

Sommaire

INTRODUCTION GÉNÉRALE.....	6
Contexte générale.....	6
1.1. Première génération.....	6
1.2. Deuxième génération.....	7
1.3. Troisième génération.....	7
2. Objectifs de l'étude.....	7
3. Organisation du mémoire.....	8
CHAPITRE 1 : GÉNÉRALITÉS SUR LE TRAITEMENT D'IMAGES.....	9
1. Introduction.....	9
2. Définition d'image.....	9
2.1. Types d'images.....	9
2.2. Formats d'image.....	18
3. Caractéristiques d'images.....	19
3.1. Pixel.....	19
3.2. La texture.....	19
3.3. Dimension.....	20
3.4. Résolution.....	20
3.5. Bruit.....	20
3.6. Histogramme.....	20
3.7. Contours et textures.....	20
3.8. Luminance.....	20
3.9. Contraste.....	21
4. Système de traitement d'image.....	21
5. Filtrage.....	21
5.1. Filtres linéaires.....	22
5.2. Filtres non linéaires.....	24
6. Segmentation :.....	25
7. Quelques applications concrètes de traitement d'images.....	26
8. Conclusion.....	27
CHAPITRE 2 : SYSTÈMES DE RECHERCHE D'IMAGE PAR CONTENU (CBIR).....	28
1. Introduction.....	28
2. Composants d'un CBIR.....	29
2.1 La base d'image.....	29

Recherche d'Image par le Contenu

2.2 L'indexation.....	32
2.3 La gestion des index.....	33
2.4 Les requêtes.....	33
2.5 Analyse de la requête.....	34
2.6 Mise en correspondance requête / base.....	34
2.7 La présentation des résultats.....	35
3. Représentation des images dans un CBIR.....	35
4. Mesures pour évaluer un système.....	35
4.1. Rappel et précision (en anglais : Recall and Precision).....	36
4.2. La courbe de rappel et précision.....	37
5. Conclusion.....	39
CHAPITRE 3 : DESCRIPTEURS D'IMAGES & MESURES DE SIMILARITÉ.....	40
1. Introduction.....	40
2. Descripteurs d'image.....	40
2.1. Descripteurs de couleur.....	40
2.2. Descripteurs des textures.....	44
2.3. Descripteurs de Formes.....	47
3. Mesures de similarité.....	50
A. Images brutes (pixels).....	50
B. Attributs visuels.....	50
3.1. Les méthodes de calcul.....	50
4. Conclusion.....	53
CHAPITRE 4 : LE TRAVAIL EFFECTUÉ.....	54
1. Introduction.....	54
2. Les caractéristiques visuelles choisies.....	54
2.1. Les modèles de couleur utilisés.....	54
2.2. Préparation des indexes.....	54
2.2.1 Les histogrammes de couleur.....	54
2.2.2 Le nombre d'objets d'image.....	55
2.2.3 Les moments statistiques.....	55
2.3. Mesures de distance utilisées.....	56
2.4. Les mesures de performance étudiées.....	56
3. Architecture de l'application.....	56
3.1 L'interface utilisateur.....	57
3.2 Le moteur de recherche.....	60

Recherche d'Image par le Contenu

3.3La base d'attributs.....	60
4. Evaluation de l'application	61
4.1La base d'image utilisée	61
4.2Processus suivi	61
5. Problèmes rencontré	65
6. perspective	65
7. conclusion	65
CONCLUSION GÉNÉRALE.....	66
BIBLIOGRAPHIE	67
SITOPGRAPHIE	71
RÉSUMÉ	72

