appelé un *voxel*, et représente un volume élémentaire. Des exemples d'images de ce type se rencontrent dans les images médicales. Les images tomographiques axiales sont ainsi des images construites à partir de plusieurs radiographies faites sous des angles de vue différents.

I.3.2. Poids de l'image

C'est la taille de l'image. Etant donné que cette dernière est représentée sous forme d'une matrice dont les valeurs représentent l'intensité (pixels), le nombre de colonne (W) multiplié par le nombre de ligne (H) donne le nombre total de pixels dans l'image.

Pour une image de 640x480 en couleur :

- ✓ Nombre de pixel = $640 \times 480 = 307200$
- ✓ Poids de chaque pixel = 3 octets
- ✓ Le poids de l'image = $307200 \times 3 = 921600$ octets = 900 Ko

I.3.3. Transparence

La transparence est une caractéristique définissant le niveau d'opacité des éléments de l'image, c'est la possibilité de voir à travers l'image des éléments graphiques situés derrière celle-ci.

I.3.4. Luminance

C'est le degré de luminosité des points de l'image. Elle est définie ainsi comme étant le quotient de l'intensité lumineuse d'une surface par l'aire apparente de cette surface.

1.3.5. Contraste

Le contraste est une propriété visuelle qui permet de distinguer, dans une image, deux régions distinctes.

1.3.6. Histogramme

Un histogramme est un graphique statistique permettant de représenter la distribution des intensités des pixels d'une image. Il fournit diverses informations comme les statistiques d'ordre (moyenne, variance,...), l'entropie, et peut permettre d'isoler des objets.

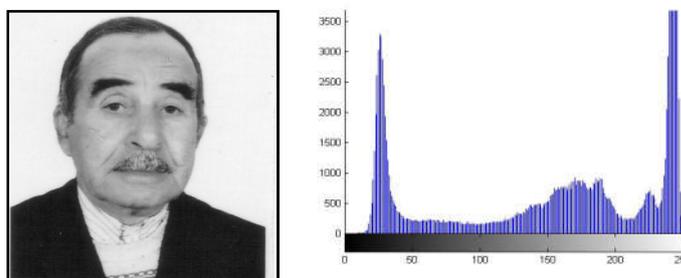


Figure 3 : Image en niveau de gris et son histogramme

L'histogramme de la **Figure 3** noté ici $H(x)$, est le nombre de pixels dont le niveau de gris est égal à x . l'histogramme cumulatif normalisé est calculé comme suit :

$$HC(x) = \frac{\sum_{i=0}^x H(x)}{W * H} \quad (1)$$

$HC(x)$ est le taux de pixels dont le niveau de gris est inférieur à x .

Nous présentons dans la suite quelques traitements d'analyse effectués uniquement à partir de l'histogramme. Retenons que certains de ces traitements sont souvent calculés au niveau des capteurs, et qu'en général leur pertinence est très intimement liée aux conditions d'acquisition.

- (1) Normalisation : exploiter toute la dynamique de codage.
- (2) Egalisation : équilibrer la dynamique de codage et augmenter le contraste.
- (3) Segmentation : simplifier l'image en regroupant les pixels selon leurs valeurs.

I.3.7. Bruit

Le bruit est un terme issu du domaine de l'acoustique et désigne un signal parasite. Que ce soit pour le son ou pour l'image, le principe est identique : sur tout signal de base vient s'adjoindre un ensemble d'informations parasites aléatoires. Si le niveau du signal est suffisant, la proportion de bruit dans le signal utile (le rapport signal/bruit) reste insignifiante. Par contre, si le niveau de bruit prend le pied sur l'information principale, le bruit sera présent.

A chaque étape de l'acquisition d'une scène, des perturbations (rayures, poussières, caméra, amplification, quantification) vont détériorer la qualité de l'image et contribuer à une mauvaise compréhension et un mauvais traitement de la scène. Le bruit peut provenir de différentes sources :

- ✓ Environnement lors de l'acquisition,
- ✓ Qualité du capteur,
- ✓ Qualité de l'échantillonnage,
- ✓ Liés à la transmission,
- ✓ Changement de format ou de stockage,
- ✓ Vieillesse de film...

Bruit lié au contexte de l'acquisition

Dans cette première catégorie, on retrouve tous les événements inattendus, modifiant les conditions de l'acquisition du signal. L'exemple le plus simple est le bougé. On peut aussi imaginer une modification ponctuelle des conditions d'éclairage, conduisant à une sous (sur) illumination de l'objet observé. Le caractère inattendu de ces événements fait qu'il est pratiquement impossible d'en corriger les effets sans l'intervention d'un opérateur humain pour activer le bon processus de restauration (correction de flou, de bougé, de dérive lumineuse, ...).

Bruit lié au capteur

Le capteur, s'il est de mauvaise qualité ou s'il est mal utilisé, peut introduire toute sorte de bruits. On peut ainsi obtenir une distorsion de la gamme des niveaux de gris, provoquant une saturation ou bien une distorsion géométrique de l'image équivalente à l'effet d'un miroir grossissant par exemple. Une mauvaise mise au point peut provoquer un flou. Bien entendu,

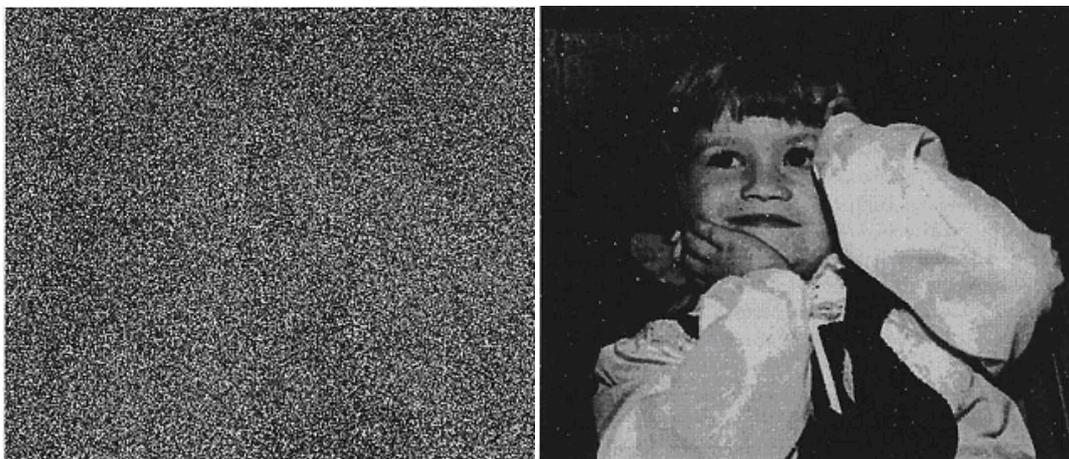
la meilleure restauration possible consiste à agir directement au niveau du capteur. Cependant, si cela n'est pas possible, la connaissance de l'appareil, permet le plus souvent, de construire un modèle à priori du phénomène de dégradation; une inversion systématique est alors envisageable. C'est le cas en microscopie électronique où il est très difficile d'obtenir des images de bonne qualité, mais où les phénomènes dégradants (produisant le plus souvent un mauvais contraste ou une dérive lumineuse) sont bien connus et modélisés.

Bruit lié à l'échantillonnage

Ce type de bruit est connu sous le nom d'effet "poivre et sel". Ce bruit est aussi généré par une texture dont les caractéristiques sont trop fines par rapport à la fréquence d'échantillonnage.

Bruit lié à la nature de la scène

L'environnement dans lequel se situe l'acquisition n'est pas toujours maîtrisable. En télédétection, la couverture nuageuse dégrade la qualité des images satellites. De même, la poussière dans un atelier de production peut être difficile à éliminer. Cependant, là encore, la connaissance a priori du phénomène perturbateur permet d'envisager une modélisation et donc une correction systématique.



(a) Image bruitée

(b) Image sans bruit correspondante à (a)

Figure 4 : L'image (a) contient des informations totalement invisibles pour l'œil humain, l'image (b) est obtenue après un traitement de débruitage.

Modèles de bruit

En général, le bruit d'image est considéré comme un champ aléatoire. Dans la plus part des cas, on considère le bruit comme additif. L'image à traiter I peut être mise sous la forme :

$$I = g + B \quad (\text{II.2})$$

où g représente l'information utile et B un champ aléatoire.

Parfois le bruit est considéré comme multiplicatif, par exemple en présence de speckle dans des images radar ou de grain sur des films radiologiques. L'image est sous la forme :

$$I = g \bullet B \quad (\text{II.3})$$

Finalement, le bruit peut être considéré comme convolutif pour décrire les dégradations dues au bougé, de mauvaise mise au point. L'image est :

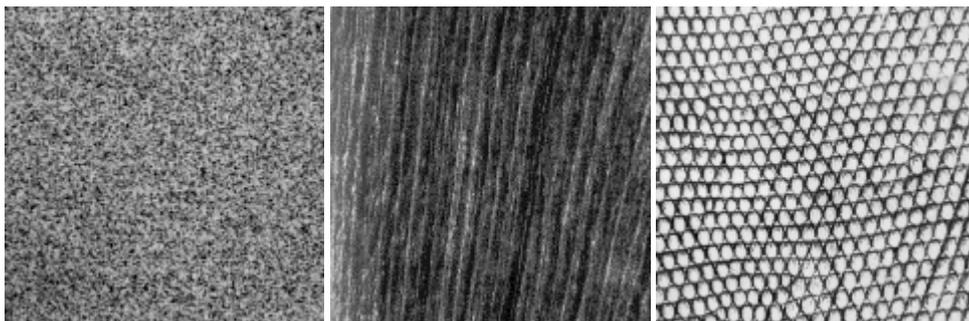
$$I = g * B \quad (\text{II.4})$$

Réduction du bruit

Pour réduire l'effet du bruit on a recours aux filtres linéaires (filtre moyonneur, filtre gaussien, filtre exponentiel), aux filtres non linéaires (filtre d'ordre, L-filtre, ...), aux filtres adaptatifs (filtre de rang adaptatif, moyenne tronquée adaptative,...), au filtrage homomorphique,...

1.3.8. Texture

En générale, une texture est le mode d'intercroisement des fibres dans les tissus. En traitement d'image elle est définie comme étant une région dans une image numérique qui a des caractéristiques homogènes. Ces caractéristiques sont par exemple un motif basique qui se répète, ou des caractéristiques fréquentielles. Une texture se compose de «texels», l'équivalent des pixels.[2]



(a) (b) (c)

Figure 5 : (a) Texture aléatoire (b) Texture directionnelle (c) Texture structurée

De nombreux travaux utilisant des caractéristiques de texture existent. Citons, pour illustration, les applications :

- Segmentation d'images aériennes (détection des arbres, des bâtiments,...) ;
- Segmentation d'images volcaniques ;
- Recherche de fruits dans un arbre ;
- Segmentation de scènes routières pour asservir un véhicule sur la route ;
- Détection de défauts sur des tôles décapées.

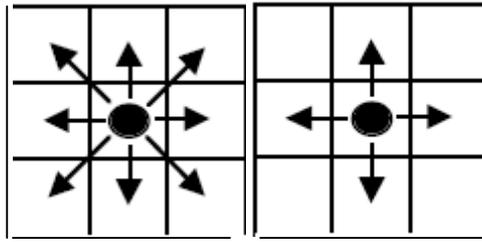
1.3.9. Homogénéité

L'homogénéité est une information locale et correspond au caractère uniforme d'une région. Une région dans une image est dite homogène si elle regroupe un ensemble de pixels qui possèdent des caractéristiques similaires ou uniformes. Ces caractéristiques peuvent être par exemple la variance du niveau de gris, la couleur, la texture,... [10]

1.3.10. Connexité

En traitant une image, on est souvent amené à se déplacer dans celle-ci. Un déplacement doit souvent obéir à des règles de voisinage, on utilise généralement deux types de voisinage : le voisinage à 4-connexité (4 pixels voisins, **Figure 6 (a)**) et le voisinage à 8-connexité (8 pixels voisins, **Figure 6 (b)**). Deux pixels seront considérés comme connexes (appartenant au même objet donc) s'ils satisfont deux critères :

- d'une part un critère de similarité (par exemple même niveau de gris)
- s'ils sont adjacents (voisins)



(a)

(b)

Figure 6 : Voisinages (a) 8-connextité (b) 4-connextité

I.3.11. Région

Une région est un ensemble de pixels connexes et homogènes. Un pixel n'appartient à une région donnée que s'il vérifie les caractéristiques de celle-ci (intensité moyenne, centre de gravité, ...). Une région est toujours limitée par un contour.

I.4. Etapes fondamentales en traitement d'images [10]

Une panoplie de traitements peut être appliquée à l'image numérique, la **Figure 7** résume l'ensemble de ces traitements.

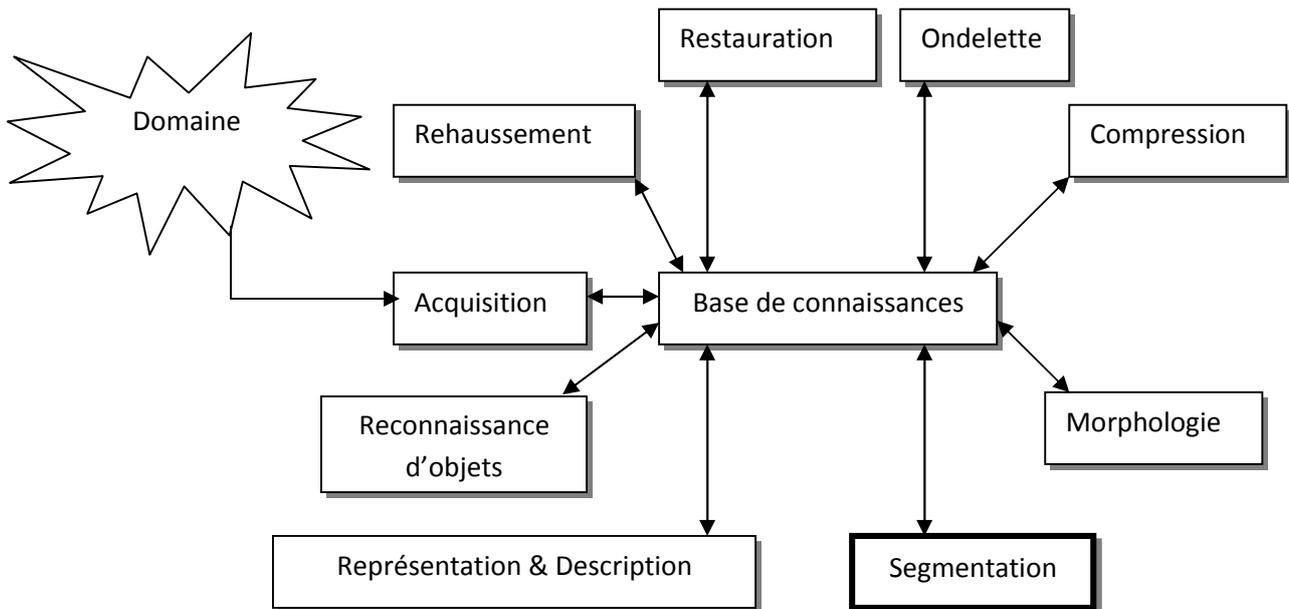


Figure 7 : Traitements fondamentaux en traitement d'images

Acquisition de l'image : Elle a pour objet de passer de la scène physique à une forme numérique observée,

Le rehaussement de l'image :

Le mécanisme de formation des images est loin d'être parfait donc présence de différentes formes de bruit, améliorer le contraste d'où l'objectif est de :

- Rehausser le niveau de gris
- Accentuer les caractéristiques
- Filtrer pour atténuer le bruit

La restauration d'image : Son but est d'améliorer la qualité d'une image, atténuer, supprimer les dégradations.

Les ondelettes : elles permettent de représenter l'image en différentes résolutions. Elles sont utilisées aussi en compression.

La compression : c'est la technique permettant de réduire la taille de l'image pour des raisons de stockage ou de bande passante pour pouvoir la transmettre sur un réseau.

La morphologie : c'est outil permettant d'extraire des composantes d'une image pour décrire et représenter différentes formes.

La segmentation : c'est une procédure permettant de partitionner l'image en ses constituants ou objets. *La segmentation automatique* est la tâche la plus difficile en traitement d'images. Plus la segmentation est meilleur plus l'étape de reconnaissance d'objets est réussite.

La représentation & description : *la représentation* est l'étape qui vient juste après la segmentation. Le résultat de la segmentation est un ensemble de pixels relatifs à une région. Ces données doivent être converties en une forme traitable par un ordinateur. On peut soit représenter la région ou sa frontière. *La description* est l'extraction d'attribut permettant de distinguer une classe d'objets d'une autre classe.

La reconnaissance : est le traitement qui affecte une étiquette (exemple : route, voiture,..) à un objet en se basant sur ses descripteurs.

La base de connaissance contient la connaissance du domaine du problème en cours du traitement. Dans son aspect le plus simple, elle peut consister en coordonnées de l'objet à traiter, ceci permet de réduire l'espace de recherche. Comme elle peut être complexe contenant toutes les défaillances que peut présenter un produit manufacturé.

Conclusion

L'image est une information riche. Dans ce chapitre nous avons présenté la notion d'image, sa définition et ses caractéristiques (notamment la couleur et la texture) ainsi que les différents traitements appliqués sur l'image numérique. L'image, richesse en information, nous entoure partout, le nombre d'images générées ne cesse d'augmenter. Le chapitre suivant traitera la problématique de la recherche d'image pour satisfaire le besoin de l'utilisateur en information.

chapitre 2

Recherche d'image

Chapitre2 : Recherche d'images

Introduction

La recherche d'image est la réponse à une requête d'un utilisateur pour trouver une ou plusieurs images similaires dans une base d'images. Cette dernière peut contenir plusieurs médias : média statique (Texte, Image) et/ou dynamique (Son et Vidéo). Le nombre d'images ne cesse d'augmenter que ce soit au niveau des particuliers ou au niveau des professionnels.

Voici quelques chiffres pour donner un ordre d'idée : Google recensait environ 2 milliard d'images présentes sur Internet le 9 août 2005. R.J. Pittmann, directeur produit des services Consumer Search de Google, estime à 1 000 000 000 000 le nombre d'images et photos disponibles en ligne sur internet à l'heure actuelle (croissance phénoménale due notamment à l'explosion de la photographie numérique et des téléphones portables). Il n'est pas imaginable de consulter toutes ces images à la main pour retrouver celles que l'on recherche. Pour utiliser efficacement ces bases d'images de manière automatique, un système d'indexation et de recherche d'images est nécessaire. C'est pourquoi le sujet de la recherche d'images devient un sujet très actif dans la communauté internationale depuis plus d'une dizaine d'années.

Les systèmes classiques de recherches d'images peuvent être classés en deux catégories :

- Les systèmes textuels de recherches d'images (TBIR),
- Les systèmes de recherches d'images par le contenu visuel (CBIR).

Ils possèdent tous les deux une phase d'indexation et une phase de recherche. Le terme « indexation » est parfois ambigu, car il est utilisé pour deux problèmes distincts :

- Le processus d'extraction des informations à partir des documents,
- La représentation de cette information.

L'indexation a donc pour but d'extraire et de représenter le sens d'un document de manière à ce qu'il puisse être retrouvé par l'utilisateur. Nous utiliserons indifféremment les deux emplois de ce terme.

Dans ce chapitre les systèmes de recherche d'image sont étudiés.

II.1. Systèmes de recherche d'images [3]

Les systèmes de recherche d'images classiques comportent plusieurs phases :

- La phase d'indexation (effectuée hors ligne),
- La phase de recherche,
- Éventuellement, une phase de bouclage de pertinence (*relevance feedback*).

La **Figure (8)** montre un schéma classique des systèmes de recherche d'information. [3]

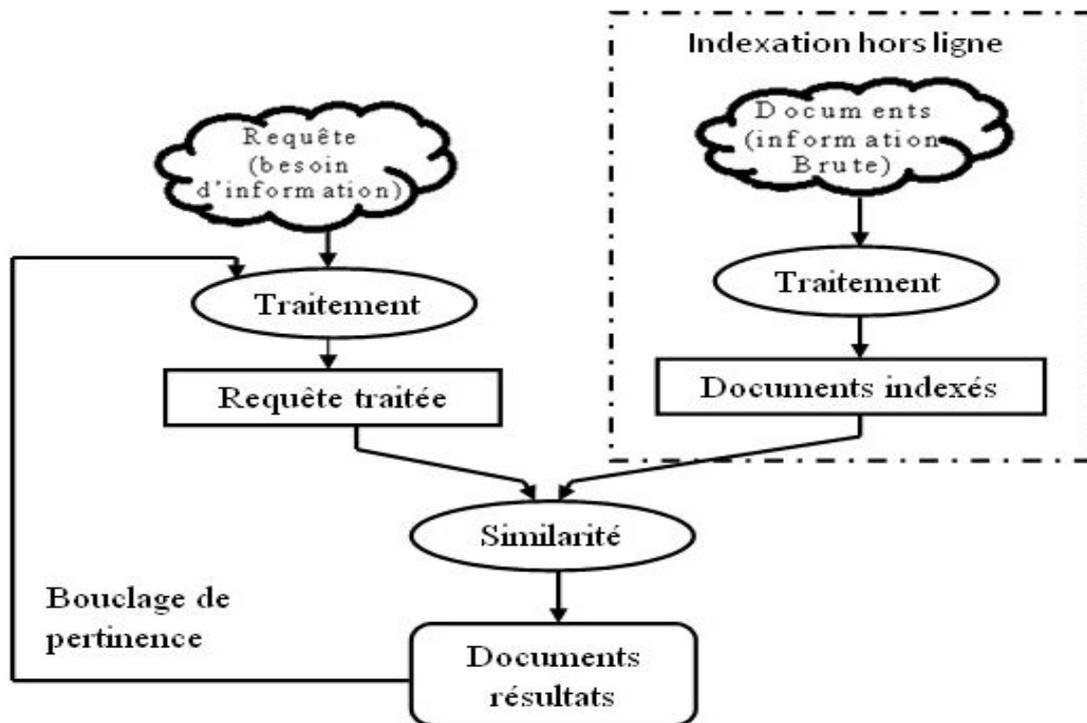


Figure 8 : Schéma d'un système classique de recherche d'information

Il existe principalement deux types de systèmes de recherche d'images :

- **Les systèmes textuels de recherche d'images** : les images sont indexées à partir du texte associé à chaque image et la requête est composée de quelques mots-clés. C'est le cas par exemple des moteurs de recherche d'images sur le web.

- **Les systèmes de recherche d'images par le contenu visuel** : des descripteurs visuels (couleurs, textures, formes...) sont extraits dans des images et la requête est composée de descripteurs visuels souvent fournis sous la forme d'une image exemple. [3]

Le **Tableau 1** résume les principales caractéristiques des trois types de systèmes de recherche d'images.

(*) Même s'il est possible d'avoir dans la base des documents uniquement textuels ou uniquement visuels, il est important, pour la phase d'apprentissage, que le nombre de documents bimodaux soient conséquents.

(**) Le système peut fournir comme résultats des textes, même dans le cas où la requête est une image (par exemple, le système annote l'image requête avec des mots et recherche alors les textes qui contiennent les mêmes mots), mais dans ce cas ce n'est plus un système de recherche d'images.

Tableau 1 : Comparaisons des caractéristiques de trois types de systèmes de recherche d'images. [3]

	Système textuel de recherche d'images	Système de recherche d'images par le contenu visuel	Système de recherche d'images fusionnant texte et images
documents contenant l'information	Textes	Images	images (*), textes (*) ou les deux
Indexation	Textuelle	Visuelle	textuelle, visuelle ou les deux
Requête	Textuelle	Visuelle	textuelle, visuelle ou les deux
documents résultats	Images	Images	images, textes ou les deux (**)

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

Un Système d'indexation et de recherche d'images permet aux utilisateurs de retrouver les images qui satisfont leurs besoins dans des bases de données d'images.

Son fonctionnement est décomposé en deux étapes : l'étape d'indexation dit, hors ligne (offline) ou le système va extraire les caractéristiques des images de la base et les stocker dans une base de données. La deuxième étape consiste à extraire les descripteurs de l'image requête et la comparer avec les descripteurs existant dans la base de données afin de trouver les images similaires à celle désirée. [6]

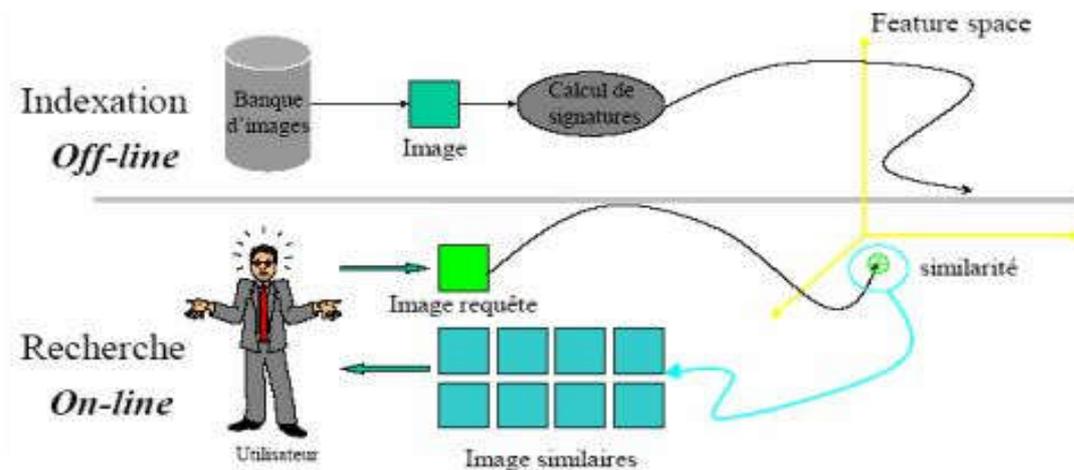


Figure 9 : Aperçu d'un système d'indexation et de recherche d'image.

Les Types de requêtes

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 :

- a. Requête par mot clés.
- b. Requête par esquisse (croquis) et
- c. Requête par exemple. [7]

La Figure 10 illustre les grandes catégories de requête :

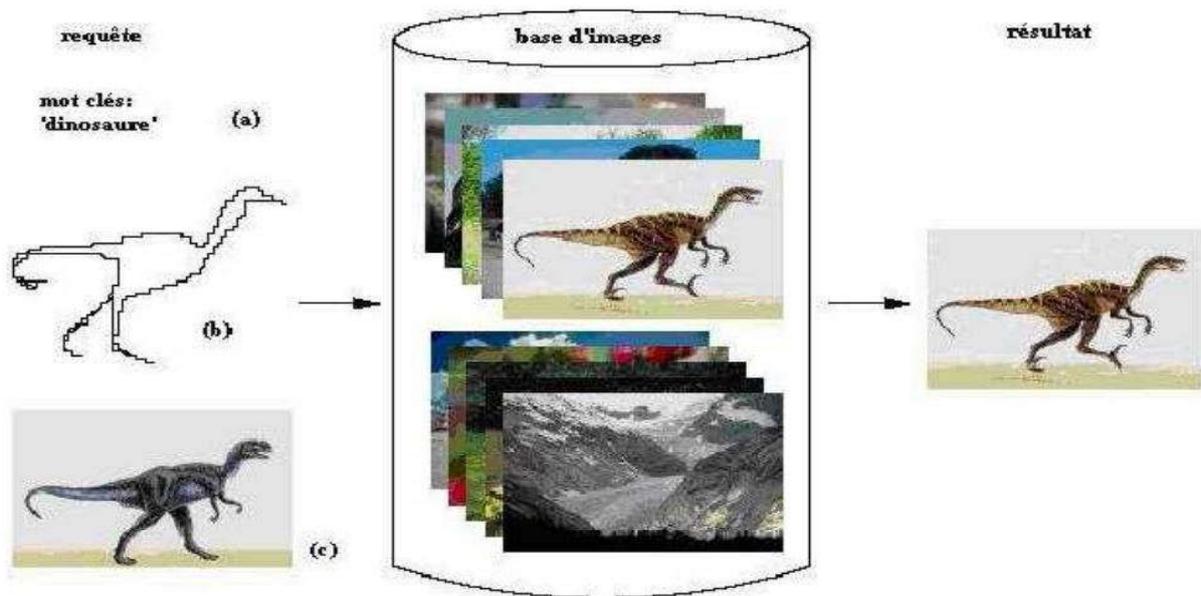


Figure 10 : Les Différents types de requêtes

A. Requête par mots clés

Les images sont recherchées suivant un ou plusieurs critères, par exemple trouver les images contenant 80% de rouge.

B. 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. [7]

C. 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. [7]

II.3. Recherche d'Image par le Contenu (CBIR)

La recherche d'images par le contenu (RIC en anglais : CBIR), est une des applications de la vision par ordinateur permettant de rechercher des images dans de grandes bases de données à partir de leurs caractéristiques visuelles. [14]

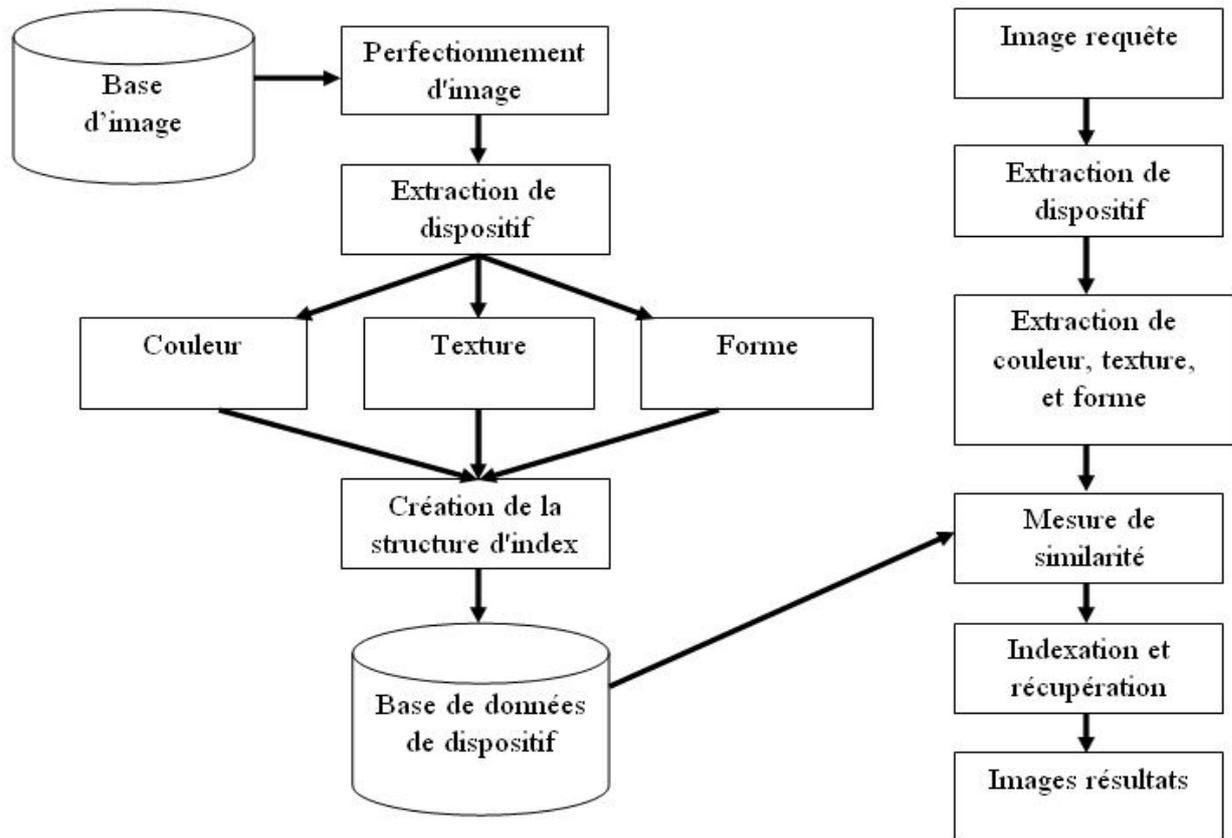


Figure 11 : Approche général concernant CBIR [17]

Par analogie avec le domaine de la RI, les images peuvent être considérées comme des documents et un système de recherche d'images est défini comme « le système qui s'intéresse à la navigation, la recherche et l'extraction d'images à partir d'une grande base de données d'images numériques » [13]

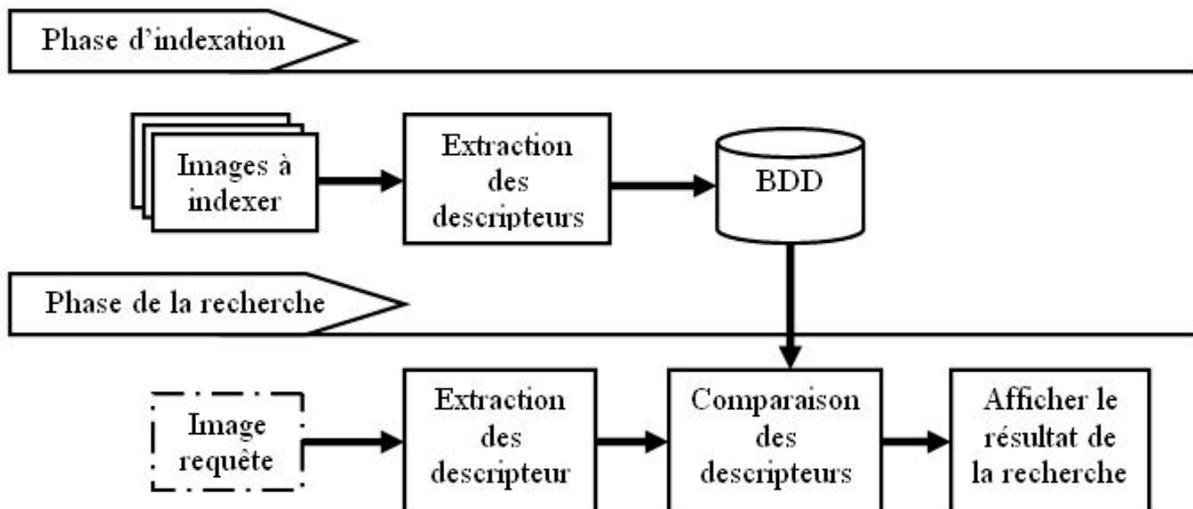


Figure 12 : Schéma standard d'une approche d'indexation par le contenu visuel

1. L'indexation

La phase d'indexation est un processus hors ligne (en anglais : offline). Il consiste à représenter les images d'une collection par leurs contenus visuels. Ainsi, chaque image est indexée sous une forme condensée représentée par des vecteurs de caractéristiques visuelles, par des histogrammes de couleurs, histogrammes de texture, etc. [13]

Dans cette phase, des caractéristiques sont automatiquement extraites à partir de l'image et stockées dans un vecteur numérique appelé descripteur visuel. Grâce aux techniques de la base de données, on peut stocker ces caractéristiques et les récupérer rapidement et efficacement. [16]

2. La Segmentation

La segmentation d'image est le processus de diviser une image numérique dans des segments. Le but de la segmentation est de simplifier et/ou changer la représentation d'une image en quelque chose qui est plus significative et plus facile à analyser. La segmentation d'image est typiquement employée pour localiser des objets et des frontières (lignes, courbes, ..., etc.) dans les images. Et avec précision, la segmentation d'image est le processus d'assigner à une étiquette à chaque Pixel dans une image tels que les Pixel avec la même chose marquent la part certaines caractéristiques. Le résultat de la segmentation d'image est un ensemble de segments qui couvrent collectivement l'image entière, ou un ensemble de découpes extraites à partir de l'image. Chacun des Pixel dans une région est semblable en ce

qui concerne une certaine caractéristique ou propriété calculée, telle qu'une couleur, intensité, ou une texture. Les régions limitrophes sont sensiblement différentes en ce qui concerne le même caractéristique(s). [18]

3. La recherche où l'appariement requête-image

Permet de mesurer la pertinence des images indexées vis-à-vis d'une requête image. Cette pertinence est représentée par un score qui peut être calculé à partir des mesures de distance entre les représentations des descripteurs visuels d'images (par exemple, distance Euclidienne, distance de Hamming, cosinus, Manhattan, Minkowski...). [13]

Dans cette étape, le système prend une ou des requêtes à l'utilisateur et lui donne le résultat correspond à une liste d'images ordonnées en fonction de la similarité entre leur descripteur visuel et celui de l'image requête en utilisant une mesure de distance. [16]

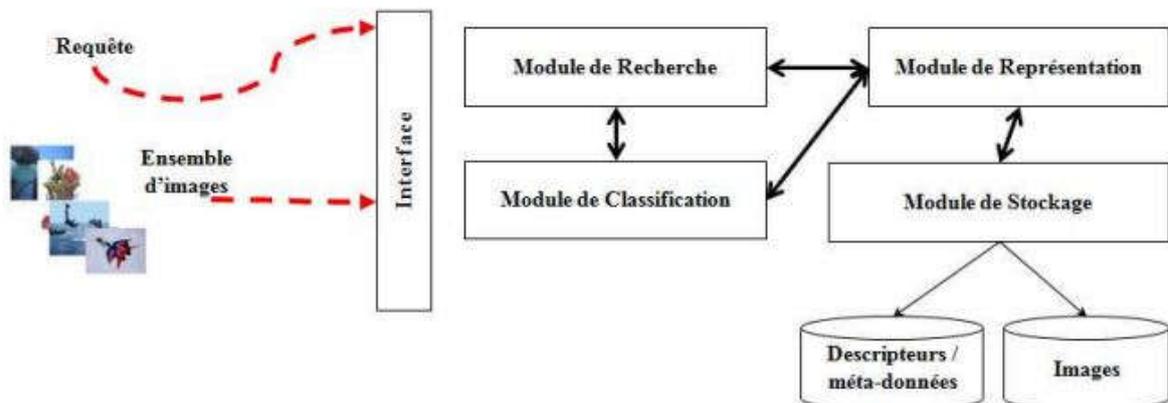


Figure 13 : Architecture en module de CBIR

Recherche d'image par le contenu basé sur plusieurs étapes :

- **Extraction des caractéristiques de l'image** : couleur, texture, forme, contraintes spatiales, régions

II.4. Descripteurs d'images

Aujourd'hui avec le développement des systèmes multimédias et le recul de l'écrit, nous utilisons de plus en plus le contenu visuel comme support de communication dans différents domaines.

Undescripteur est défini comme la connaissance utilisée pour caractériser l'information contenue dans les images. De nombreux descripteurs sont utilisés dans les systèmes de recherche pour décrire les images. Ceux-ci peuvent être différenciés selon deux niveaux : [4]

1. **Les descripteurs bas niveau** : ils décrivent le contenu bas niveau de l'image, principalement en termes de couleurs, textures et de formes. Ce sont les descripteurs les plus utilisés dans les systèmes actuels, car ils sont plus simples à mettre en place.
2. **Les descripteurs haut niveau** : ils décrivent le contenu sémantique de l'image, et sont principalement des mots clés fournis par l'utilisateur lors de l'indexation. Ces attributssémantiques peuvent être le résultat d'un processus d'analyse sémantique automatisé utilisant des approches de classification ou des ontologies spécialisées.

II.4.1 Descripteurs de couleurs

La couleur est l'information visuelle la plus utilisée dans les systèmes de recherche par le contenu.

II.4.1.1. L'espace de couleurs

Chaque pixel d'image peut être représenté comme un point dans un espace 3D. Les espaces les plus communément utilisés dans les CBIR sont : l'espace RGB, Cie L*a*b*, CIE L*u*v*, HSV (ou HSL, HSB).

Les images sont codées sur trois canaux contre un seul pour les monochromes. [1]

1. **Système RGB**

Le système le plus couramment utilisé est le système RGB qui est le système des trois couleurs fondamentales. Il associe à chaque couleur trois composantes (ou canaux), qui correspondent aux intensités respectives de trois couleurs primaires de la synthèse additive. Le blanc correspond à la valeur maximale pour chaque canal, tandis que le noir correspond aux trois composantes nulles. [1]

Par la variation de leur combinaison, d'autres couleurs peuvent être obtenues. Ce modèle peut être représenté sous forme d'un cube unitaire (**Figure14**) avec : noir (0,0,0), blanc(1,1,1), rouge (1,0,0), vert (0,1,0), et le bleu (0,0,1). Et les couleurs secondaires : cyan (0, 1,1), magenta (1, 0,1), et le jaune (1, 1,0). [2]

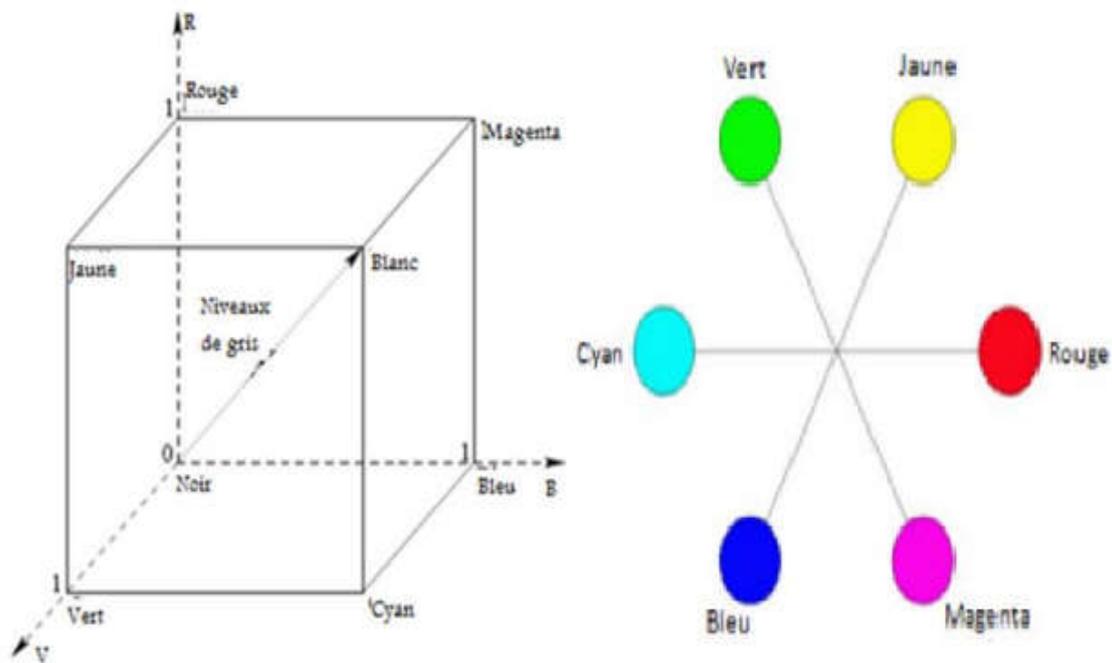


Figure 14 : Le modèle de couleur RVB

2. Système XYZ [5]

La CIE a défini un autre espace de représentation d'une couleur qui prend en compte la sensibilité de l'œil. Il s'agit de l'espace XYZ. Il s'obtient simplement à partir du système RGB

et à l'aide d'une matrice de passage. Y contient la luminance, X et Z l'information de chrominance. Cet espace lui aussi n'est pas perceptuellement uniforme ce qui le rend rarement utilisé dans les systèmes de recherches d'images.