Table des figures

Figure 1 : Le spectre électromagnétique.....	11
Figure 2:Les trois couleurs primaires additives de lumière : Rouge, Vert, et Bleu	12
Figure 3 : Le modèle de couleur RVB	12
Figure 4 : Les couleurs primaires de lumière et leur complémentaires (couleurs secondaires)	13
Figure 5 : La teinte, la saturation, et luminosité représentées sur une roue de couleurs sous forme d'un cône.	14
Figure 6 : Comparaison entre la saturation et la valeur dans le modèle TSL	15
Figure 7 : Représentation du modèle CIE XYZ par le triangle de Maxwell	15
Figure 8 : Le modèle CIE LAB.	16
Figure 9 : Application du filtre moyen (lissage).....	22
Figure 10 : Rehaussement de contour	24
Figure 11 : Principe du filtre médian	25
Figure 12 : Principe du filtre maximum	25
Figure 13 : Principe du filtre minimum.....	25
Figure 14 : Principaux composants d'un Système de Recherche par le Contenu	29
Figure 15 :10 classes de la base de Wang	30
Figure 16 : Les objets utilisés dans COIL-100	31
Figure 17 : Quelques images exemples de la base de Pollens	31
Figure 18 : Quelques images exemples dans la base de CURET(CURET).....	32
Figure 19 : Quelques images exemples dans la base de Fei-Fei.....	32
Figure 20 : Le rappel et la précision pour une requête.....	37
Figure 21 : Une courbe de rappel et précision.....	38
Figure 22 : Comparaison entre histogrammes	43
Figure 23 : Des textures différentes	46
Figure 24: Exemples d'histogrammes.....	54
Figure 25: architecture de l'application	57
Figure 26: L'interface utilisateur.....	58
Figure 27: Le choix de l'image requête et de la classe.	59
Figure 28: indexation de l'image requête.....	59
Figure 29: Un exemple d'une recherche effectué.....	60
Figure 30: Un échantillon de la base d'images utilisée.	61
Figure 31: schémas représentative du processus	62
Figure 32: un exemple d'un ensemble d'image requête.....	63
Figure 33: les précisions moyennes obtenues.....	64

Liste des Abréviations

CBIR : Content Based Image Retrieval.

CIE : Commission Internationale de l'Eclairage.

EMD: Earth Mover Distance.

RGB: Red Green Blue.

RI : Recherche d'Information.

RIC : Recherche d'Information par le Contenu.

SRI : Systèmes de Recherche d'Information.

SRIC : Systèmes de recherche d'image par contenu, Content Based Image Retrieval

TH : Transformée de Hough.

INTRODUCTION GÉNÉRALE

“Quand on commence, on ne sait pas comment ça finira.”

La recherche d'images a évolué dans le temps selon les besoins des utilisateurs. Les progrès technologiques ont beaucoup influencé cette évolution.

Contexte générale

La recherche d'information est le processus de trouver un objet (en général un document) de nature non structurée (souvent) qui satisfait un besoin d'information au sein de collections de données de tailles très importantes stockées sur des supports de stockage.

La recherche d'information n'a pas commencé avec le Web mais plutôt avant. En réponse à différents problèmes de fourniture d'accès à l'information, le domaine de la recherche d'information a évolué pour donner de nouveaux principes et de nouvelles approches pour la recherche d'information de nature variée (l'information recherchée peut être un document texte, une image, une vidéo etc.).

Chaque jours des millions d'octets sont générés et stockés sur des supports du stockage. La recherche de l'information voulue devient de plus en plus difficile et présente des difficultés majeures, c'est comme trouver une aiguille dans une meule de foin.

Nous nous intéressons dans notre travail à la recherche d'images. Les images numériques possèdent une position prédominante parmi les différents types de données multimédia. À la différence de la vidéo et l'audio, qui sont surtout utilisés pour le divertissement et l'industrie des informations (newscasting), les images sont au cœur d'un large éventail de domaines (l'histoire artistique, la médecine, l'astronomie, l'exploration pétrolière, et les prévisions météorologiques...).

L'imagerie numérique joue un rôle important dans de nombreuses activités humaines, telles que : l'application de la loi (surtout l'investigation des scènes de crime), l'agriculture et la gestion forestière, les sciences de la terre, la planification urbaine, ainsi que les sports, et le divertissement.

La recherche d'images par le contenu a beaucoup évolué dans le temps, de ce fait [7] divise cette évolution en trois générations différentes:

1.1. Première génération

Dans la première génération des systèmes de recherche d'images, les images sont stockées dans une base de données, et peuvent être trouvées par le biais de chaînes de caractères qui lui sont liées. Ces chaînes de caractères peuvent être liées à un élément de l'image, ou à l'image elle-même. Ces chaînes sont stockées et peuvent être recherchées de manière structurée, comme dans les bases de données SQL. La technique utilisée dans ces systèmes souffre de limitations importantes :

Les descripteurs textuels dépendent de ce que l'utilisateur peut saisir lors de la création de la base d'images. Des utilisateurs différents peuvent décrire la même image d'une manière

différente, et le même utilisateur peut décrire la même image d'une manière différente en analysant l'image une deuxième fois.