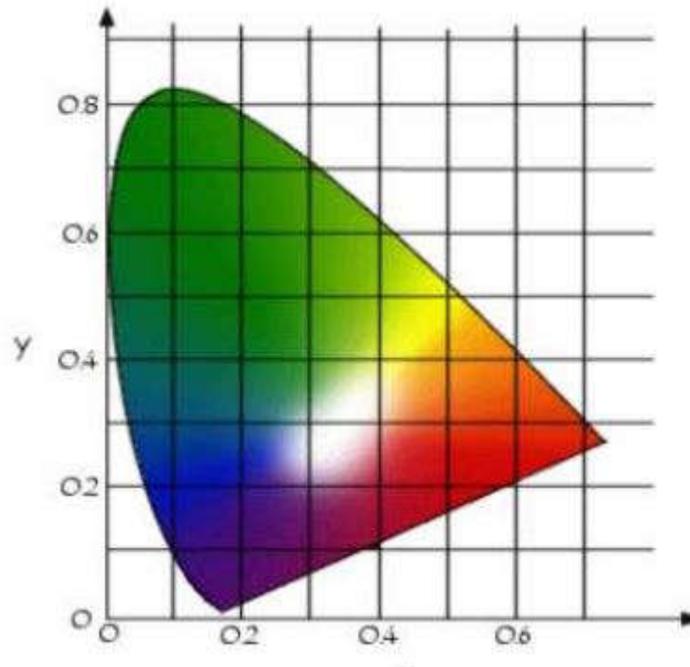


Figure 15 : Espace de couleur XYZ.

3. Système L^*u^*v (CIE Luv) [5]

C'est un système perceptuellement uniforme que l'on peut obtenir à partir du système XYZ. Il permet l'utilisation d'une distance Euclidienne pour mesurer l'écart entre deux couleurs. La composante L représente la réponse de l'œil à un niveau de luminance, u et v représentent la chrominance.

4. Système $L^*a^*b^*$ [4]

$L^*a^*b^*$ (CIELab), dans lequel une couleur est repérée par trois valeurs :

- L, la luminance, exprimée en pourcentage (0 pour le noir à 100 pour le blanc)

- a et b deux gammes de couleur allant respectivement du vert au rouge et du bleu au jaune avec des valeurs allant de -120 à +120.

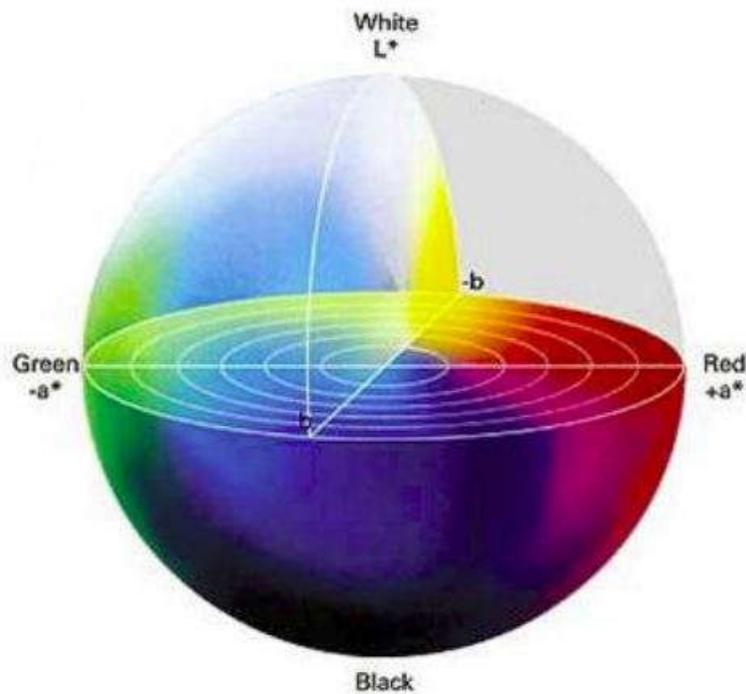


Figure 16 : Espace de couleur $L^*a^*b^*$

5. Système HSV (HSI) [5]

La projection d'une image dans cet espace permet de séparer les trois informations relatives à la teinte (H), qui représente la couleur pure (rouge, jaune, vert, bleu,...etc.), la saturation (S) qui donne des informations sur la quantité de blanc dans la couleur pure et la valeur (V) correspondant à l'intensité lumineuse qui indique si la couleur est claire ou sombre.

On représente cet espace par un cône, dont la teinte est représentée par les angles du cône, la saturation est donnée par les rayons du cône et la valeur est définie par hauteur du cône, illustré par la **figure (17)**.

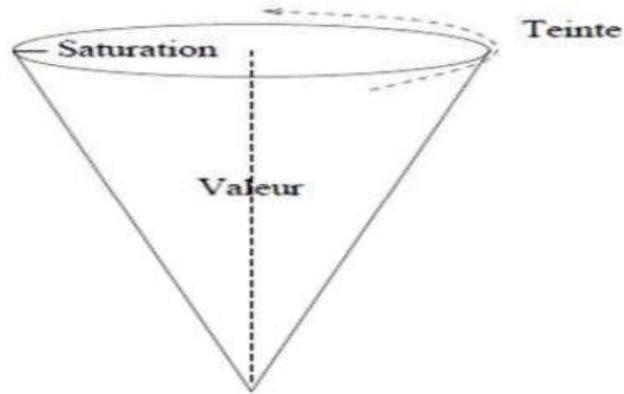


Figure 17 : Espace de couleur HSV.

6. Système YCbCr[4]

L'œil humain est plus sensible à la luminance (où luminosité) qu'à la chrominance d'une image. Or la luminance est présente dans les trois couleurs rouges, vert, bleu. Ainsi on transforme les composantes RGB de l'image en une composante de luminance notée Y et en deux composantes de chrominance notées Cb et Cr . Pour cela on utilise des formules de passage (pour un pixel) :

$$Y = 0.2989 \times R + 0.5866 \times G + 0.1145 \times B$$

$$Cb = -0.1688 \times R - 0.3312 \times G + 0.5000 \times B$$

$$Cr = 0.5000 \times R - 0.4184 \times G - 0.0816 \times B$$

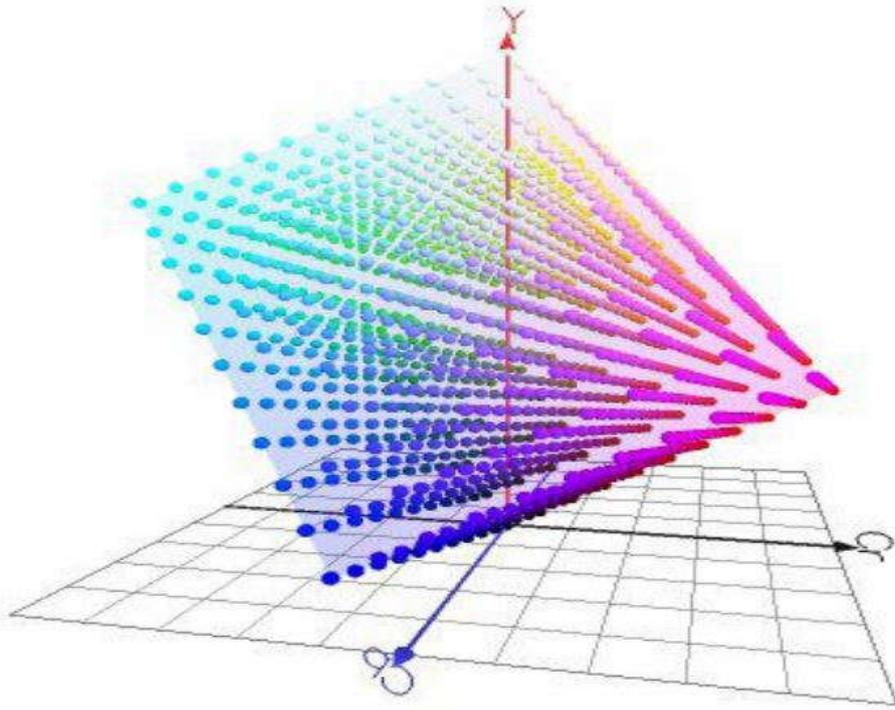


Figure 18 : Espaces de couleur YCbCr

Cette transformation des composantes RGB en composantes YCbCr est linéaire et bijective. Ainsi il n'y a pas de pertes de données pendant cette étape.

II.4.1.2. Histogrammes

L'histogramme de couleurs est le descripteur de couleurs le plus courant et le plus rencontré dans la littérature. L'histogramme représente la répartition des pixels en fonction de leur niveau de gris. [4]

Un histogramme est un outil statistique qui permet d'estimer la densité de probabilité d'une distribution à partir d'échantillons. L'intervalle des valeurs possibles est divisé en classes, puis pour chacune d'elles on compte le nombre d'échantillons associés.

Comme un système de couleur est composé de plusieurs canaux, son histogramme a plusieurs dimensions. L'histogramme peut alors être calculé de deux façons :

- Indépendamment sur les différents canaux. Il contient alors $\sum_{i=1}^n n_i$, si on note n le nombre de canaux et n_i le nombre de classes pour chaque canal.

- Par une discrétisation de l'espace complet. Le nombre de classes est alors $\prod_{i=1}^n n_i$. [1]

Les histogrammes sont faciles et rapides à calculer, et robustes à la rotation et à la translation. Cependant l'utilisation d'histogrammes pour l'indexation et la recherche d'images pose quatre problèmes :

1. Ils sont de grandes tailles, donc par conséquent il est difficile de créer une indexation rapide et efficace en les utilisant tels qu'ils sont.
2. Ils ne possèdent pas d'informations spatiales sur les positions des couleurs.
3. Ils sont sensibles à de petits changements de luminosité, ce qui est problématique pour comparer des images similaires, mais acquises dans des conditions différentes.
4. Ils sont inutilisables pour la comparaison partielle des images (objet particulier dans une image), puisque calculés globalement sur toute l'image. [2]

II.4.1.3. Les moments statistiques

Les moments de couleur ont été utilisés dans plusieurs systèmes de recherche d'images par le contenu tel que QBIC, mathématiquement les trois premiers moments sont définis par :

$$E_i = \frac{1}{n} \sum_{j=1}^N P_{ij}$$

$$\delta_i = \left(\frac{1}{N} \sum_{j=1}^N (p_{ij} - E_i)^2 \right)^{\frac{1}{2}}$$

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

Où P_{ij} est la valeur de la ième composante chromatique du pixel j, et N le nombre de pixels de l'image [2], E_i est l'espérance, δ_i la variance et s_i est le troisième moment qui correspond à la composante couleur i. [1]

II.4.1.4. Cohérence spatiale

Ce descripteur a pour but de combler en partie, l'absence d'information spatio-colorimétrique de l'image dans le descripteur précédent. La cohérence spatiale est calculée pour chaque classe de couleur identifiée. Tout d'abord un histogramme de connexité est calculé :

$$H_I(c) = \sum_{i=0}^{X=I} \sum_{j=0}^{X=I} \delta(I(i,j), c) \alpha(i,j)$$

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

$$\alpha(i,j) = \begin{cases} 1 & \text{si } \forall k, \hat{k} \in (-w, w) I(i+k, j+\hat{k}) = I(i,j) \\ 0 & \text{sinon} \end{cases}$$

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

$$SCR(c) = \frac{H_I(c)}{H(c)}$$

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

II.4.1.5. Couleurs dominantes [4]

Il est défini par :

$$F = \{(C_i, P_i, V_i, S)\}, (i = 1, 2, \dots, N)$$

Ou N désigne le nombre de couleurs dominantes, C_i la valeur de la $i^{\text{ème}}$ 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.

Considérons les deux descripteurs ci-dessous :

$$F_1 = \{(C_{1i}, P_{1i}, V_{1i}, S_1)\}, (i = 1, 2, \dots, N_1)$$

$$F_2 = \{(C_{2i}, P_{2i}, V_{2i}, S_2)\}, (i = 1, 2, \dots, N_2)$$

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

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

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

II.4.1.6. Corrélogramme [4]

Un corrélogramme présente les corrélations spatiales entre les niveaux de gris en fonction de la distance inter pixel. Pour une image A de N pixels, quantifiée sur n niveaux de gris $\{c_1, c_2, \dots, c_n\}$.

Chaque pixel p en (x, y) est doté d'un niveau de gris $A(p)$:

$$P \in A_c = A(P) = c$$

Pour une distance $d \in \{1, 2, \dots, N\}$ fixée à priori, le corrélogramme du niveau de gris i par rapport à j est défini par :

$$\gamma_{ci,cj}^{(k)} = \text{prob} \left| \begin{array}{l} p_1 \in A_{cj} / |p_1 - p_2| = k \\ p_2 \in A_{ci} \end{array} \right|$$

$$\forall k \in \{1, 2, \dots, d\}$$

Pour tout pixel de niveau de gris c_i dans l'image, la valeur $\gamma_{c_i, c_j}^{(k)}$ donne la probabilité de trouver à une distance k un pixel de niveau de gris c_j .

Le corrélogramme couleur exprime comment la corrélation spatiale de couples de couleurs évolue en fonction de la distance dans l'image. Il prend simultanément en compte la distribution globale des couleurs des pixels ainsi que la corrélation des couleurs entre pixels voisins.

II.4.2. Descripteurs de textures

Au même titre que la couleur, la texture est une caractéristique fondamentale des images car elle concerne un élément important de la vision humaine. De nombreuses recherches ont été menées à la fois dans les domaines de l'analyse et de la synthèse de texture. [2]

La texture est une information de plus en plus utilisée en indexation et la recherche d'images par le contenu, car elle permet de pallier certains problèmes posés par l'indexation par la couleur, notamment lorsque les distributions de couleur sont très proches. [1]

L'étude de la texture des objets d'une image peut avoir des objectifs très divers :

- Obtenir des informations sur la nature d'un l'objet,
- Segmenter l'image en régions homogènes,
- Identifier la texture afin de la réduire à un ensemble de paramètres (compression d'images),
- Recherche d'image par contenu,
- etc. [1] [2]

La notion de texture est liée à trois concepts principaux [1] [2] :

1. Un certain ordre local qui se répète dans une région de taille assez grande,
2. Cet ordre est défini par un arrangement structuré de ses constituants élémentaires,
3. Ces constituants élémentaires représentent des entités uniformes qui se caractérisent par des dimensions semblables dans toute la région considérée.

Il existe un grand nombre de textures. On peut les séparer en deux classes :

- A. Les textures structurées (macro textures) et
- B. Les textures aléatoires (micro textures). [1] [2]

Les attributs texturaux peuvent être obtenus à partir d'un ensemble assez vaste de différentes théories mathématiques. Citons notamment :

- Les attributs fondés sur des calculs statistiques effectués sur les niveaux de gris des pixels de l'image. C'est le cas des statistiques classiques, et des matrices de cooccurrences ou de longueurs de plages ainsi que les méthodes utilisant directement la fonction de covariance ou les statistiques d'ordre supérieur.
- Les attributs obtenus à la suite de transformations orthogonales appliquées aux images (transformées de Fourier, Ondelettes, ... etc.). Les attributs texturaux seront alors calculés dans des domaines différents de celui de la grille spatiale des luminances (domaine spectral par exemple).
- D'autres méthodes, basées par exemple sur la morphologie mathématique, les intégrales curvilignes, l'application de filtres ou sur la théorie fractale peuvent également être employées pour obtenir des attributs texturaux. [1] [2]

II.4.2.1. Schéma général [1]

Le principe le plus général de l'analyse statistique des textures est le suivant :

1. On définit un voisinage V_{ij} de tout pixel (i, j) , de taille et de forme appropriés.
2. Sur le voisinage V_{ij} on mesure des propriétés particulières de l'image : soit $\prod k$ la mesure attachée à la configuration parmi les N mesurées. Ces mesures seront attachées au pixel (i, j) .
3. On classe les pixels à partir du vecteur formé par les $\prod k$ par l'une des nombreuses méthodes de la reconnaissance des formes.
4. Eventuellement, avant cette classification, et si le vecteur de mesures est trop grand, on réduit la dimension de l'espace par une sélection judicieuse des composantes les plus significatives ou par analyse en composantes principales.

II.4.2.2. La fenêtre d'analyse [1]

La dimension du voisinage V_{ij} est importante. C'est l'un des choix délicats de l'analyse des textures. Elle doit comprendre au moins un motif de base de la texture mesurée pour

fournir des statistiques homogènes sur la texture, mais, si V_{ij} est trop grand, la précision de localisation des frontières des textures détectées sera médiocre.

II.4.2.3. Matrices de cooccurrences [1]

En 1973, Haralick a proposé une méthode en se basant sur les matrices de cooccurrences de niveaux de gris. La texture d'une image peut être interprétée comme la régularité d'apparition de couples de niveaux de gris selon une distance donnée dans l'image. La matrice de cooccurrences contient les fréquences spatiales relatives d'apparition des niveaux de gris selon quatre directions :

$$\theta = 0, \theta = \frac{\pi}{4}, \theta = \frac{\pi}{2}, \theta = \frac{3\pi}{4}$$

Une matrice de cooccurrences est définie au moyen d'une relation géométrique \mathcal{R} entre deux pixels (x_1, y_1) et (x_2, y_2) .

La matrice de cooccurrences $P_{d,\theta}(i, j)$ est carrée et de dimension $\Delta * \Delta$, où Δ est le nombre de niveaux de gris présents dans I . Les indices de la matrice de cooccurrences sont donc les niveaux de gris de la texture étudiée.

On définit la matrice de cooccurrences :

$$P_{d,\theta} \text{ par } P_{d,\theta} = P_{d,\theta}(i, j)$$

$P_{d,\theta}(i, j)$ Représente le nombre de fois où un couple de points séparés par la distance d dans la direction θ ayant les niveaux de gris I_i et I_j . Pour obtenir de véritables fréquences relatives, il faut normaliser les éléments de la matrice en les divisant par le nombre total de paires de points élémentaires séparés par la distance d dans la direction θ dans toute l'image.

Algorithme. Matrice de cooccurrences [1]

Soit $L_x = \{1, 2, \dots, N_x\}$ et $L_y = \{1, 2, \dots, N_y\}$ les domaines spatiaux verticaux et horizontaux de l'image et soit $\Delta = \{0, 1, 2, \dots, F - 1\}$ les niveaux de gris.

L'image numérique $I: L_x * L_y \rightarrow \Delta$ affecte un niveau de gris à chaque pixel.

Soit ' q ' une direction, et une distance ' d ' d'étude.