Plusieurs attributs d'image, par exemple, la texture et la distribution de la couleur, sont difficiles à décrire sans ambiguïté en utilisant les descripteurs textuels.

La saisie de chaînes de caractères dans une base de données exige beaucoup d'effort, et nécessite trop de temps.

La recherche dans ce cas se fait selon l'approche textuelle traditionnelle des bases de données relationnelles.

1.2. Deuxième génération

La deuxième génération de systèmes de recherche d'images offre différentes manières d'interroger la base d'images, permettant des recherches selon les attributs visuels des images comme la texture, la forme et la couleur. Cette approche (basée sur les attributs visuels) peut être combinée avec la recherche basée sur le texte.

La recherche dans les systèmes de cette génération se base sur une mesure de similarité qui classe les images selon leur degré de similitude à l'image requête.

1.3. Troisième génération

Ces systèmes sont encore en cours de réalisation. Ils sont supposés être capables de travailler d'une manière intelligente, semblable au fonctionnement du système visuel humain. Ces systèmes devaient apprendre à partir de quelques exemples et de tirer des conclusions fondées sur l'expérience. Ces systèmes restent encore hypothétiques car la connaissance du système visuel humain est limitée.

Nous nous intéressons dans notre travail à la recherche d'images basées sur leurs attributs visuels. Comme nous l'avons déjà mentionné, la recherche basée sur le contenu visuel se base sur l'extraction des attributs visuels des images de la base, et les comparer à ceux de l'image requête.

2. Objectifs de l'étude

L'objectif principal de ce mémoire est d'étudier le principe de la recherche d'images par le contenu.

Le deuxième objectif consiste à utiliser les connaissances tirées de cette étude et construire un prototype d'un système de recherche d'images par le contenu.

Nous nous intéressons dans notre cas aux images couleurs. Les attributs de couleur sont considérés parmi les attributs bas niveau les plus utilisés pour la recherche d'images dans les bases d'images. Ils ont été utilisés dans plusieurs systèmes de recherche par le contenu. L'utilisation de la couleur dans ce domaine est motivée par deux facteurs. D'abord, la couleur est un descripteur puissant qui facilite l'identification et l'extraction d'objets à partir d'une scène. De plus, les humains peuvent discerner des milliers de nuances et d'intensités de

couleurs différentes. Nous avons essayé par le présent travail de construire une application d'un système de recherche d'images couleurs par le contenu.

3. Organisation du mémoire

Ce mémoire est structuré en quatre chapitres.

Le premier chapitre porte sur une généralité de traitement d'image. Ce chapitre décrit la définition d'une image, leurs caractéristiques, le principe du système de traitement d'image, filtrage, segmentation et quelques applications concrets sur le traitement d'image.

Le deuxième chapitre concerne la définition du domaine, il s'agit de la recherche d'images par le contenu. Ce chapitre explique le principe de fonctionnement des systèmes de recherche d'images par le contenu, leurs composants et les mesures pour évaluer un système de recherche d'image par contenu. Le troisième chapitre décrit les descripteurs des images tels que la couleur, la texture et la forme, ce chapitre représente aussi un tour d'horizon portant sur les mesures de similarité entre les images de la base.

Le quatrième chapitre présente les différentes étapes de notre travail.

En conclusion nous présentons les points importants de ce travail et quelques perspectives qui découlent de cette étude.

CHAPITRE 1 : GÉNÉRALITÉS SUR LE TRAITEMENT D'IMAGES



1. Introduction

Le traitement d'images désigne une discipline des mathématiques appliquées qui étudie les images numériques et leurs transformations, dans le but d'améliorer leur qualité ou d'en extraire de l'information.

Il s'agit donc d'un sous-ensemble du traitement du signal dédié aux images et aux données dérivées comme la vidéo (par opposition aux parties du traitement du signal consacrées à d'autres types de données : son et autres signaux monodimensionnels notamment), tout en opérant dans le domaine numérique (par opposition aux techniques analogiques de traitement du signal, comme la photographie ou la télévision traditionnelles).

Dans ce chapitre nous présentons quelques notions de base du domaine de traitement d'image numérique tels que : la définition d'image, les types d'image, formats d'images, caractéristiques d'image, système de traitement d'image, analyse élémentaire, filtrage, segmentation et en fin quelques exemples concrets de traitement d'images.

2. Définition d'image

- 🌐 L'image est définie par le nombre de points qui la composent. En imagerie numérique, cela correspond au nombre de pixels qui compose l'image en hauteur (axe vertical) et en largeur (axe horizontal) : *200 pixels par 450 pixels* par exemple, abrégé en « 200×450 ».
- 🌐 Une image est simplement la représentation d'une scène. Elle est issue du contact des rayons lumineux provenant des objets formant la scène avec un capteur (caméra, scanner, rayons X, ...). Il s'agit en réalité que d'une représentation spatiale de la lumière.

2.1. Types d'images

On peut classer les images numériques selon deux critères:

2.1.1. Selon la représentation numérique :

Ils existent deux types:

a) Images matricielles (ou images bitmap)

Elle est composée comme son nom l'indique d'une matrice (tableau) de points à plusieurs dimensions, chaque dimension représentant une dimension spatiale (hauteur, largeur, profondeur), temporelle (durée) ou autre (par exemple, un niveau de résolution).

b) Images vectorielles

Le principe est de représenter les données de l'image par des formules géométriques qui vont pouvoir être décrites d'un point de vue mathématique. Cela signifie qu'au lieu de mémoriser une mosaïque de points élémentaires, on stocke la succession d'opérations conduisant au tracé. Par exemple, un dessin peut être mémorisé par l'ordinateur comme « une droite tracée entre les points (x1, y1) et (x2, y2) », puis « un cercle tracé de centre (x3, y3) et de rayon 30 de couleur rouge ».

L'avantage de ce type d'image est la possibilité de l'agrandir indéfiniment sans perdre la qualité initiale, ainsi qu'un faible encombrement. L'usage de prédilection de ce type d'images concerne les schémas qu'il est possible de générer avec certains logiciels de DAO (Dessin Assisté par Ordinateur). Ce type d'images est aussi utilisé pour les animations Flash, utilisées sur Internet pour la création de bannières publicitaires, l'introduction de sites web, voire des sites web complet.

2.1.2. Selon la représentation des couleurs:

Il existe plusieurs modes de codage informatique des couleurs, le plus utilisé pour le maniement des images est l'espace colorimétrique Rouge, Vert, Bleu (RVB ou RGB - Red green Blue).

2.1.2.1. LES MODÈLES DE COULEUR

La couleur est l'une des composantes principales pour la description et l'analyse des images couleurs dans le domaine de la recherche d'images par le contenu. Comme nous le savons, les couleurs perceptibles par l'œil humain représente une petite gamme de l'ensemble du spectre électromagnétique qui représente tous les rayons, depuis les rayons cosmiques et les rayons X jusqu'aux ondes électriques (Figure 1).

Comme le démontre la figure 1 du spectre électromagnétique, les couleurs visibles par l'œil humain s'étendent entre les longueurs d'ondes allant de 4000 à 7000 angströms, représentant respectivement les couleurs : violet et rouge et toutes les couleurs entre les deux. Toutes les autres ondes s'étendant des rayons cosmiques des étoiles jusqu'aux ondes FM de nos radios ne peuvent pas être perçues par l'œil humain. Cette petite gamme du spectre est l'espace de couleur perceptible par l'œil humain. [27]