- i- Initialisation de la matrice : $\forall i, j \in [0, L]: P_{d,\theta}(i, j) = 0$
- ii- Remplissage de la matrice. Pour tout point (x_1, y_1) d'intensité $I(x_1, y_1)$, le couple distance-angle (d, θ) ; Si la relation R entre deux pixels (x_1, y_1) et (x_2, y_2) est respectée :

$$P_{d,\theta}(i, j) = \sum_{x_1 \in L_x} \sum_{y_1 \in L_y} \delta(I(x_1, y_1) = i \text{ et } I(x_2, y_2) = j)$$

Exemple : Soit l'image suivante comportant $\Delta = 4$ niveaux de gris $\{0, 1, 2, 3\}$;

$$I(x, y) = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 2 & 2 & 3 \\ 2 & 2 & 3 & 3 \end{bmatrix}$$

$$P_{d,0^\circ}(i, j) = \{(x_1, y_1), (x_2, y_2) \in I / y_1 = y_2, |x_2 - x_1| = d, I(x_1, y_1) = i \text{ et } I(x_2, y_2) = j\}$$

Les matrices $P_{1,0^\circ}$ et $P_{1,90^\circ}$ sont donc de dimension 4 et valent

$$P_{1,0^\circ} = \begin{bmatrix} 6 & 2 & 1 & 0 \\ 2 & 2 & 0 & 0 \\ 1 & 0 & 4 & 2 \\ 0 & 2 & 2 & 2 \end{bmatrix} \quad P_{1,90^\circ} = \begin{bmatrix} 6 & 1 & 2 & 0 \\ 1 & 2 & 1 & 1 \\ 2 & 1 & 2 & 1 \\ 0 & 1 & 1 & 2 \end{bmatrix}$$

II.4.3. Descripteurs texturaux [1]

Haralick a proposé 14 caractéristiques statistiques extraites à partir de cette matrice. Actuellement, seulement les quatre caractéristiques les plus appropriées sont largement utilisées :

L'homogénéité h : $h = \sum_{i=1}^n \sum_{j=1}^n P_{d,\theta}(i,j)^2$

L'entropie e : $e = \sum_{i=1}^n \sum_{j=1}^n P_{d,\theta}(i,j) \log_2 P_{d,\theta}(i,j)$

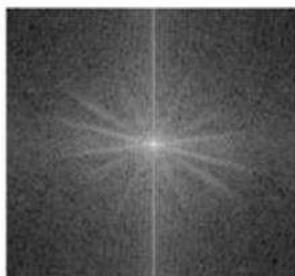
Le contraste c : $e = \sum_{k=0}^{n-1} \left(\sum_{i=1}^n \sum_{j=1}^n P_{d,\theta}(i,j) \right), |i-j| = n$

Moment des Différences inverses m : $m = \sum_i \sum_j \frac{1}{1+(i-j)^2} P_{d,\theta}(i,j)$

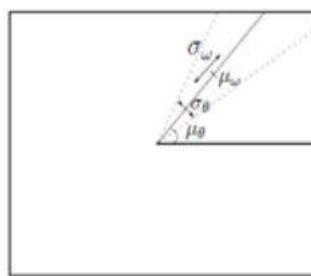
Un inconvénient de cette approche est le coût en termes de temps de calcul. Comme pour les corrélogrammes, la complexité algorithmique de la création de toutes les matrices est assez lourde, et nécessite une optimisation.

II.4.3.1. Filtres de Gabor [1]

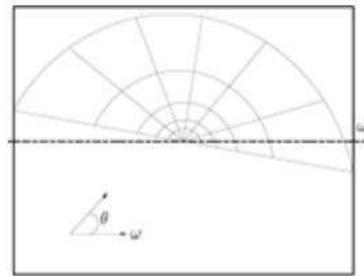
Les filtres de Gabor sont très utilisés en indexation, pour la description de la texture. Ils sont notamment utilisés par la norme MPEG-7. Ces filtres sont généralement exploités dans l'espace de Fourier dans le but de caractériser des textures locales.



Transformée de Fourier



Paramètres d'un filtre



Découpage du plan de Fourier

Figure 19 : Description d'un filtre de Gabor

L'espace de Fourier est découpé en plusieurs régions. Pour une région donnée (ω, θ) , les termes μ_θ et μ_ω représentent respectivement l'angle par rapport à l'axe d'origine et la position radiale, σ_θ et σ_ω représentent leur tolérance. Ce découpage est utilisé par la norme MPEG-7. Ces images ont été recentrées pour que l'origine du plan de Fourier corresponde au centre de l'image.

L'utilisation des filtres de Gabor consiste à analyser indépendamment différentes parties de l'espace de Fourier, à l'aide de plusieurs filtres. Chacun d'eux est défini par de plus fortes valeurs de coefficients à une fréquence et une orientation donnée : ainsi il est très sensible aux fréquences situées au centre de la région d'intérêt, et pratiquement pas partout ailleurs. Tous les filtres sont modélisés par la même loi, une gaussienne, avec des paramètres différents. Si on découpe l'espace de Fourier en M positions radiales et N angles, l'équation du filtre de Gabor ayant un indice de position r et un indice d'angle a est donnée, en coordonnées polaires sur l'espace de Fourier, et en associant l'origine au coefficient des basses fréquences, par :

$$G_{r,a}(\omega, \theta) = \exp\left(\frac{-(\omega - \mu_{\omega_r})^2}{2\sigma_{\omega_r}}\right) \exp\left(\frac{-(\theta - \mu_{\theta_a})^2}{2\sigma_{\theta_a}}\right)$$

II.4.3.2. Transformée en ondelettes [1]

La transformée en ondelettes est à la base de nombreuses analyses de texture, telles que les filtres de Haar. La description de texture à base d'ondelettes est utilisée dans pour la recherche d'images.

L'approche continue des ondelettes pour un signal 2D est trop complexe pour être applicable rapidement sur des images. Pour résoudre ce problème, Mallat considère l'analyse en ondelettes comme une décomposition du signal par une cascade de filtres, en utilisant une paire de filtres pour chaque niveau de résolution (un filtre passe-haut et un filtre passe-bas). Il propose ainsi la DWT qui permet d'obtenir une transformée rapide. Le choix de l'ondelette mère est alors remplacé par le choix du filtre. Pour calculer une transformée en ondelettes, on n'a alors besoin que des deux filtres : au lieu de calculer le produit scalaire de l'ondelette avec le signal, on réalise un produit de convolution du signal avec ces filtres.

Les filtres de Haar sont fréquemment employés en apprentissage pour obtenir la description d'un objet (comme un visage ou une personne).

II.4.4. Les descripteurs des points d'intérêts [1]

Les points anguleux (ou points d'intérêt, points saillants,...) sont des points « qui contiennent beaucoup d'information » relativement à l'image. Ce sont des points aux voisinages desquels l'image varie significativement dans plusieurs directions.

Une mesure des variations locales de l'image I au point (x, y) associée à un déplacement $(\Delta x, \Delta y)$ est fournie par la fonction d'auto corrélation :

$$x(x, y) = \sum_{(x_k, y_k) \in W} (I(x_k, y_k) - I(x_k + \Delta x, y_k + \Delta y))^2$$

Où W est une fenêtre centrée au point (x, y) .

Sachant que, en utilisant une approximation du premier ordre :

$$I(x_k + \Delta x, y_k + \Delta y) \approx I(x_k, y_k) + \left(\frac{\partial I}{\partial x}(x_k, y_k) \frac{\partial I}{\partial y}(x_k, y_k) \right) \cdot \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix}$$

Et donc :

$$x(x, y) = \sum_{(x_k, y_k) \in W} \left(\left(\frac{\partial I}{\partial x}(x_k, y_k) \frac{\partial I}{\partial y}(x_k, y_k) \right) \cdot \begin{pmatrix} \Delta x \\ \Delta y \end{pmatrix} \right)^2$$

$x(x, y)$

$$= (\Delta x \ \Delta y) \begin{pmatrix} \sum_{(x_k, y_k) \in W} \left(\frac{\partial I}{\partial x}(x_k, y_k) \right)^2 & \sum_{(x_k, y_k) \in W} \frac{\partial I}{\partial x}(x_k, y_k) \frac{\partial I}{\partial y}(x_k, y_k) \\ \sum_{(x_k, y_k) \in W} \frac{\partial I}{\partial x}(x_k, y_k) \frac{\partial I}{\partial y}(x_k, y_k) & \sum_{(x_k, y_k) \in W} \left(\frac{\partial I}{\partial y}(x_k, y_k) \right)^2 \end{pmatrix}$$

$E(x, y)$: Matrice d'autocorrélation de l'image I en (x, y)

$$E(x, y) = \begin{pmatrix} \sum_{(x_k, y_k) \in W} \left(\frac{\partial I}{\partial x}(x_k, y_k) \right)^2 & \sum_{(x_k, y_k) \in W} \frac{\partial I}{\partial x}(x_k, y_k) \frac{\partial I}{\partial y}(x_k, y_k) \\ \sum_{(x_k, y_k) \in W} \frac{\partial I}{\partial x}(x_k, y_k) \frac{\partial I}{\partial y}(x_k, y_k) & \sum_{(x_k, y_k) \in W} \left(\frac{\partial I}{\partial y}(x_k, y_k) \right)^2 \end{pmatrix}$$

La matrice d'auto corrélation X représente la variation locale de l'image I en (x, y) .

(x, y) sera considéré comme un point anguleux de I si pour tous les déplacements $(\Delta x, \Delta y)$, la quantité $(\Delta x, \Delta y) \cdot \varepsilon(x, y) \cdot (\Delta x, \Delta y)^t$ est grande.

Les points anguleux sont les points (x, y) pour lesquels la matrice d'auto corrélation $X(x, y)$ a deux valeurs propres grandes.

Cela correspond aux points pour lesquels il existe localement une base de vecteurs propres de X décrivant des variations locales importantes de l'image.

Méthodes de détection.

Le détecteur de Moravec (1980).

Le détecteur de Harris (1988).

Extraction de caractéristiques des points d'intérêt par SIFT.

1. Points SIFT.
2. Implémentation.
- 3.

II.4.5. Descripteurs de formes [2]

La forme est un descripteur très important dans l'indexation des images. La forme désigne l'aspect général d'un objet, son contour.

II.4.5.1. Les attributs géométriques de région [2]

Les attributs géométriques de forme permettent de distinguer les différents types de forme que peuvent prendre les objets d'une scène. Ils nécessitent une segmentation en région préalable de l'image. Ils sont ensuite calculés sur les différentes régions de l'image.

La surface relative (ou normalisée) d'une région \mathfrak{R}_k de l'image I est le nombre de pixels contenus dans cette région par rapport au nombre total de pixels de l'image :

$$S_k = \frac{\text{card}(\mathfrak{R}_k)}{\text{hauteur}(I) * \text{largeur}(I)}$$

Le centre de masse des pixels de la région est défini par :

$$P = (P_i, P_j) = \left(\frac{\sum_{i \in \mathfrak{R}_k} i / \text{card}(\mathfrak{R}_k)}{\text{largeur}(I)}, \frac{\sum_{j \in \mathfrak{R}_k} j / \text{card}(\mathfrak{R}_k)}{\text{hauteur}(I)} \right)$$

La longueur du contour de la région est le nombre de pixels en bordure de la région :

$$l_k = \text{card}(\text{contour}(\mathfrak{R}_k))$$

La compacité traduit le regroupement des pixels de la région en zones homogènes et non trouées :

$$C_k = \frac{l_k^2}{S_k}$$

Ces attributs très simples permettent d'obtenir des informations sur la géométrie des régions de l'image. Il existe d'autres attributs de forme, basés sur des statistiques sur les pixels des régions de l'image.

II.4.5.2. Les moments géométriques [1] [2]

Les moments géométriques permettent de décrire une forme à l'aide de propriétés statistiques. Ils sont simples à manipuler mais leur temps de calcul est très long.

Formule générale des moments :

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

L'ordre du moment est $p + q$. 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 :

$$x_c = \frac{m_{1,0}}{m_{0,0}} \text{ et } y_c = \frac{m_{0,1}}{m_{0,0}}$$

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 :

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

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

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

À partir des moments géométriques, Hu a introduit sept invariants aux translations, rotations et changement d'échelle, appelés moments de Hu.

$$M_1 = \mu_{20} + \mu_{02}$$

$$M_2 = (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2$$

$$M_3 = (\mu_{30} - 3\mu_{21})^2 + (3\mu_{21} - \mu_{03})^2$$

$$M_4 = (\mu_{30} - \mu_{21})^2 + (\mu_{21} + \mu_{03})^2$$

$$\begin{aligned} M_5 &= (\mu_{30} - 3\mu_{21})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{21})^2 - 3(\mu_{21} + \mu_{03})^2] \\ &\quad + (3\mu_{21} - \mu_{03})(\mu_{21} + \mu_{03})[3(\mu_{30} + \mu_{12})^2 \\ &\quad - (\mu_{21} + \mu_{03})^2]M_6 \\ &= (\mu_{20} - \mu_{02}) [(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] \\ &\quad + 4\mu_{11}(\mu_{30} + \mu_{12})(\mu_{030} + \mu_{21}) \end{aligned}$$

$$\begin{aligned} M_7 &= (3\mu_{21} - \mu_{03})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})] \\ &\quad - (\mu_{30} - 3\mu_{21})(\mu_{12} + \mu_{03})[3(\mu_{30} + \mu_{12})^2 - (\mu_{12} + \mu_{03})^2] \end{aligned}$$

Les moments de Hu offrent d'excellents descripteurs invariants en translation, rotation et changement d'échelle pour décrire une image. 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.

II.4.5.3. Transformée de Hough

Soit \mathcal{R}^n l'espace image, et ξ un ensemble de N points sélectionnés par un prétraitement :

$$\xi = \{M_i, i = 1 \dots \bar{N}\} \in \mathcal{R}^n$$

Un point M de \mathcal{R}^n est repéré par ses coordonnées x .

Soit $p \Omega \subset \mathcal{R}^p$ un espace de paramètres et F une famille de courbes dans $n \mathcal{R}^n$ paramétrée par a :

$$F = \{x: f(x, a) = 0, \quad x \in \mathcal{R}^n, a \in \Omega\}$$

On appelle transformation de Hough associée à la famille F une transformation qui fait correspondre [1] à l'ensemble x . [2]

Algorithme : Transformée de Hough Standard. [2]

1. Choisir une espace discret (ρ, θ) .

2. Initialiser un accumulateur $A(\dim(\rho), \dim(\theta))$ à zéro.
3. Pour chaque pixel :
 - Si $I(u, v) = 1$
 - Calculer $\theta = u * \cos(\theta) + v * \sin(\theta)$,
 - Incréments $A(\text{round}(\rho), \text{round}(\theta))$.
4. Recherche-le ou les maximums dans la matrice A .
5. Les couples (ρ, θ) sélectionnés caractérisent les lignes de l'image.

Les formes représentent un descripteur puissant pour décrire les objets contenus dans l'image. [2]

- Représentation de ces caractéristiques en vue de leur stockage dans la base de données.
- Recherche sur ces informations extraites. [11]

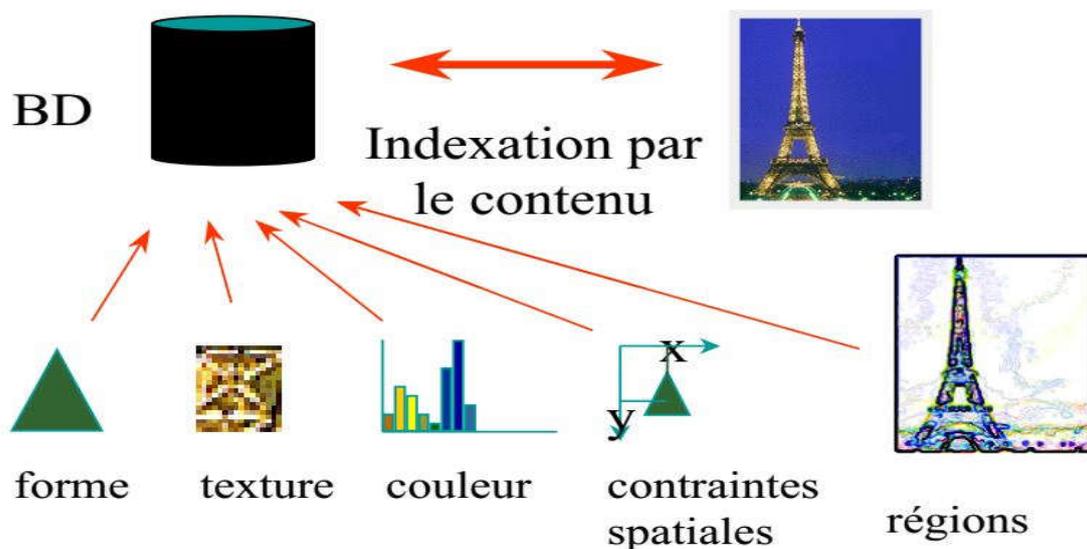


Figure 20 : recherche d'image basée sur le contenu

II.5. Mesures de similarité

Afin de déterminer les descripteurs d'images, on peut calculer la valeur de similarité entre les images de base et notre image requête.

Plusieurs mesures de similarité sont basées sur la distance L_p entre deux points. Pour deux points donnés (x, y) dans R^k la distance L_p est définie par :

$$L_P = \left(\sum_{i=1}^k |x_i - y_i|^P \right)^{1/P}$$

Où $P = 1, 2$ ou ∞

Le choix de la mesure de similarité la plus appropriée dépend du niveau d'abstraction de la représentation de l'image : images brutes (pixels) ou attributs visuels.

- A. Images brutes (pixels).
- B. Attributs visuels [2].

II.5.1. Les méthodes de calcul

II.5.1.1. Distance de Mahalanobis [1]

Cette distance prend en compte la corrélation entre les distributions des classes. Elle est ainsi définie par : $D_M = \sqrt{(f_1 - f_2)^T C^{-1} (f_1 - f_2)}$

Où C est la matrice de covariance. Dans les cas où les dimensions des caractéristiques sont indépendantes, C ne comporte que des variances et la distance de Mahalanobis se simplifie sous la forme :

$$D_M = \frac{\sum (f_1(i) - f_2(i))^2}{c_i}$$

Si C est la matrice identité, D_M est la distance euclidienne.

II.5.1.2. Intersection d'histogrammes [1] [2]

Cette mesure est l'une des premières distances utilisées dans la recherche d'image par le contenu. Elle a été proposée par Swain et Ballard mesurant la partie commune entre deux histogrammes. Etant donné deux histogrammes h_1 et h_2 .

$$D_{Inter\ sec} = \frac{\sum_i \min(h_1(i), h_2(i))}{\sum_i h_2(i)}$$

Deux images présentant une intersection normalisée d'histogrammes proche de I sont considérées comme similaires. Cette mesure n'est pas une métrique parce que non symétrique. Cependant il en existe des versions symétriques telles que celle proposée par Smith.

II.5.1.3. EarthMover Distance [1]

EMD consiste à minimiser le coût de transformation d'une distribution en une autre sous certaines contraintes de déplacement des classes de descripteurs. EMD requiert une optimisation linéaire.

$$D_{EMD} = \frac{\sum_{ij} g_{ij} d_{ij}}{\sum_{ij} g_{ij}}$$

Où d_{ij} représente la distance de similarité entre deux indices (i, j) et g_{ij} est le flot optimal entre deux distributions dont le coût total est :

Le coût est minimisé sous les contraintes suivantes :

$$D_{EMD} \sum_{ij} g_{ij} d_{ij}$$

$$g_{ij} \geq 0, \forall i, j$$

$$\sum_i g_{ij} \leq f_2(j), \forall j$$

$$\sum_j g_{ij} \leq f_1(i), \forall i$$

$$\sum_i \sum_j g_{ij} = \min(f_1(i), f_2(j))$$

II.5.1.4. Distance de Minkowski [1]

La distance de Minkowski est une famille de distances vectorielles. Soit f_1, f_2 deux vecteurs de caractéristiques, elle s'exprime par :

$$d^p(f_1, f_2) = \left(\sum_{i=1}^n |f_1(i) - f_2(i)|^p \right)^{1/p}$$

p est le facteur de Minkowski et n la dimension de l'espace caractéristique. La distance Euclidienne est un cas particulier de cette distance où $p=2$, de même que la distance de Manhattan ($p=1$).

II.5.1.5. Distance quadratique [2]

La distance de Minkowski traite les éléments du vecteur de caractéristique d'une manière équitable.

La distance quadratique en revanche favorise les éléments les plus ressemblants. Sa forme générale est donnée par : $D_Q = \sqrt{(f_1 - f_2)^T A (f_1 - f_2)}$ où $A = [a_{ij}]$ est la matrice de similarité. a_{ij} Représente la distance entre deux éléments des vecteurs f_1 et f_2 . Hafner et al propose la formule suivante pour construire la matrice A .

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

Les propriétés de cette distance la rendraient proche de la perception humaine de la couleur, ce qui en fait une métrique attractive pour les systèmes de Recherche d'images couleur par le contenu.

II.5.1.6. Distance de Bhattacharya [1]

La distance de Bhattacharya exploite la séparabilité entre deux distributions gaussiennes représentées par leur covariance Σ :

$$D_B = \frac{1}{8}(\mu_1 - \mu_2)^T \Sigma^{-1} (\mu_1 - \mu_2) + \frac{1}{2} \ln \frac{\det(\Sigma)}{\sqrt{\det(\Sigma_1) \det(\Sigma_2)}}$$

Où $\Sigma = 0.5 \times (\Sigma_1 + \Sigma_2)$ La séparabilité entre classes est estimée par la distance des moyennes et des matrices de covariance de chaque classe.