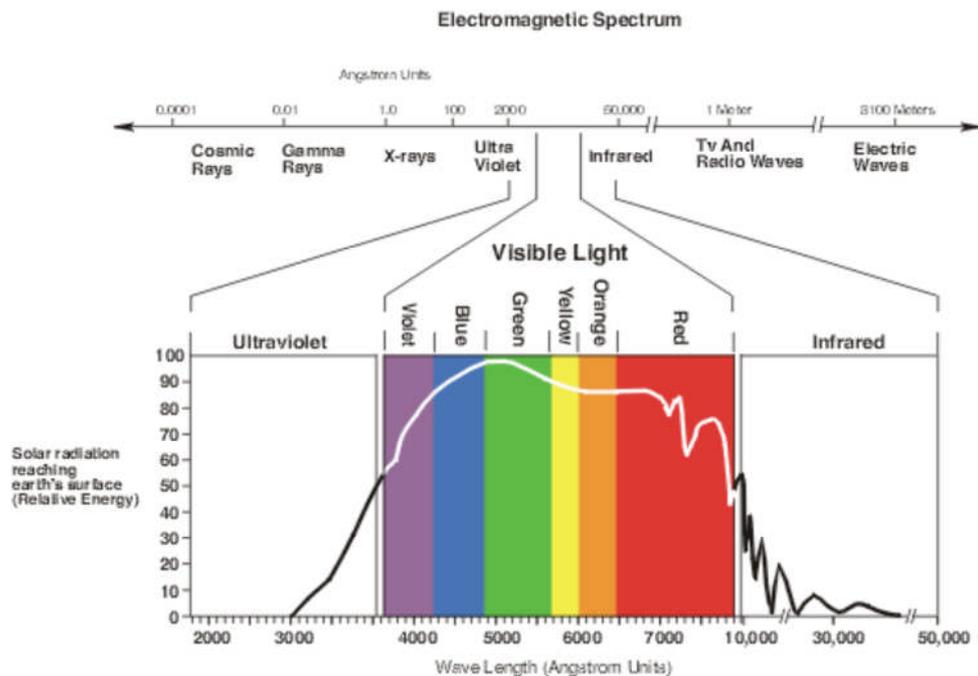


Figure 1 : Le spectre électromagnétique.

2.1.2.1.1. DÉFINITION DU MODÈLE DE COULEUR

Comme toutes les représentations mathématiques des phénomènes physiques, les couleurs peuvent être exprimées selon différentes manières, chacune ayant ses avantages et ses inconvénients.

Un modèle de couleur (modèle chromatique) est un modèle mathématique abstrait décrivant la façon dont les couleurs peuvent être représentées en tant que tuples de nombres, généralement de trois ou quatre valeur ou composantes de couleur. Le but d'un modèle de couleur est de faciliter la spécification des couleurs d'une manière. [27]

Les modèles sont des abstractions qui ne peuvent pas décrire une couleur spécifique sans que la référence ne soit définie au préalable.

La première distinction majeure entre les modèles de couleur est la dépendance du dispositif utilisé. Les coordonnées d'une couleur dans un modèle indépendant sont toujours les mêmes sur tous les dispositifs d'affichage. Un exemple de cette catégorie est le modèle XYZ de la CIE (Commission Internationale de l'Eclairage). D'un autre côté, un modèle qui dépend du dispositif de sortie aura différentes coordonnées pour la même couleur pour des dispositifs différents. RVB (Rouge, Vert, Bleu) et CMJ (Cyan, Magenta, Jaune) sont des exemples de cette catégorie.

La deuxième distinction est l'uniformité perceptuelle des couleurs. Ceci veut dire qu'une même variation de la valeur des composantes est toujours perçue comme la même variation

de couleur. En d'autres termes, la mesure de la variation perçue par un humain est égale à la mesure de la distance mathématique.

Dans ce qui suit nous allons présenter quelques modèles de couleur.

A) LE MODÈLE RVB(en anglais RGB) :

Le modèle RVB est le modèle de couleur le plus utilisé pour la représentation de la couleur. Il est composé des trois couleurs primaires : rouge, vert et bleu. Ce modèle est aussi le plus utilisé pour reproduction de la couleur sur les dispositifs d'affichage tels que la télévision et les écrans des ordinateurs. Ces trois couleurs sont appelées les couleurs primaires additives (Figure 2).

Par la variation de leur combinaison, d'autres couleurs peuvent être obtenues. Ce modèle peut être représenté sous forme d'un cube unitaire (Figure 3) 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 (Figure 4): cyan (0, 1,1), magenta (1, 0,1), et le jaune (1, 1,0).

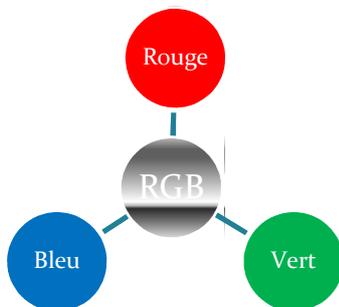


Figure 2:Les trois couleurs primaires additives de lumière : Rouge, Vert, et Bleu

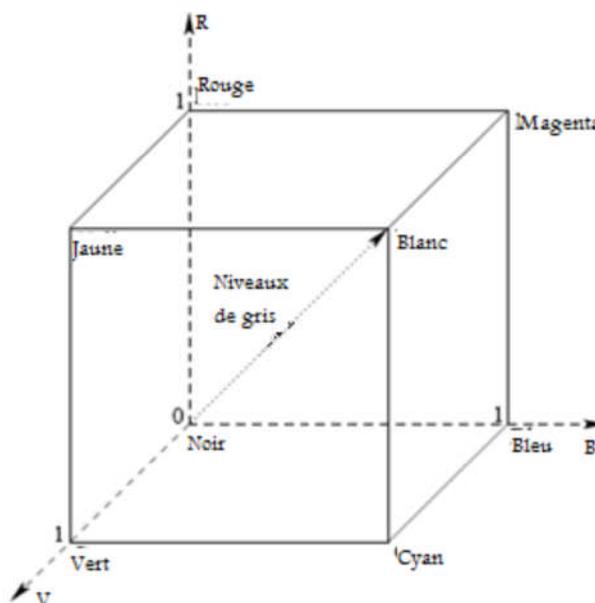


Figure 3 : Le modèle de couleur RVB

B) LE MODÈLE CMJ (CYAN, MAGENTA, JAUNE)

C'est un modèle basé sur les couleurs secondaires du modèle RVB et il est utilisé surtout pour les tâches d'impression (sur papier). Cyan, magenta et le jaune sont les couleurs complémentaires des couleurs : rouge, vert et bleu. Comparé au modèle RVB, aucune des composantes du CMJ est une couleur pure. Il est impossible de créer un noir pur en utilisant ce modèle. Pour surpasser ce problème, le modèle CMJ a été étendu vers un autre modèle nommé CMJN qui utilise le noir (N) comme une quatrième couleur.

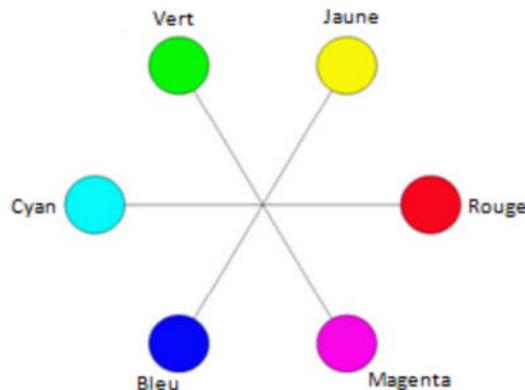


Figure 4 : Les couleurs primaires de lumière et leur complémentaires (couleurs secondaires)

C) LE MODÈLE TSL (TEINTE, SATURATION, LUMINOSITÉ)

Un modèle de couleur dérivé de la façon dont les couleurs apparaissent à un observateur (humain) est le modèle TSL. Ce modèle peut être représenté sous forme d'un cylindre, mais généralement il est représenté sous forme d'un cône. Comme le démontre la Figure 5, la teinte représente le composant chromatique dans ce modèle et c'est la définition d'une couleur par la combinaison des couleurs primaires. La saturation se rapporte à la dominance d'une teinte particulière dans une couleur. Une couleur moins saturée est plus proche du blanc tandis qu'une couleur plus saturée est plus proche de la couleur pure trouvée sur l'arête externe du diagramme de la roue de couleurs de TSL (vers les couleurs pures). La luminosité d'une couleur se rapporte à l'intensité (la légèreté ou la noirceur de la couleur). Malgré que les deux composants semblent être similaires, ils ont des effets différents sur la visibilité d'une couleur (Figure 6).

D) LES MODÈLES DE LA CIE (COMMISSION INTERNATIONALE DEL'ÉCLAIRAGE)

Dans le but de favoriser la compréhension de la couleur, la CIE (Commission International de l'Eclairage), autorité spécialisée dans tout ce qui touche à la lumière, à l'éclairage, à la couleur et aux espaces chromatiques, a adopté des normes dans les années 1930 à propos de plusieurs espaces chromatiques représentant le spectre visible. Cela a rendu la comparaison possible entre les différents espaces chromatiques des différents écrans d'affichage et périphériques sur le marché.

La CIE a mené un ensemble de tests sur un grand nombre de personnes afin de définir un observateur humain moyen hypothétique et sa réaction à la couleur, soit un profil désigné comme `` l'observateur de référence``.

L'œil humain possédant trois types de capteurs de couleurs qui réagissent à des plages de longueurs d'ondes différents, il faudrait un diagramme tridimensionnel pour obtenir une représentation complète de la totalité des couleurs visibles.