II.5.1.7. Distance de KullbackLeiber (KL) [2]

La divergence de KL exprime l'entropie relative de deux distributions :

$$D_{KL} = \sum_i f_1(i) \log \frac{f_1(i)}{f_2(i)}$$

II.5.1.8. Divergence de Jeffrey (JD) [1]

La DJ est définie par : $D_{JD} = \sum_i f_1(i) \log \frac{f_1(i)}{\hat{f}_i} + f_2(i) \log \frac{f_2(i)}{\hat{f}_i}$ Où

$\hat{f}_i = (f_1(i) + f_2(i))/2$ la différence de la mesure KL. JD est symétrique et plus stable.

II.5.1.9. Distance de Kolmogorov Smirnov [2]

Cette distance est appliquée aux distributions cumulées $f^c(i)$:

$$D_{KS} = \max_i |f_1^c(i) - f_2^c(i)|$$

II.5.1.10. Distance de Cramer Von Mises [1]

La distance de Cramer Von Mises s'applique également sur des distributions cumulées, elle est définie par :

$$D_{CVM} = \sum_i (f_1^c(i) - f_2^c(i))^2$$

II.6. Domaines D'application de CBIR

La technologie de CBIR a été employée dans plusieurs applications telles que l'identification d'empreinte digitale, systèmes d'information de biodiversité, bibliothèques numériques, empêchement de crime, la médecine, recherche historique, et d'autres. [19]

II.7. Méthodes d'évaluation d'exécution de CBIR (Précision, Rappel)

Il est important de savoir le degré d'exactitude de récupération réalisé par un système son exécution. Si l'exactitude est satisfaisante et les résultats de la technique sont prometteurs alors la technique peuvent être employés comme repère. Dans CBIR, les métriques statistiques précision se rappellent à la note ou valident l'exactitude de recherche. [12]

II.7.1. Le rappel (en anglais : 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. [16]

$$Rappel = \frac{|R_a|}{|R|}$$

II.7.2. La Précision (en anglais : Precision)

La précision est le rapport entre le nombre d'images pertinentes dans l'ensemble des images trouvées et le nombre d'images trouvées. [16]

$$Précision = \frac{|R_a|}{|A|}$$

Où :

I : une image requête

R : l'ensemble d'images pertinentes dans la base d'images utilisée pour évaluer.

$|R|$: le nombre d'images pertinentes dans la base d'images.

A : l'ensemble des réponses.

$|A|$: le nombre d'images dans l'ensemble des réponses.

$|Ra|$: le nombre d'images pertinentes dans l'ensemble des réponses.

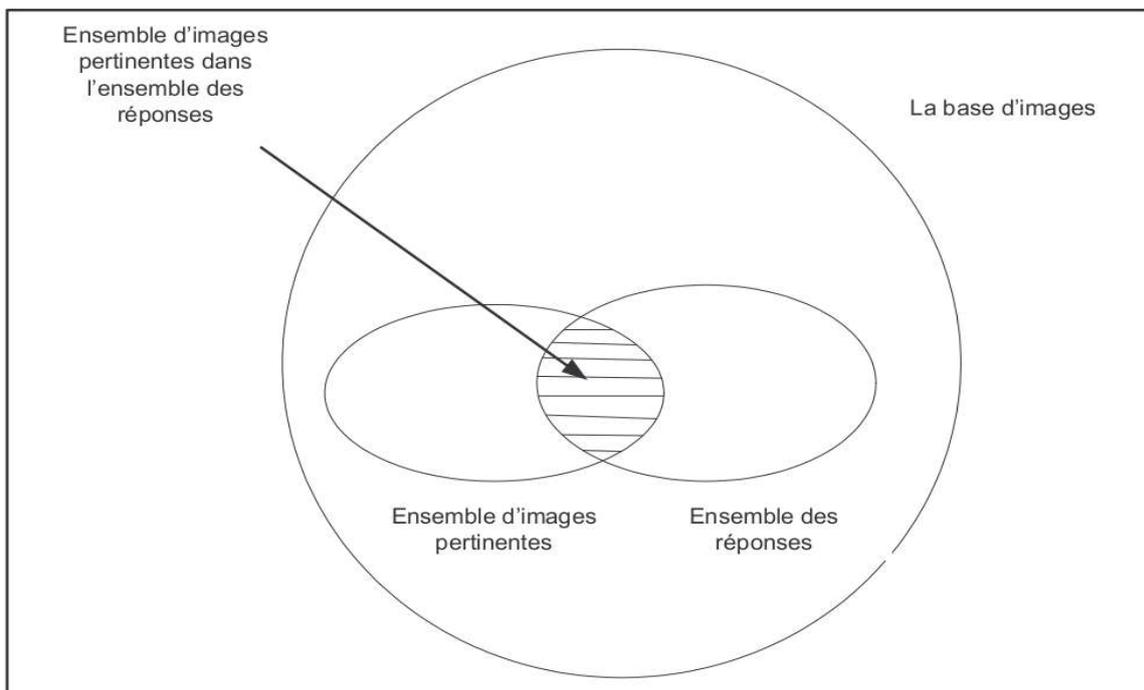


Figure 21. Le rappel et la précision pour une requête (Yates, 1999)

Le rappel et la précision sont très utiles parce qu'ils nous permettent d'évaluer quantitativement la qualité de la réponse globale et la largeur de l'algorithme de recherche. Mais ils ont 4 désavantages :

- ❖ L'estimation de la valeur maximum de rappel exige de savoir toutes les connaissances de la base d'images. Quand la base d'images devient de plus en plus grande, ces connaissances ne sont pas disponibles, ce qui veut dire que l'évaluation n'est pas bien estimée.

- ❖ Le rappel et la précision sont reliés. Donc dans quelques cas, les deux mesures ne sont pas suffisantes. Dans ces cas, l'utilisation d'autres mesures qui combinent le rappel et précision pourrait être plus appropriée.
- ❖ Ces mesures travaillent bien sur un ensemble de requêtes par lots (non-interactif). Cependant, les systèmes modernes ne travaillent pas dans ce mode. Ils permettent de communiquer avec l'utilisateur donc dans ces systèmes d'autres mesures qui sont plus appropriées doivent être utilisées.
- ❖ Le rappel et la précision sont faciles à définir quand l'ordre des images est linéaire. Ces mesures ne sont pas appropriées pour les systèmes qui ont un ordre faible. [16]

II.7.3. La courbe de rappel et précision

La précision est une fonction décroissante du rappel. Précision et rappel sont dans l'intervalle [0..1].

Le rappel et la précision sont les mesures importantes, mais si on voit seulement une paire de valeurs de rappel et précision, cette paire de valeurs ne peut pas indiquer la performance du système. C'est pourquoi on donne souvent une distribution de rappel et précision sous en forme de courbe. Pour dessiner cette courbe, on doit calculer plusieurs paires de rappel et précision et les interpoler. [16]

Les travaux de recherche dans le domaine des moteurs de recherche d'images se scindent en trois axes : [8]

- Recherche d'image basée-texte (1970-1990)

Les images sont d'abord annotées avec du texte puis recherchées en utilisant une approche basée-texte des systèmes de gestion de base de données. La difficulté réside dans le temps énorme pour annoter manuellement les images qui est subjective et ne peut être valable pour des données qui deviennent plus en plus massive (Big Data)

- Recherche d'image basée-contenu (1990-2000)

L'extraction de l'information visuelle, les requêtes utilisateurs et l'interaction, et les techniques d'indexation ont été développés. Ce qui a limité cette approche c'est le gap sémantique entre les caractéristiques de bas niveau et les concepts de haut niveau.

- Recherche d'image basée-web (2000 – Jusqu'à présent)

Les chercheurs se sont intéressés par les images web à grande échelle pour des problèmes spécifiques :

- Comment peut-on générer automatiquement des mots clés pour annoter les images ?
- Comment construire un index efficace pour la scalabilité d'un système de recherche ?
- Comment collecter le feedback utilisateur pour améliorer les performances d'un système de recherche ?
- Comment donner plus de modalités pour les requêtes de l'utilisateur pour mieux exprimer son besoin ?

Le volume et la variété des données (dimensions du big data) disponibles sont des facteurs qui alimentent la réussite des moteurs de recherche d'images.

II.8. Big Data [9]

Le big data, littéralement « grosses données », ou méga données (recommandé), parfois appelées données massives, désignent des ensembles de données qui deviennent tellement volumineux qu'ils en deviennent difficiles à travailler avec des outils classiques de gestion de base de données ou de gestion de l'information.

Divers experts, grandes institutions (comme le MIT aux États-Unis), administrations et spécialistes sur le terrain des technologies ou des usages considèrent le phénomène big data comme l'un des grands défis informatiques de la décennie 2010-2020 et en ont fait une de leurs nouvelles priorités de recherche et développement.

II.8.1. Dimensions

Le big data s'accompagne du développement d'applications à visée analytique, qui traitent les données pour en tirer du sens. Ces analyses sont appelées BigAnalytics ou « broyage de données ». Elles portent sur des données quantitatives complexes à l'aide de méthodes de calcul distribué et de statistiques.

En 2001, un rapport de recherche du META Group (devenu Gartner) définit les enjeux inhérents à la croissance des données comme étant tridimensionnels : les analyses complexes

répondent en effet à la règle dite « des 3V » (volume, vitesse et variété). Ce modèle est encore largement utilisé aujourd'hui pour décrire ce phénomène.

Le taux de croissance annuel moyen mondial du marché de la technologie et des services du Big Data sur la période 2011-2016 devrait être de 31,7%. Ce marché devrait ainsi atteindre 23,8 milliards de dollars en 2016 (d'après IDC mars 2013). Le Big Data devrait également représenter 8% du PIB européen en 2020 (AFDEL février 2013).

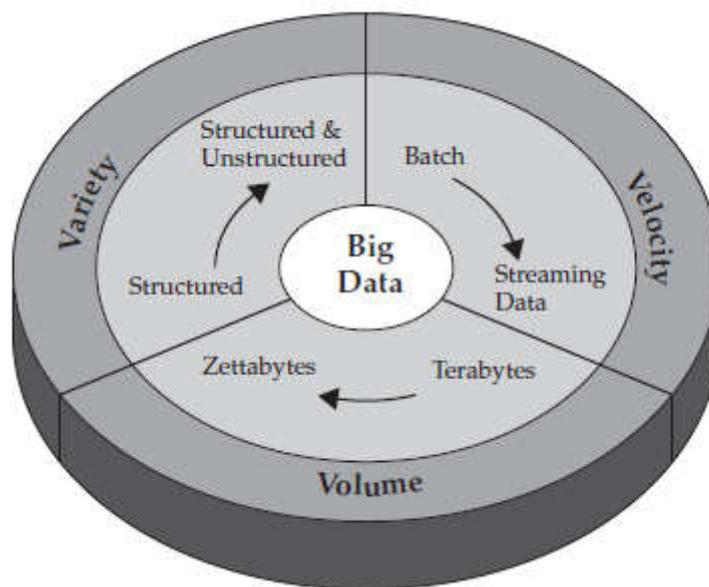


Figure 22. Caractéristique du Big Data (Les 3 V)

II.8.1.1. Volume

C'est une dimension relative : le Big Data comme le notait Lev Manovitch en 2011 définissait autrefois « les ensembles de données suffisamment grands pour nécessiter des super-ordinateurs », mais il est rapidement (dans les années 1990/2000) devenu possible d'utiliser des logiciels standards sur des ordinateurs de bureau pour analyser ou co-analyser de vastes ensembles de données.

Le volume des données stockées est en pleine expansion : les données numériques créées dans le monde seraient passées de 1,2 zettaoctets par an en 2010 à 1,8 zettaoctets en 2011, puis 2,8 zettaoctets en 2012 et s'élèveront à 40 zettaoctets en 2020. À titre d'exemple, Twitter génère en janvier 2013, 7 téraoctets (TB) de données chaque jour et Facebook 10 téraoctets (TB).

Ce sont pourtant les installations technico-scientifiques (météorologie par exemple) qui produiraient le plus de données. De nombreux projets, de dimension pharaonique, sont ainsi en cours. Le radiotélescope "Square Kilometre Array" par exemple, produira 50 téraoctets de données analysées par jour, à un rythme de 7 000 téraoctets de données brutes par seconde

II.8.1.2. Variété

Le volume des Big Data met les data centers devant un réel défi : la variété des données. Il ne s'agit pas de données relationnelles traditionnelles, ces données sont brutes, semi-structurées voire non structurées (cependant, les données non-structurées devront, pour utilisation, être structurées). Ce sont des données complexes provenant du web (Web Mining), au format texte (Text Mining) et images (Image Mining). Elles peuvent être publiques (Open Data, Web des données), géo-démographiques par îlot (adresses IP), ou relever de la propriété des consommateurs (Profils 360°). Ce qui les rend difficilement utilisables avec les outils traditionnels.

La démultiplication des outils de collecte sur les individus et sur les objets permettent d'amasser toujours plus de données. Et les analyses sont d'autant plus complexes qu'elles portent de plus en plus sur les liens entre des données de natures différentes.

II.8.1.3. Vitesse

La vitesse représente à la fois la fréquence à laquelle les données sont générées, capturées, partagées et mises à jour.

Des flux croissants de données doivent être analysés en quasi-temps réel (fouille de flots de données) pour répondre aux besoins des processus chrono-sensibles. Par exemple, les systèmes mis en place par la bourse et les entreprises doivent être capables de traiter ces données avant qu'un nouveau cycle de génération n'ait commencé, avec le risque pour l'Homme de perdre une grande partie de la maîtrise du système quand les principaux opérateurs deviennent des "robots" capables de lancer des ordres d'achat ou de vente de l'ordre de la nanoseconde (Trading haute fréquence), sans disposer de tous les critères pertinents d'analyse pour le moyen et long terme.

II.8.1.4. Différence avec l'informatique décisionnelle

Si la définition du Gartner en 3V est encore largement reprise (voire augmentée de “V” supplémentaires selon l’inspiration des services marketing), la maturation du sujet fait apparaître un autre critère plus fondamental de différence avec l'informatique décisionnelle et concernant les données et leur utilisation :

- Informatique décisionnelle : utilisation de statistique descriptive, sur des données à forte densité en information afin de mesurer des phénomènes, détecter des tendances... ;
- Big Data : utilisation de statistique inférentielle, sur des données à faible densité en information dont le grand volume permet d’inférer des lois (régressions....) donnant dès lors (avec les limites de l’inférence) au big data des capacités prédictives.

Synthétiquement :

- L’informatique traditionnelle, informatique décisionnelle comprise, est basée sur un modèle du monde ;
- Le big data vise à ce que les mathématiques trouvent un modèle dans les données

La solution que nous proposons est une approche mixte qui combine entre les approches basées texte celles basées contenu. L'utilisateur peut exprimer sa requête par des mots clés ou par une image exemple.

Conclusion

Dans ce chapitre, les systèmes de recherche d'images ont été abordés avec leur typologie. La notion de BIG-DATA est présentée. Une brève présentation des approches de recherche d'images (basée texte, basée contenu, et basée web) est donnée et enfin nous avons adopté une approche mixte combinant les basée texte et basée contenu pour notre système de recherche d'images. Le chapitre suivant donnera le détail de ce système sur des exemples de notre base d'images.

chapitre 3

Expérimentation et résultats

Chapitre3 : Expérimentation et résultats

Introduction

Nous présentons dans ce dernier chapitre, après avoir étudié le domaine de CBIR, les différents outils utilisés pour le développement de notre application, nous allons exposer les différentes étapes par lesquelles nous sommes passés pour la réalisation de cette application, et l'ensemble des expérimentations que nous avons menées durant cette thèse.

III.1. Les outils de développement

Pour l'implémentation de notre application de recherche d'image, nous avons utilisé le langage de programmation Matlab version R2013a avec le système d'exploitation *Windows 7*.

III.1.1. Matlab

Matlab est un logiciel de calcul numérique commercialisé par la société *MathWorks*. Il a été initialement développé à la fin des années 70 par *Cleve Moler*, professeur de mathématique à l'université du Nouveau-Mexique puis à Stanford.

Matlab est utilisé pour le calcul scientifique, l'analyse de données, leur visualisation, l'implémentation rapide d'algorithmes. Son interface propose, d'une part, une fenêtre interactive type console pour l'exécution de commandes, et d'autre part, un environnement de développement intégré (IDE) pour la programmation d'applications. [15]

III.2. Conception

III.2.1. Les caractéristiques visuelles choisies

III.2.1.1. Les modèles de couleur utilisés

Dans notre application, nous avons intégré l'espace couleur RGB, Il est composé des trois couleurs primaires (rouge, vert et bleu) et le Système HSV (HSI), qui représente (la teinte (H), saturation (S), valeur (V)).

III.2.1.2. Préparation des indexes

Dans l'étape de préparation des indexes nous avons choisi les descripteurs qui sont définis dans le tableau suivant (**Tableau 2**) avec le nombre de composantes.

Tableau 2 : les dimensions des descripteurs

Descripteurs	Dimensions
Histogrammes	32
Corrélogramme	64
Descripteurs de formes	03
Energie	24
Amplitude	24
Moments ondelettes	40
Descripteurs texturaux	04
Moments statistiques	06
Dimension vecteur visuelle	197

III.2.2. Les caractéristiques textuelles choisies

Dans cette étape on a utilisé le modèle vectoriel documentaire « sac de mots » [21], pour présenter les vecteurs textuels, on a chargé la maîtrise de plusieurs attributs, chaque une représente le nombre répétitif du mot dans le texte de la requête textuelle et on a calculé les fréquences de chaque attribut de la matrice, on utilise un dictionnaire intermédiaire.

III.2.2.1. Extraction de données

Dans la classification automatique de textes, le choix de représentation des documents est une étape cruciale. Une approche fréquente consiste à faire appel à une représentation dite « sac de mots », où la seule information utilisée est la présence et/ou la fréquence de certains mots. Dans notre contexte, nous utilisons une représentation vectorielle des documents selon le modèle de Salton, de manière à pouvoir représenter les documents sous forme de vecteurs de mots pondérés.

L'extraction de données est une tâche qui s'est développée dans le domaine de Traitement Automatique des Langues (TAL). Elle consiste à identifier et extraire d'un texte les éléments pertinents contenant des informations dont la nature est spécifiée à l'avance. Elle vise donc à transformer un texte de son format initial (une suite de chaînes de caractères) à une représentation structurée et donc un format qui soit compréhensible par l'ordinateur. Elle se fait en reconnaissant dans le texte des unités lexicales particulières [12].

Dans cette étape on a décomposé le texte en unités lexicales (mots, ponctuation, mots vides), après cette tâche on a éliminé la ponctuation et les mots vides on laisse seulement les mots des noms et les verbes, pour les ajouter dans un dictionnaire intermédiaire.

III.2.2.2. Calcul de poids

Les données collectées du texte n'ont pas toutes la même valeur informative, Certaines données sont plus importantes que d'autres pour la classification car elles apparaissent de façon répétée dans le texte, d'autres le sont moins à cause de leur caractère commun dans la langue. Pour permettre une meilleure classification, il faut refléter ce degré d'importance dans l'algorithme. C'est ainsi que nous attribuons à chaque mot un poids.

IL existe plusieurs méthodes de calcul de poids, en voici deux les plus connues :

- **TF** (TermFrequency) : C'est la fréquence d'apparition d'un mot dans le texte, elle est égale au nombre d'occurrences de ce mot divisé par le nombre total de mots du texte.
- **TF – IDF** (TermFrequency, Inverse Document Frequency): C'est une mesure statistique qui permet d'évaluer l'importance d'un mot dans un texte relativement à tout un corpus. Le poids ne dépend pas seulement de la fréquence du mot dans le texte en question mais aussi de la fréquence du mot dans tout le corpus. Ainsi un mot qui se répète dans tous les documents devient inutile pour la classification.

TF_IDF = TF × IDF.

Où **IDF = $\log \left(\frac{\text{nombre totale des documents}}{\text{nombre de document contenant le mot}} \right)$.**

III.2.2.3. Préparation des indexes

Nos descripteurs textuels correspondent aux rangs de ses mots dans le dictionnaire, les valeurs de ce descripteur sont les fréquences de chaque mot dans le texte (TF_IDF).

III.2.3. Les requêtes

On a donné trois façons à l'utilisateur d'émettre sa requête :

- utiliser une requête visuelle par une image de recherche.
- utiliser une requête textuelle (une suite de mots).
- utiliser une requête combinant texte et image.

III.2.4. Mesure de similarité utilisée

Mesure de similarité utilisée est la distance euclidienne définie par :

$$D = \sqrt{\sum_{i=1}^k (x_i - y_i)^2}$$

Où :

$X = [x_1, x_2, \dots, x_k]$ et $Y = [y_1, y_2, \dots, y_k]$ étant les vecteurs à comparer.

III.2.5. Architecture de l'application

Nous avons adopté Matlab version R2013a comme langage de programmation, ce choix est justifié par plusieurs facteurs :

- Puissance du langage dans le traitement d'images (toolbox IPT).
- Puissance du langage en matière de structures de données (Vecteurs et Matrices dynamiques...).
- Facilité d'interaction grâce au mode interpréteur.

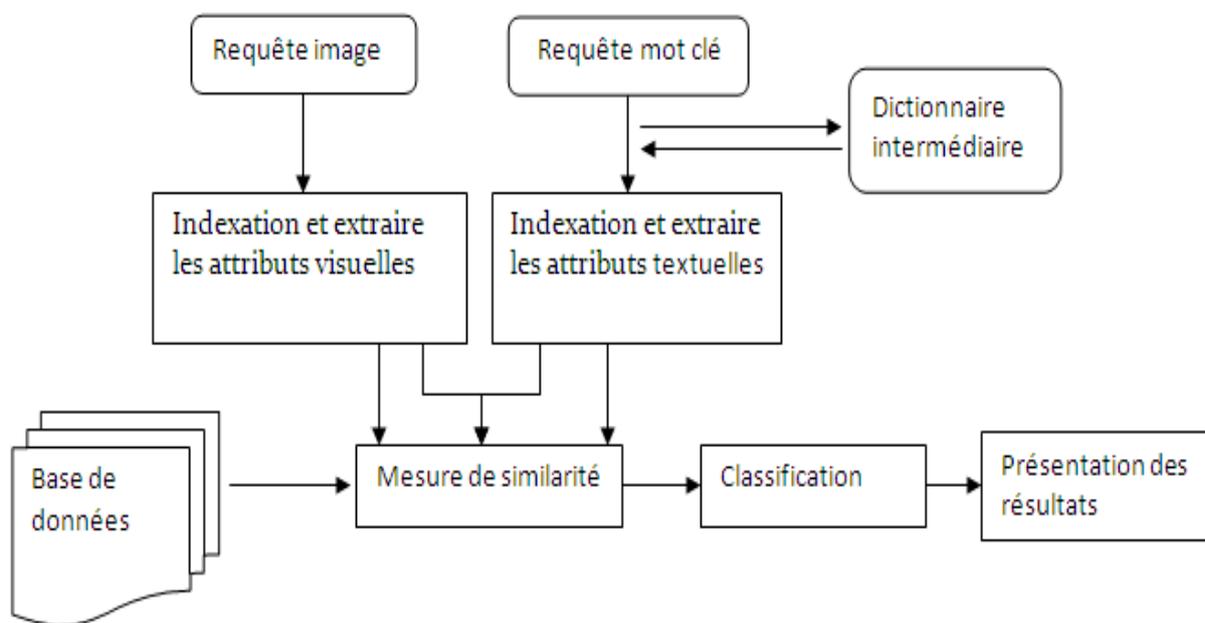


Figure 23 : Architecture de l'application

III.3. Réalisation

Notre application est composée de deux interfaces, l'interface utilisateur et l'interface de présentation des résultats.

III.3.1. L'interface utilisateur

Permet à l'utilisateur d'introduire ses requêtes (visuelle, textuelle ou les deux). De créer sa base d'images ou éventuellement d'y ajouter une nouvelle image.

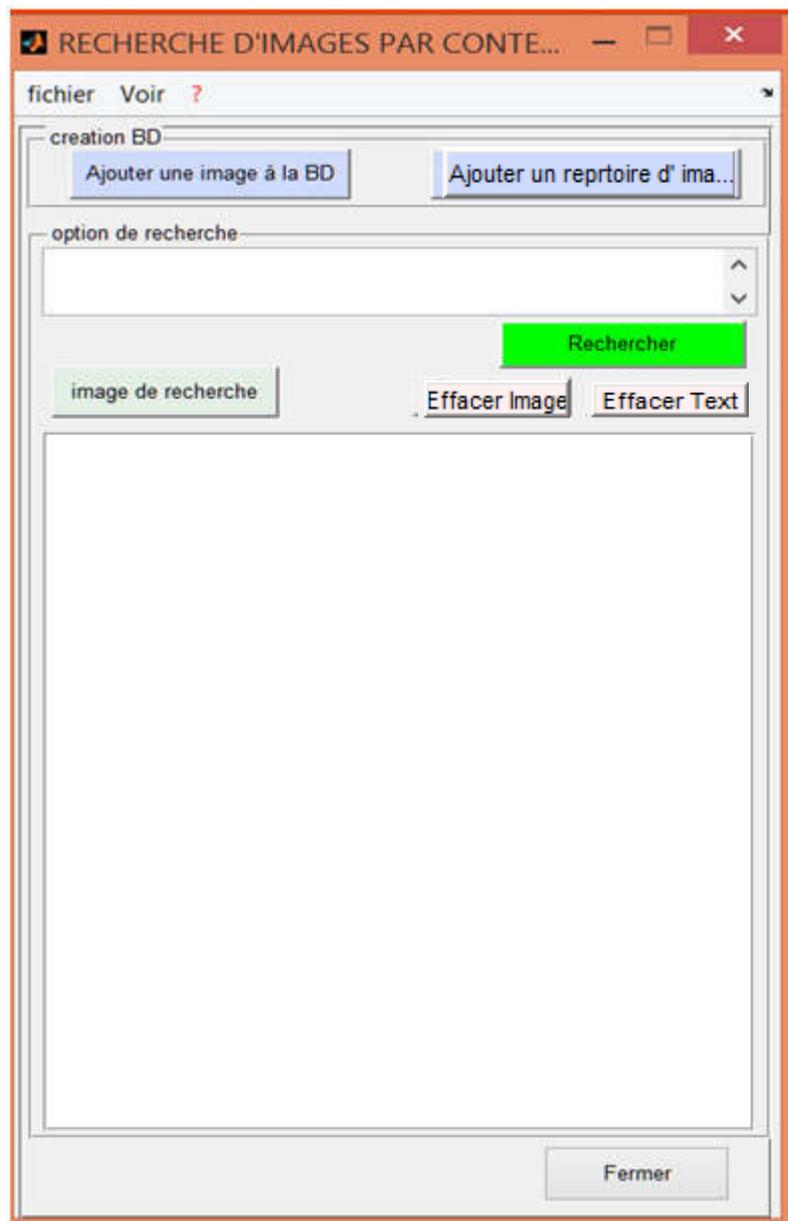


Figure 24 : Interface Utilisateur

III.3.2. La présentation des résultats

La plupart des systèmes disponibles du même genre que CBIR, la présentation de résultat d'une requête se fait sous la forme d'une liste d'images (réduites à des vignettes) ordonnées par pertinence décroissante.

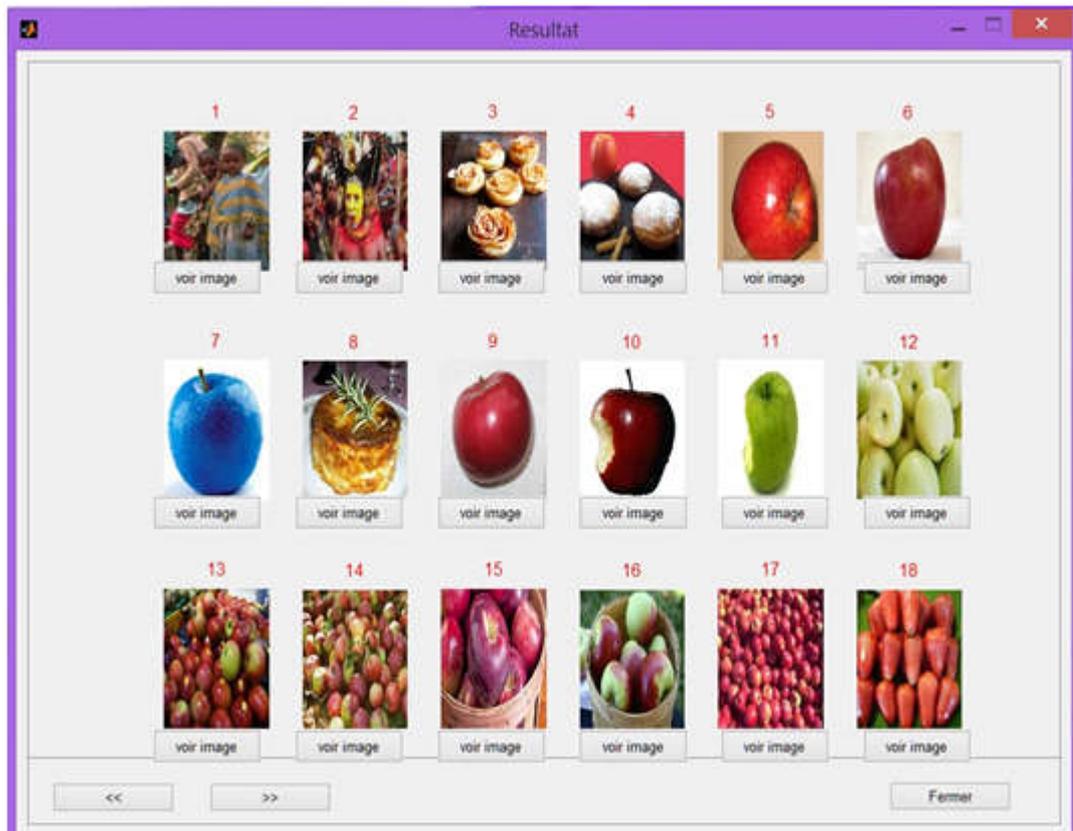


Figure 25 : Présentation des résultats

III.4. Evaluation de l'application

Nous avons choisi les trois types de requête (visuelle, textuelle, visuelle et textuelle) comme une technique de base pour la recherche des images dans notre application, nous avons testé la performance de ces techniques et dans cette section nous avons présenté les différents résultats.

III.4.1. la Base des images

La base d'images utilisée comporte 90 images couleur. Elle a été téléchargée depuis le site web [20]. La base d'images originale comporte 1000 images divisées en 10 classes. Nous avons optimisé pour l'utilisation de 90 images et 5 classes pour des raisons de temps de calcul. Chaque classe représente un thème défini : Rose, Bus, Dinosaures, Chevaux et Couches de Soleil. Un échantillon de la base est présenté en **Figure 26**.

Rose

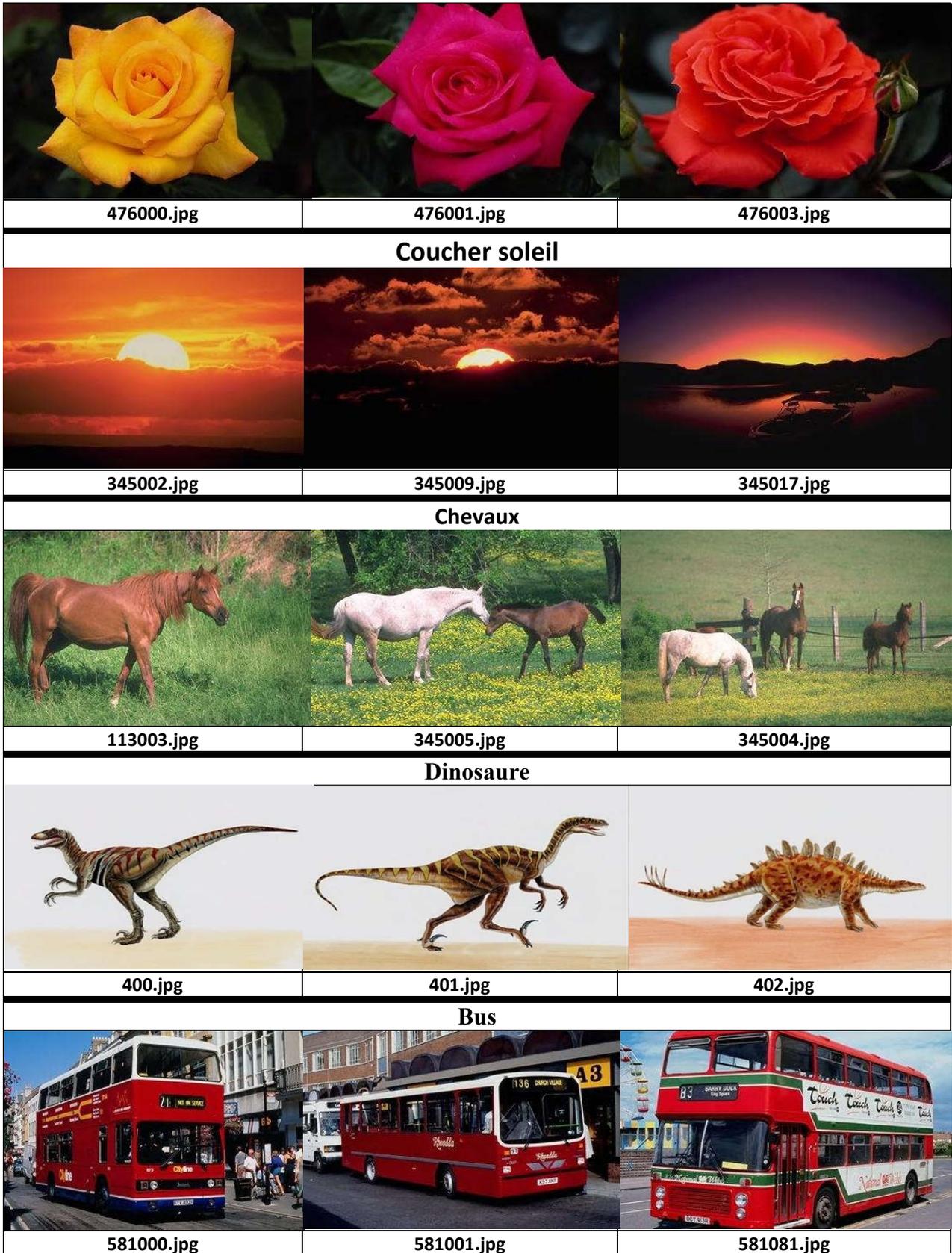


Figure 26 : Un échantillon de la base

III.4.2. Processus suivi

Dans cette partie nous avons évalué l'application avec le calcul de précision pour chaque type de requête. Le principe de fonctionnement de ce processus d'évaluation peut être d'écrite a par la figure 32.

Les images requêtent son sélectionnées de manière aléatoires (sans répétition de la même image requête). La figure 31 représente un exemple d'un ensemble d'images requête.

III.4.2.1. Requête visuelle

Remarque : il est clair que pour un système soit parfait, la valeur de précision égale à 1, dans notre cas on a 15 requêtes, dont les valeurs de précision varient entre 0 et 1, donc la précision moyenne globale sera calculée comme suite :

$$\text{Précision Moyenne Globale} = \frac{\sum Pr * 100}{15}$$

Pr : valeur de précision par requête.

Alors que la précision moyenne pour chaque classe doit être calculé par rapport à la précision moyenne globale :

$$\text{Précision Moyenne Classe} = \frac{\sum Pr * 100}{Pmg}$$

Pmg : Valeur de précision moyenne globale

Tableau 3 : Requêtevisuelle, Seuil = 50% du maximum de la distance Euclidienne.

Classe	Image requête	Total trouvé	Images pertinentes	Images non pertinentes	Précision	Rappel	Précision moyenne par classe
Rose :	476000.jpg	398	90	308	0,22	1	12,41 %
	476001.jpg	398	90	308	0,22	1	
	476003.jpg	398	90	308	0,22	1	
Coucher soleil :	345002.jpg	392	87	305	0,22	0,96	11,97 %
	345009.jpg	386	90	296	0,23	1	
	345017.jpg	391	78	313	0,19	0,86	
Chevaux :	113003.jpg	397	87	310	0,21	0,96	12,03%
	113005.jpg	400	87	313	0,21	0,96	
	113004.jpg	394	87	307	0,22	0,96	
Dinosaure :	400.jpg	409	90	319	0,22	1	11,81%
	401.jpg	406	90	316	0,22	1	
	402.jpg	402	82	320	0,20	0,91	
Bus :	581000.jpg	85	83	2	0,97	0,92	51,57%
	581001.jpg	86	83	3	0,96	0,92	
	581081.jpg	97	86	11	0,88	0,95	
Précision moyenne globale					36.42%		

Tableau 4 : Requêtevisuelle, Seuil = 25% du maximum de la distance Euclidienne.

Classe	Image requête	Total trouvé	Images pertinentes	Images non pertinentes	Précision	Rappel	Précision moyenne par classe
Rose :	476000.jpg	263	89	174	0,33	0,98	12.90 %
	476001.jpg	222	83	139	0,37	0,92	
	476003.jpg	252	88	164	0,34	0,97	
Coucher soleil :	345002.jpg	251	78	173	0,31	0,86	13.46 %
	345009.jpg	198	77	121	0,38	0,85	
	345017.jpg	191	78	113	0,40	0,86	
Chevaux :	113003.jpg	274	80	194	0,29	0,88	17.80 %
	113005.jpg	283	81	202	0,28	0,9	
	113004.jpg	53	47	6	0,88	0,52	
Dinosaure :	400.jpg	151	90	61	0,59	1	19.37 %
	401.jpg	138	90	48	0,65	1	
	402.jpg	243	84	159	0,34	0,93	
Bus :	581000.jpg	5	5	0	1	0,05	36.45 %
	581001.jpg	17	17	0	1	0,18	
	581081.jpg	23	23	0	1	0,25	
Précision moyenne globale					54.85 %		

Interprétation des résultats

D'après les résultats obtenus et illustrés par les tableaux ci-dessus, avec 25 % de la distance euclidienne maximal une précision moyenne globale des 5 classes égale à 54,85% est accomplie. Les deux meilleures précisions en été obtenu pour les classes « bus » et « dinosaures » car elle présente une hétérogénéité dans la distribution de la couleur et les Forme d'objet dans le fond.

La classe « Dinosaures » présente la plus grande précision (19.37%). Les images de cette classe ne sont pas des photographies mais des images synthétiques. Elles ont toutes un arrière-plan uni blanc. Cet arrière-plan permet de les distinguer des images des autres classes. Les cas d'erreur introduits sont des images qui contiennent une grande quantité de

couleur blanche. La classe « Bus » présentent une bonne précision (36.45%) pour la même raison.

III.4.2.2. Requête textuelle

Tableau 5 : Requête textuelle, Seuil = 94 % du maximum de la distance Euclidienne.

Classe	Image requête	Total trouvé	Images pertinentes	Images non pertinentes	Précision	Rappel	Précision moyenne par classe
Rose :	Roserouge	50	6	44	0,12	8,33	16.12 %
	Rose jaune	50	2	48	0,04	25	
	Rose blanc	50	2	48	0,04	25	
Coucher soleil :	Coucher soleil Algérie	50	6	44	0,12	8,33	16.12 %
	Coucher soleil Maroc	50	3	47	0,06	16,66	
	Coucher soleil Tunis	50	1	49	0,02	50	
Chevaux :	Chevaux marron	50	10	40	0,2	5	25.80 %
	Chevaux marron et gris	50	3	47	0,06	16,66	
	Chevaux marron blanc	50	3	47	0,06	16,66	
Dinosaure :	Dinosaure marron	50	4	46	0,08	12,5	14.51 %
	Dinosaure vert	50	3	47	0,06	16,66	
	Dinosaure gris	50	2	48	0,04	25	
Bus :	Bus rouge	50	7	43	0,14	7,14	27.41 %
	Bus rouge et blanc	50	8	42	0,16	6,25	
	Bus bleu	50	2	48	0,04	25	
Précision moyenne globale					8.26%		

Tableau 6 : tableau d'évaluation par requête textuelle, Seuil = 0 % du maximum de la distance.

Classe	Image requête	Total trouvé	Images pertinentes	Images non pertinentes	Précision	Rappel	Précision moyenne par classe
Rose :	Rose rouge	6	6	0	1	1	20 %
	Rose jaune	2	2	0	1	1	
	Rose blanc	1	1	0	1	1	
Coucher soleil :	Coucher soleil Algérie	5	5	0	1	1	20 %
	Coucher soleil Maroc	3	3	0	1	1	
	Coucher soleil Tunis	1	1	0	1	1	
Chevaux :	Chevaux marron	4	4	0	1	1	20 %
	Chevaux marron et gris	3	3	0	1	1	
	Chevaux marron blanc	3	3	0	1	1	
Dinosaure :	Dinosaure marron	4	4	0	1	1	20 %
	Dinosaure vert	1	1	0	1	1	
	Dinosaure gris	1	1	0	1	1	
Bus :	Bus rouge	4	4	0	1	1	20 %
	Bus rouge et blanc	3	3	0	1	1	
	Bus bleu	1	1	0	1	1	
Précision moyenne globale					100 %		

Interprétation des résultats

D'après les résultats obtenus et illustrés par les tableaux ci-dessus, la Précision moyenne globale est meilleure égale 100% lorsque la valeur de la distance maximum égale à 0%, ce qui donne une bonne représentation de l'information, mais les systèmes classiques de recherche d'information possèdent des problèmes persistants, par exemple la requête de recherche est souvent exprimée par une liste de termes souvent très réduite qui exprime souvent mal les besoins en information de l'utilisateur et pour une même requête le système retourne les mêmes résultats, alors que des utilisateurs différents peuvent formuler la même requête pour exprimer des besoins différents.

III.4.2.3. Requête visuelle et textuelle

Tableau 7 : tableau d'évaluation par requête visuelle et textuelle, Seuil = 94 % du maximum de la distance.

Classe	Image requête	Total trouvé	Images pertinentes	Images non pertinentes	Précision	Rappel	Précision moyenne par classe
Rose	476000.jpg+rose rouge	17	10	7	0,58	0,18	14.09 %
	476001.jpg +rose rouge	17	10	7	0,58	0,18	
	476003.jpg +rose jaune	396	90	306	0,22	4,4	
Coucher soleil	345002.jpg+ coucher soleil Algérie	10	10	0	1	0,11	22.07 %
	345009.jpg +coucher soleil Algérie	10	10	0	1	0,11	
	345017.jpg +coucher soleil Maroc	453	90	363	0,19	5,03	
Chevaux	113003.jpg+chevaux marron	14	10	4	0,71	0,15	20.30 %
	113005.jpg +chevaux marron	14	10	4	0,71	0,15	
	113004.jpg +chevaux gris	5	3	2	0,6	0,05	
Dinosaure	400.jpg+dinosaure marron	14	10	4	0,71	0,15	17.69 %
	401.jpg +dinosaure marron	14	10	4	0,71	0,15	
	402.jpg +dinosaure gris	3	1	2	0,33	0,03	
Bus	581000.jpg+bus rouge	9	9	0	1	0,1	25.77 %
	581001.jpg +bus rouge	10	9	1	0,9	0,11	
	581081.jpg +bus bleu	3	2	1	0,66	0,03	
Précision moyenne globale					66.39 %		

Tableau 8 : tableau d'évaluation par requête visuelle et textuelle, Seuil = 92 % du maximum de la distance.

Classe	Image requête	Total trouvé	Images pertinentes	Images non pertinentes	Précision	Rappel	Précision moyenne par classe
Rose :	476000.jpg+rose rouge	10	10	0	1	0,11	20 %
	476001.jpg +rose rouge	10	10	0	1	0,11	
	476003.jpg +rose jaune	2	2	0	1	0,02	
Coucher soleil :	345002.jpg+ coucher soleil Algérie	9	9	0	1	0,1	20 %
	345009.jpg +coucher soleil Algérie	10	10	0	1	0,11	
	345017.jpg +coucher soleil Maroc	10	10	0	1	0,11	
Chevaux :	113003.jpg+chevaux marron	10	10	0	1	0,11	20 %
	113005.jpg +chevaux marron	10	10	0	1	0,11	
	113004.jpg +chevaux gris	3	3	0	1	0,03	
Dinosaure :	400.jpg+dinosaure marron	4	4	0	1	0,04	20 %
	401.jpg +dinosaure marron	4	4	0	1	0,04	
	402.jpg +dinosaure gris	2	2	0	1	0,02	
Bus :	581000.jpg+bus rouge	5	5	0	1	0,05	20 %
	581001.jpg +bus rouge	6	6	0	1	0,06	
	581081.jpg +bus bleu	1	1	0	1	0,01	
Précision moyenne globale					100%		

Interprétation des résultats

D'après les résultats obtenus et illustrés par les tableaux ci-dessus, la Précision moyenne globale est meilleur égale 100% lorsque la valeur de la distance maximum égale à 92%, mais le problème reste dans la formulation du requête utilisateur ou la combinaison entre les deux

requêtes, comment choisir une requête d'image et sa requête textuelle, pour obtenir de bon résultat souhaiter.

III.4.3. Problèmes rencontrés

La première problématique qui s'est imposée durant la réalisation de notre application est le choix des descripteurs discriminants d'image pour s'assurer de l'obtention des meilleurs résultats et pour une évaluation significative.

- L'annotation manuelle des images de la base de données (La saisie de chaînes de caractères) exige beaucoup d'effort (description pertinente de la scène), et nécessite trop de temps.
- Aussi le choix du modèle de couleur et la mesure de distance à utiliser.

Conclusion

Nous avons présenté dans ce chapitre les caractéristiques textuelles et les caractéristiques visuelles choisis (modèle de couleur, préparation de l'index, la mesure de similarité utilisé et les mesures de performance étudiées). Après, on a présenté les grandes parties de notre application et les différents résultats obtenus pendant l'évaluation de notre application de recherche d'image par le contenu.

Chapitre3 : Expérimentation et résultats

Introduction

Nous présentons dans ce dernier chapitre, après avoir étudié le domaine de CBIR, les différents outils utilisés pour le développement de notre application, nous allons exposer les différentes étapes par lesquelles nous sommes passés pour la réalisation de cette application, et l'ensemble des expérimentations que nous avons menées durant cette thèse.

III.1. Les outils de développement

Pour l'implémentation de notre application de recherche d'image, nous avons utilisé le langage de programmation Matlab version R2013a avec le système d'exploitation *Windows 7*.

III.1.1. Matlab

Matlab est un logiciel de calcul numérique commercialisé par la société *MathWorks*. Il a été initialement développé à la fin des années 70 par *Cleve Moler*, professeur de mathématique à l'université du Nouveau-Mexique puis à Stanford.

Matlab est utilisé pour le calcul scientifique, l'analyse de données, leur visualisation, l'implémentation rapide d'algorithmes. Son interface propose, d'une part, une fenêtre interactive type console pour l'exécution de commandes, et d'autre part, un environnement de développement intégré (IDE) pour la programmation d'applications. [15]

III.2. Conception

III.2.1. Les caractéristiques visuelles choisies

III.2.1.1. Les modèles de couleur utilisés

Dans notre application, nous avons intégré l'espace couleur RGB, Il est composé des trois couleurs primaires (rouge, vert et bleu) et le Système HSV (HSI), qui représente (la teinte (H), saturation (S), valeur (V)).

III.2.1.2. Préparation des indexes

Dans l'étape de préparation des indexes nous avons choisi les descripteurs qui sont définis dans le tableau suivant (**Tableau 2**) avec le nombre de composantes.

Tableau 2 : les dimensions des descripteurs

Descripteurs	Dimensions
Histogrammes	32
Corrélogramme	64
Descripteurs de formes	03
Energie	24
Amplitude	24
Moments ondelettes	40
Descripteurs texturaux	04
Moments statistiques	06
Dimension vecteur visuelle	197

III.2.2. Les caractéristiques textuelles choisies

Dans cette étape on a utilisé le modèle vectoriel documentaire « sac de mots » [21], pour présenter les vecteurs textuels, on a chargé la maîtrise de plusieurs attributs, chaque une représente le nombre répétitif du mot dans le texte de la requête textuelle et on a calculé les fréquences de chaque attribut de la matrice, on utilise un dictionnaire intermédiaire.

III.2.2.1. Extraction de données

Dans la classification automatique de textes, le choix de représentation des documents est une étape cruciale. Une approche fréquente consiste à faire appel à une représentation dite « sac de mots », où la seule information utilisée est la présence et/ou la fréquence de certains mots. Dans notre contexte, nous utilisons une représentation vectorielle des documents selon le modèle de Salton, de manière à pouvoir représenter les documents sous forme de vecteurs de mots pondérés.

L'extraction de données est une tâche qui s'est développée dans le domaine de Traitement Automatique des Langues (TAL). Elle consiste à identifier et extraire d'un texte les éléments pertinents contenant des informations dont la nature est spécifiée à l'avance. Elle vise donc à transformer un texte de son format initial (une suite de chaînes de caractères) à une représentation structurée et donc un format qui soit compréhensible par l'ordinateur. Elle se fait en reconnaissant dans le texte des unités lexicales particulières [12].

Dans cette étape on a décomposé le texte en unités lexicales (mots, ponctuation, mots vides), après cette tâche on a éliminé la ponctuation et les mots vides on laisse seulement les mots des noms et les verbes, pour les ajouter dans un dictionnaire intermédiaire.

III.2.2.2. Calcul de poids

Les données collectées du texte n'ont pas toutes la même valeur informative, Certaines données sont plus importantes que d'autres pour la classification car elles apparaissent de façon répétée dans le texte, d'autres le sont moins à cause de leur caractère commun dans la langue. Pour permettre une meilleure classification, il faut refléter ce degré d'importance dans l'algorithme. C'est ainsi que nous attribuons à chaque mot un poids.

IL existe plusieurs méthodes de calcul de poids, en voici deux les plus connues :

- **TF** (TermFrequency) : C'est la fréquence d'apparition d'un mot dans le texte, elle est égale au nombre d'occurrences de ce mot divisé par le nombre total de mots du texte.
- **TF – IDF** (TermFrequency, Inverse Document Frequency): C'est une mesure statistique qui permet d'évaluer l'importance d'un mot dans un texte relativement à tout un corpus. Le poids ne dépend pas seulement de la fréquence du mot dans le texte en question mais aussi de la fréquence du mot dans tout le corpus. Ainsi un mot qui se répète dans tous les documents devient inutile pour la classification.

$$\mathbf{TF_IDF} = \mathbf{TF} \times \mathbf{IDF}.$$

$$\text{Où } \mathbf{IDF} = \log \left(\frac{\text{nombre totale des documents}}{\text{nombre de document contenant le mot}} \right).$$

III.2.2.3. Préparation des indexes

Nos descripteurs textuels correspondent aux rangs de ses mots dans le dictionnaire, les valeurs de ce descripteur sont les fréquences de chaque mot dans le texte (TF_IDF).

III.2.3. Les requêtes

On a donné trois façons à l'utilisateur d'émettre une requête :

- utiliser une requête visuelle par une image de recherche.
- utiliser une requête textuelle (une suite de mots).
- utiliser une requête combinant texte et image.

III.2.4. Mesure de similarité utilisée

Mesure de similarité utilisée est la distance euclidienne définie par :

$$D = \sqrt{\sum_{i=1}^k (x_i - y_i)^2}$$

Où :

$X = [x_1, x_2, \dots, x_k]$ et $Y = [y_1, y_2, \dots, y_k]$ étant les vecteurs à comparer.

III.2.5. Architecture de l'application

Nous avons adopté Matlab version R2013a comme langage de programmation, ce choix est justifié par plusieurs facteurs :

- Puissance du langage dans le traitement d'images (toolbox IPT).
- Puissance du langage en matière de structures de données (Vecteurs et Matrices dynamiques...).
- Facilité d'interaction grâce au mode interpréteur.

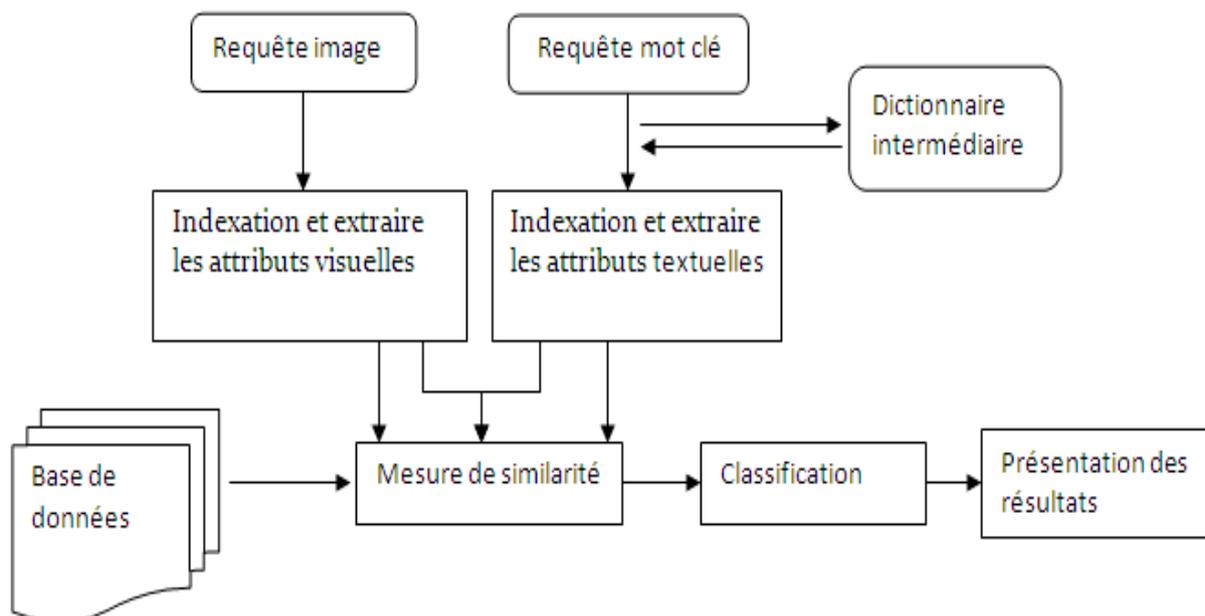


Figure 23 : Architecture de l'application

III.3. Réalisation

Notre application est composée de deux interfaces, l'interface utilisateur et l'interface de présentation des résultats.

III.3.1. L'interface utilisateur

Permet à l'utilisateur d'introduire ses requêtes (visuelle, textuelle ou les deux). De créer sa base d'images ou éventuellement d'y ajouter une nouvelle image.

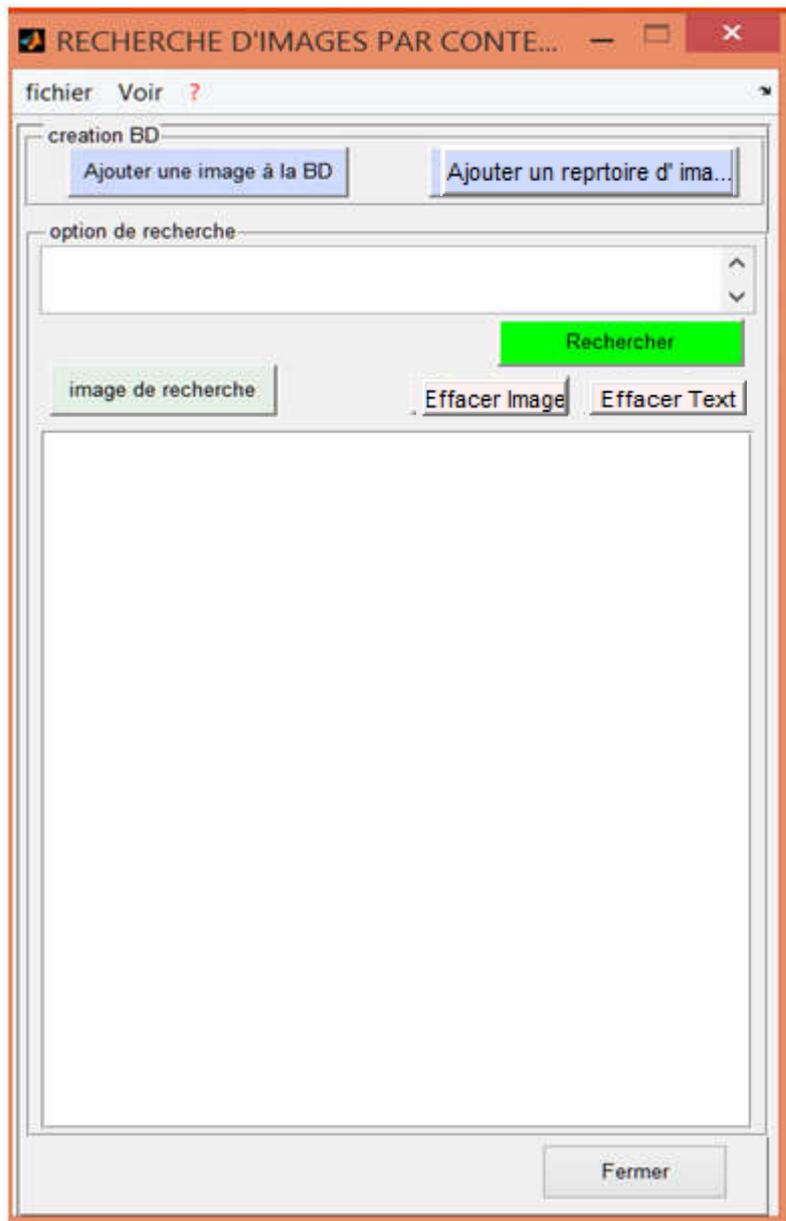


Figure 24 : Interface Utilisateur

III.3.2. La présentation des résultats

La plupart des systèmes disponibles du même genre que CBIR, la présentation de résultat d'une requête se fait sous la forme d'une liste d'images (réduites à des vignettes) ordonnées par pertinence décroissante.

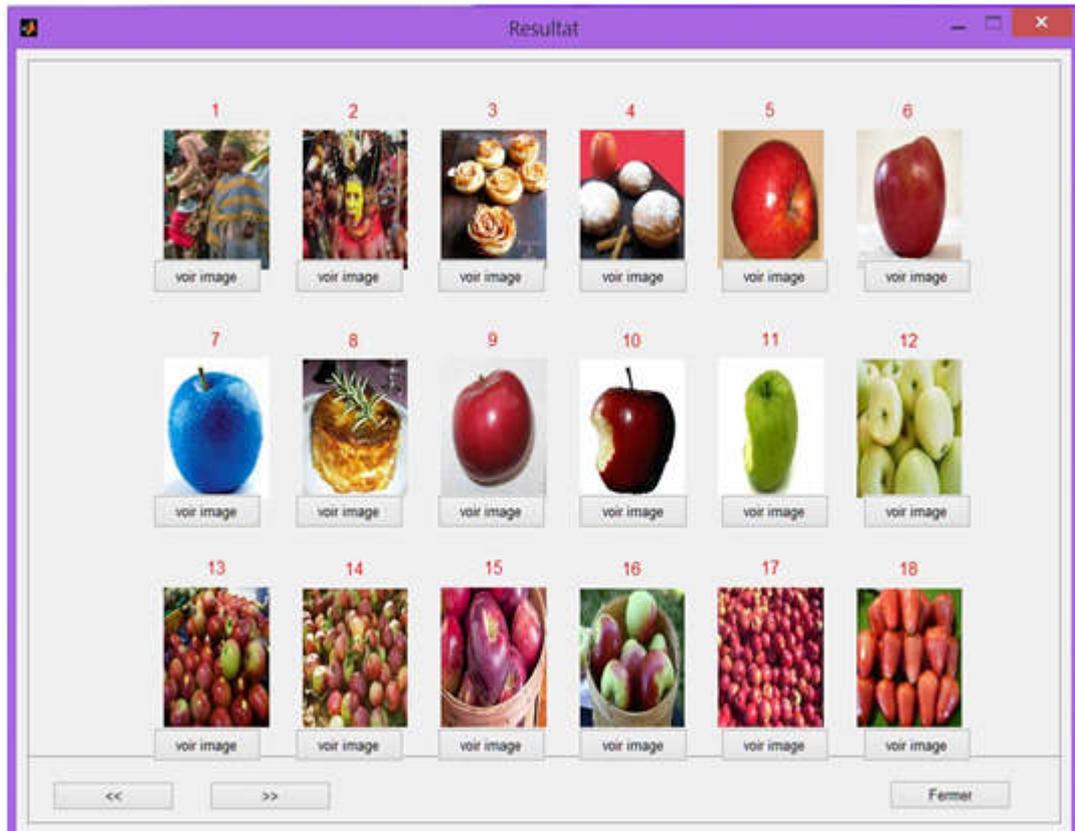


Figure 25 : Présentation des résultats

III.4. Evaluation de l'application

Nous avons choisi les trois types de requête (visuelle, textuelle, visuelle et textuelle) comme une technique de base pour la recherche des images dans notre application, nous avons testé la performance de ces techniques et dans cette section nous avons présenté les différents résultats.

III.4.1. la Base des images

La base d'images utilisée comporte 90 images couleur. Elle a été téléchargée depuis le site web [20]. La base d'images originale comporte 1000 images divisées en 10 classes. Nous avons optimisé pour l'utilisation de 90 images et 5 classes pour des raisons de temps de calcul. Chaque classe représente un thème défini : Rose, Bus, Dinosaures, Chevaux et Couches de Soleil. Un échantillon de la base est présenté en **Figure 26**.

Rose

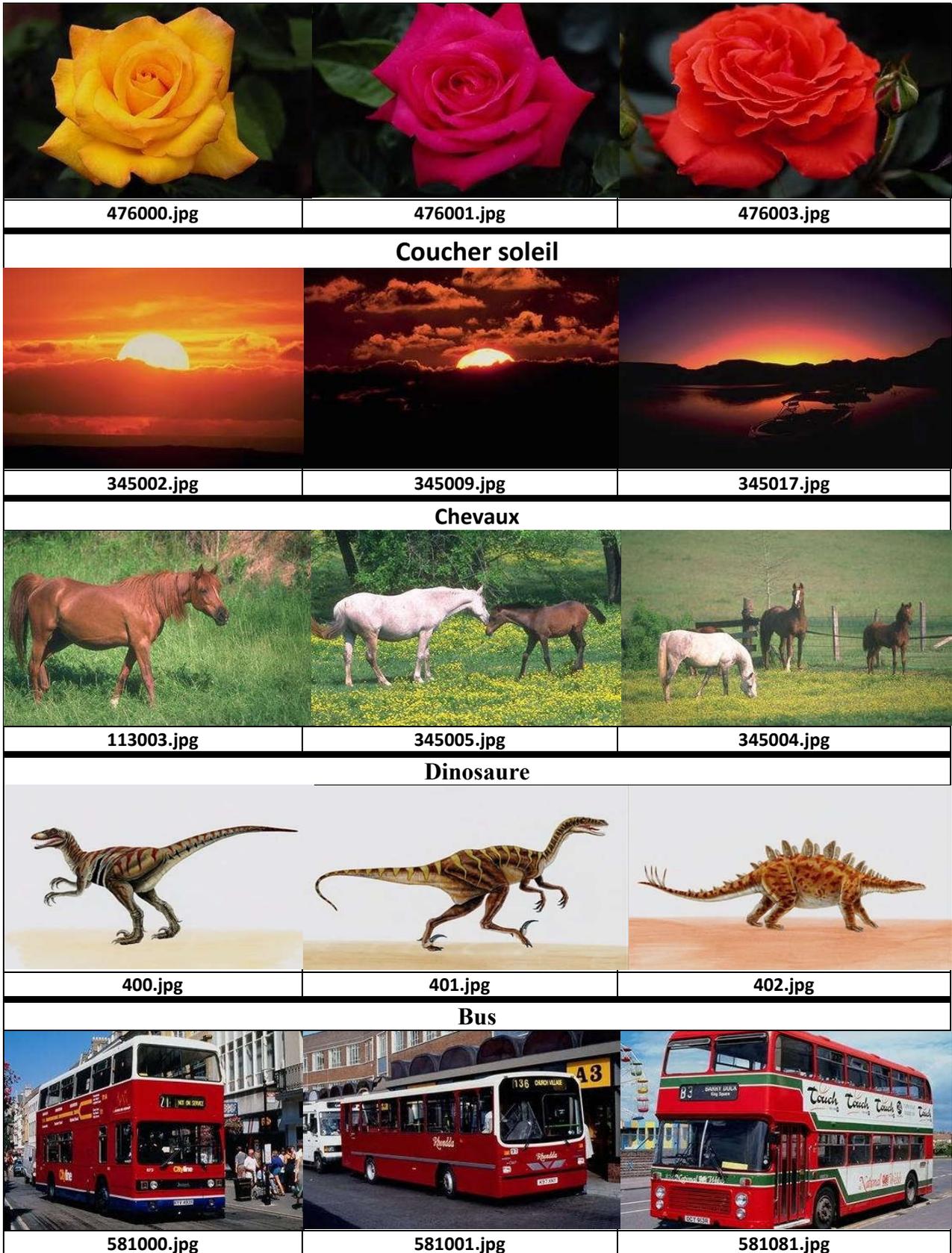


Figure 26 : Un échantillon de la base

III.4.2. Processus suivi

Dans cette partie nous avons évalué l'application avec le calcul de précision pour chaque type de requête. Le principe de fonctionnement de ce processus d'évaluation peut être d'écrite a par la figure 32.

Les images requêtent son sélectionnées de manière aléatoires (sans répétition de la même image requête). La figure 31 représente un exemple d'un ensemble d'images requête.

III.4.2.1. Requête visuelle

Remarque : il est clair que pour un système soit parfait, la valeur de précision égale à 1, dans notre cas on a 15 requêtes, dont les valeurs de précision varient entre 0 et 1, donc la précision moyenne globale sera calculée comme suite :

$$\text{Précision Moyenne Globale} = \frac{\sum Pr * 100}{15}$$

Pr : valeur de précision par requête.

Alors que la précision moyenne pour chaque classe doit être calculé par rapport à la précision moyenne globale :

$$\text{Précision Moyenne Classe} = \frac{\sum Pr * 100}{Pmg}$$

Pmg : Valeur de précision moyenne globale

Tableau 3 : Requêtevisuelle, Seuil = 50% du maximum de la distance Euclidienne.

Classe	Image requête	Total trouvé	Images pertinentes	Images non pertinentes	Précision	Rappel	Précision moyenne par classe
Rose :	476000.jpg	398	90	308	0,22	1	12,41 %
	476001.jpg	398	90	308	0,22	1	
	476003.jpg	398	90	308	0,22	1	
Coucher soleil :	345002.jpg	392	87	305	0,22	0,96	11,97 %
	345009.jpg	386	90	296	0,23	1	
	345017.jpg	391	78	313	0,19	0,86	
Chevaux :	113003.jpg	397	87	310	0,21	0,96	12,03%
	113005.jpg	400	87	313	0,21	0,96	
	113004.jpg	394	87	307	0,22	0,96	
Dinosaure :	400.jpg	409	90	319	0,22	1	11,81%
	401.jpg	406	90	316	0,22	1	
	402.jpg	402	82	320	0,20	0,91	
Bus :	581000.jpg	85	83	2	0,97	0,92	51,57%
	581001.jpg	86	83	3	0,96	0,92	
	581081.jpg	97	86	11	0,88	0,95	
Précision moyenne globale					36.42%		

Tableau 4 : Requêtevisuelle, Seuil = 25% du maximum de la distance Euclidienne.

Classe	Image requête	Total trouvé	Images pertinentes	Images non pertinentes	Précision	Rappel	Précision moyenne par classe
Rose :	476000.jpg	263	89	174	0,33	0,98	12.90 %
	476001.jpg	222	83	139	0,37	0,92	
	476003.jpg	252	88	164	0,34	0,97	
Coucher soleil :	345002.jpg	251	78	173	0,31	0,86	13.46 %
	345009.jpg	198	77	121	0,38	0,85	
	345017.jpg	191	78	113	0,40	0,86	
Chevaux :	113003.jpg	274	80	194	0,29	0,88	17.80 %
	113005.jpg	283	81	202	0,28	0,9	
	113004.jpg	53	47	6	0,88	0,52	
Dinosaure :	400.jpg	151	90	61	0,59	1	19.37 %
	401.jpg	138	90	48	0,65	1	
	402.jpg	243	84	159	0,34	0,93	
Bus :	581000.jpg	5	5	0	1	0,05	36.45 %
	581001.jpg	17	17	0	1	0,18	
	581081.jpg	23	23	0	1	0,25	
Précision moyenne globale					54.85 %		

Interprétation des résultats

D'après les résultats obtenus et illustrés par les tableaux ci-dessus, avec 25 % de la distance euclidienne maximal une précision moyenne globale des 5 classes égale à 54,85% est accomplie. Les deux meilleures précisions en été obtenu pour les classes « bus » et « dinosaures » car elle présente une hétérogénéité dans la distribution de la couleur et les Forme d'objet dans le fond.

La classe « Dinosaures » présente la plus grande précision (19.37%). Les images de cette classe ne sont pas des photographies mais des images synthétiques. Elles ont toutes un arrière-plan uni blanc. Cet arrière-plan permet de les distinguer des images des autres classes. Les cas d'erreur introduits sont des images qui contiennent une grande quantité de

couleur blanche. La classe « Bus » présentent une bonne précision (36.45%) pour la même raison.

III.4.2.2. Requête textuelle

Tableau 5 : Requête textuelle, Seuil = 94 % du maximum de la distance Euclidienne.

Classe	Image requête	Total trouvé	Images pertinentes	Images non pertinentes	Précision	Rappel	Précision moyenne par classe
Rose :	Roserouge	50	6	44	0,12	8,33	16.12 %
	Rose jaune	50	2	48	0,04	25	
	Rose blanc	50	2	48	0,04	25	
Coucher soleil :	Coucher soleil Algérie	50	6	44	0,12	8,33	16.12 %
	Coucher soleil Maroc	50	3	47	0,06	16,66	
	Coucher soleil Tunis	50	1	49	0,02	50	
Chevaux :	Chevaux marron	50	10	40	0,2	5	25.80 %
	Chevaux marron et gris	50	3	47	0,06	16,66	
	Chevaux marron blanc	50	3	47	0,06	16,66	
Dinosaure :	Dinosaure marron	50	4	46	0,08	12,5	14.51 %
	Dinosaure vert	50	3	47	0,06	16,66	
	Dinosaure gris	50	2	48	0,04	25	
Bus :	Bus rouge	50	7	43	0,14	7,14	27.41 %
	Bus rouge et blanc	50	8	42	0,16	6,25	
	Bus bleu	50	2	48	0,04	25	
Précision moyenne globale					8.26%		

Tableau 6 : tableau d'évaluation par requête textuelle, Seuil = 0 % du maximum de la distance.

Classe	Image requête	Total trouvé	Images pertinentes	Images non pertinentes	Précision	Rappel	Précision moyenne par classe
Rose :	Rose rouge	6	6	0	1	1	20 %
	Rose jaune	2	2	0	1	1	
	Rose blanc	1	1	0	1	1	
Coucher soleil :	Coucher soleil Algérie	5	5	0	1	1	20 %
	Coucher soleil Maroc	3	3	0	1	1	
	Coucher soleil Tunis	1	1	0	1	1	
Chevaux :	Chevaux marron	4	4	0	1	1	20 %
	Chevaux marron et gris	3	3	0	1	1	
	Chevaux marron blanc	3	3	0	1	1	
Dinosaure :	Dinosaure marron	4	4	0	1	1	20 %
	Dinosaure vert	1	1	0	1	1	
	Dinosaure gris	1	1	0	1	1	
Bus :	Bus rouge	4	4	0	1	1	20 %
	Bus rouge et blanc	3	3	0	1	1	
	Bus bleu	1	1	0	1	1	
Précision moyenne globale					100 %		

Interprétation des résultats

D'après les résultats obtenus et illustrés par les tableaux ci-dessus, la Précision moyenne globale est meilleure égale 100% lorsque la valeur de la distance maximum égale à 0%, ce qui donne une bonne représentation de l'information, mais les systèmes classiques de recherche d'information possèdent des problèmes persistants, par exemple la requête de recherche est souvent exprimée par une liste de termes souvent très réduite qui exprime souvent mal les besoins en information de l'utilisateur et pour une même requête le système retourne les mêmes résultats, alors que des utilisateurs différents peuvent formuler la même requête pour exprimer des besoins différents.

III.4.2.3. Requête visuelle et textuelle

Tableau 7 : tableau d'évaluation par requête visuelle et textuelle, Seuil = 94 % du maximum de la distance.

Classe	Image requête	Total trouvé	Images pertinentes	Images non pertinentes	Précision	Rappel	Précision moyenne par classe
Rose	476000.jpg+rose rouge	17	10	7	0,58	0,18	14.09 %
	476001.jpg +rose rouge	17	10	7	0,58	0,18	
	476003.jpg +rose jaune	396	90	306	0,22	4,4	
Coucher soleil	345002.jpg+ coucher soleil Algérie	10	10	0	1	0,11	22.07 %
	345009.jpg +coucher soleil Algérie	10	10	0	1	0,11	
	345017.jpg +coucher soleil Maroc	453	90	363	0,19	5,03	
Chevaux	113003.jpg+chevaux marron	14	10	4	0,71	0,15	20.30 %
	113005.jpg +chevaux marron	14	10	4	0,71	0,15	
	113004.jpg +chevaux gris	5	3	2	0,6	0,05	
Dinosaure	400.jpg+dinosaure marron	14	10	4	0,71	0,15	17.69 %
	401.jpg +dinosaure marron	14	10	4	0,71	0,15	
	402.jpg +dinosaure gris	3	1	2	0,33	0,03	
Bus	581000.jpg+bus rouge	9	9	0	1	0,1	25.77 %
	581001.jpg +bus rouge	10	9	1	0,9	0,11	
	581081.jpg +bus bleu	3	2	1	0,66	0,03	
Précision moyenne globale					66.39 %		

Tableau 8 : tableau d'évaluation par requête visuelle et textuelle, Seuil = 92 % du maximum de la distance.

Classe	Image requête	Total trouvé	Images pertinentes	Images non pertinentes	Précision	Rappel	Précision moyenne par classe
Rose :	476000.jpg+rose rouge	10	10	0	1	0,11	20 %
	476001.jpg +rose rouge	10	10	0	1	0,11	
	476003.jpg +rose jaune	2	2	0	1	0,02	
Coucher soleil :	345002.jpg+ coucher soleil Algérie	9	9	0	1	0,1	20 %
	345009.jpg +coucher soleil Algérie	10	10	0	1	0,11	
	345017.jpg +coucher soleil Maroc	10	10	0	1	0,11	
Chevaux :	113003.jpg+chevaux marron	10	10	0	1	0,11	20 %
	113005.jpg +chevaux marron	10	10	0	1	0,11	
	113004.jpg +chevaux gris	3	3	0	1	0,03	
Dinosaure :	400.jpg+dinosaure marron	4	4	0	1	0,04	20 %
	401.jpg +dinosaure marron	4	4	0	1	0,04	
	402.jpg +dinosaure gris	2	2	0	1	0,02	
Bus :	581000.jpg+bus rouge	5	5	0	1	0,05	20 %
	581001.jpg +bus rouge	6	6	0	1	0,06	
	581081.jpg +bus bleu	1	1	0	1	0,01	
Précision moyenne globale					100%		

Interprétation des résultats

D'après les résultats obtenus et illustrés par les tableaux ci-dessus, la Précision moyenne globale est meilleur égale 100% lorsque la valeur de la distance maximum égale à 92%, mais le problème reste dans la formulation du requête utilisateur ou la combinaison entre les deux

requêtes, comment choisir une requête d'image et sa requête textuelle, pour obtenir de bon résultat souhaiter.

III.4.3. Problèmes rencontrés

La première problématique qui s'est imposée durant la réalisation de notre application est le choix des descripteurs discriminants d'image pour s'assurer de l'obtention des meilleurs résultats et pour une évaluation significative.

- L'annotation manuelle des images de la base de données (La saisie de chaînes de caractères) exige beaucoup d'effort (description pertinente de la scène), et nécessite trop de temps.
- Aussi le choix du modèle de couleur et la mesure de distance à utiliser.

Conclusion

Nous avons présenté dans ce chapitre les caractéristiques textuelles et les caractéristiques visuelles choisis (modèle de couleur, préparation de l'index, la mesure de similarité utilisé et les mesures de performance étudiées). Après, on a présenté les grandes parties de notre application et les différents résultats obtenus pendant l'évaluation de notre application de recherche d'image par le contenu.

Conclusion Générale

Les méthodes de recherche d'images fournissent un moyen d'aide à la décision dans de nombreux domaines d'application. Une application de type CBIR a été développée. Les différentes composantes de son architecture ont été exposées. Les deux modalités de recherche ont été utilisées à savoir une requête image ou texte. La distance euclidienne a été choisie comme mesure de similarité. L'annotation manuelle des images de base de données est effectuée. Pour l'indexation des images, un ensemble de 197 descripteurs de bas niveau basé sur la couleur (modèles RGB et HSV), texture et forme est adopté. Pour les requêtes de type texte nous avons expérimenté le modèle vectoriel.

Plusieurs tests ont été effectués avec des seuils de pourcentage sur la distance euclidienne maximale (10% et 95%).

Une étude approfondie des descripteurs de bas niveau pris individuellement ainsi que la combinaison de ces derniers avec les descripteurs textuels doit être entreprise. Comme perspectives, Nous proposons :

- Utiliser d'autres descripteurs (basé sur des concepts haut niveau) afin de réduire le gap sémantique.
- Appliquer les descripteurs de bas niveau à des régions d'intérêt de l'image.
- **Travailler avec tous les extensions des images existant.**
- Utiliser d'autres mesures de similarité entre les descripteurs d'images.
- Tester l'application avec d'autres bases de données (autres classes d'images).
- Exploiter le potentiel que peut procurer le Big-Data, spécifiquement pour l'annotation automatique des bases de données massives.

Nous espérons que notre travail constituera une base solide pour les prochains étudiants qui pourront approfondir les points cités en perspective.

Bibliographie

- [1]. Kamel Houari, Recherche d'images par le contenu Doctorat kin, Université Mentouri de Constantine, 2010.
- [2]. MERABET Nabila, MAHLIA Meriem, Recherche d'images par le contenu Master en Informatique, Université Abou BakrBelkaid– Tlemcen, 2010-2011.
- [4]. Imane NEDJAR, CMBIR (Content MedicalBased Image Retrieval) développement d'outil logiciel d'annotation d'images médicales utilisant les méthodes d'indexation par descripteurs invariants de contenus Magister en Informatique, Université Abou BakrBelkaid– Tlemcen, 2011-2012.
- [5]. Amrouni Samia, Recherche et classement d'images dans une base de données de grande taille à partir de requêtes relatives à leurs descripteurs. Magister en Automatique, Université Mouloud Mammeri, Tizi-Ouzou.
- [6]. Z. Guellil et L. Zaoui, Architecture des bases d'images généralistes organisées en clusters. Université des sciences et de la technologie d'Oran MB, Université Mohamed Boudiaf USTO -BP 1505 El Mnaouer -ORAN – Algérie.
- [7]. M^{elle}. BEDOUHENE Saïda, Recherche d'images par le contenu MEMOIRE DE MAGISTER en Automatique, UNIVERSITE MOULOUD MAMMERI, TIZI-OUZOU.
- [8]. L. Zhang and Y. Rui, Image Search-From Thousands to Billions in 20 Years, ACM Trans. Multimedia Computing, Communications, and Applications (TOMCCAP), vol. 9, no. 1s, 2013, article no. 36.
- [10]. Sabriakram, Benali Moustafa, Développement d'une application de traitement d'images, Mémoire Licence en Informatique, 2013-2014.
- [11]. Maude Manouvrier; Bases de Données Images, LAMSADE- Université Paris Dauphine.
- [12]. Sylvie Szulman, LIPN, Université Paris 13, Construction de ressources sémantiques à partir de textes.
- [13]. Madalina MITRAN, Annotation d'images via leur contexte spatio-temporel et les métadonnées du Web, DOCTORAT DE L'UNIVERSITE DE TOULOUSE, 15/07/2014.
- [14]. Simon PIERARD, Développement d'une méthode de représentation d'images sur base d'un vocabulaire visuel, Travail de fin d'études, FACULTÉ POLYTECHNIQUE DE MONS.

- [15]. Yassine Ariba, Jérôme Cadieux, Manuel Matlab, Départements GEI & Mécanique, Icam de Toulouse, p 06.
- [16]. Sous la direction de Dr. Alain BOUCHER, Indexation et Recherche d'images par le contenu, Mémoire de Master, Institut Polytechnique de HANOI, 2005.
- [17]. Avinash N Bhute, B. B. Meshram, Content Based Image Indexing and Retrieval, VJTI, Matunga, Mumbai.
- [18]. Shruti S. Kulkarni, Nileshsingh V. Thakur, An Approach for Image Segmentation based on Region Connectivity, PRM College of Engg. & Mgmt, Badnera, India, ISSN (Print): 2278-8948, Volume-4 Issue-3, 2015.
- [19]. Ricardo da Silva Torres, Alexandre Xavier Falcão, Content-Based Image Retrieval: Theory and Applications, Institute of Computing, State University of Campinas, Campinas, SP, Brazil. rortorres.afalcao@ic.unicamp.br
- [21] Cédric Lemaître, Christophe Moulin, Cécile Barat, Christophe Ducottet, Combinaison d'information visuelle et textuelle pour la recherche, Lyon, France.
- d'information multimédia, Englewood Cliffs, Nj, 1971.
- [22]. Abeer Esa A M Alomairi, Ghazali Sulong, An Overview Of Content-Based Image Retrieval Techniques, UTM-IRDA Digital Media Centre (MaGIC-X), Faculty of Computing, University Technology, Journal of Theoretical and Applied Information Technology, 20 Février 2016, Vol.84, N°2.
- [23]. Soumendra Mohanty, Madhu Jagadeesh, Harsha Srivatsa, Big Data Imperatives, Enterprise Big Data Warehouse, BI Implementations and Analytics, 2013.

Webographie

- [3]. <http://www-ia.lip6.fr/~tollaris/ARTICLES/THESE/node4> visité le 14/04/2016.
- [9]. https://fr.wikipedia.org/wiki/Big_data visité le 02/08/2016.
- [20]. <http://wang.ist.psu.edu/docs/related/> visité le 15/05/2016.