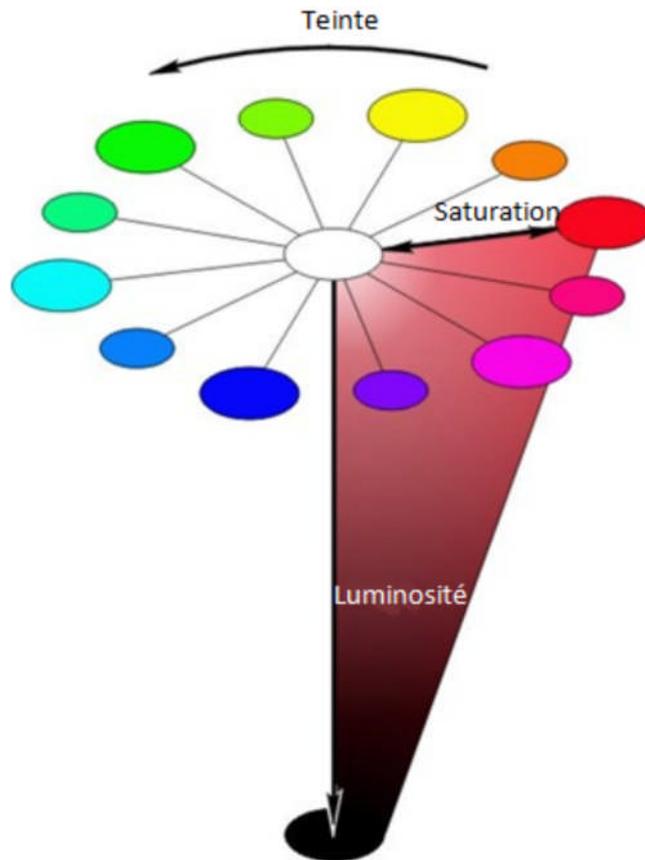


Figure 5 : La teinte, la saturation, et luminosité représentées sur une roue de couleurs sous forme d'un cône.

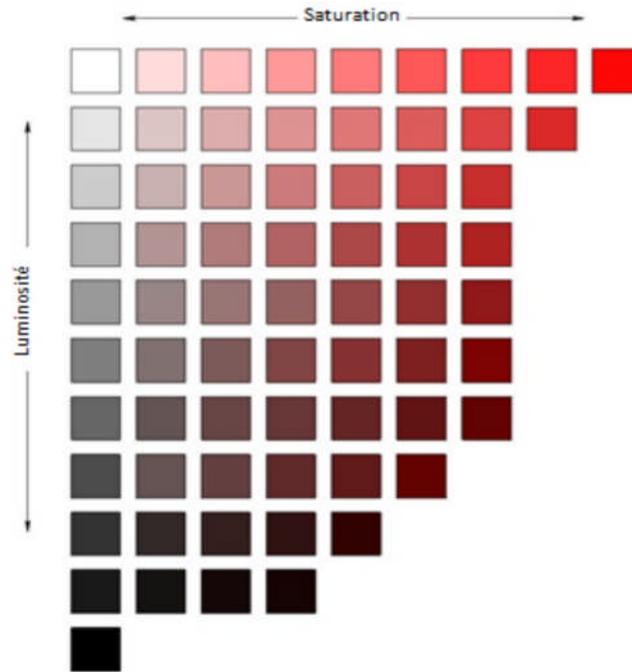


Figure 6 : Comparaison entre la saturation et la valeur dans le modèle TSL

E) LE MODELE CIE XYZ

La CIE a développé le système chromatique XYZ (Figure 7), également connu sous l'appellation "système chromatique de référence". Il est toujours utilisé comme système de référence habituel pour la définition des couleurs perçues par l'œil humain et pour d'autres espaces chromatiques. Tout comme le modèle chromatique RGB et ses primaires additives, le système CIE-XYZ (utilise trois primaires imaginaires définies spectralement (XYZ représentent trois couleurs primaires fictives ou imaginaires qui n'ont pas de corrélat physique réel i.e. il n'existe pas de sources de lumières réelles ou lampes qui aient ces couleurs. XYZ sont des entités mathématiques conçus pour faciliter les calculs de repérage en couleur.), X, Y et Z, qui sont la représentation de couleurs (ondes électromagnétiques) pouvant se combiner pour décrire toutes les couleurs visibles pour "l'observateur de référence".

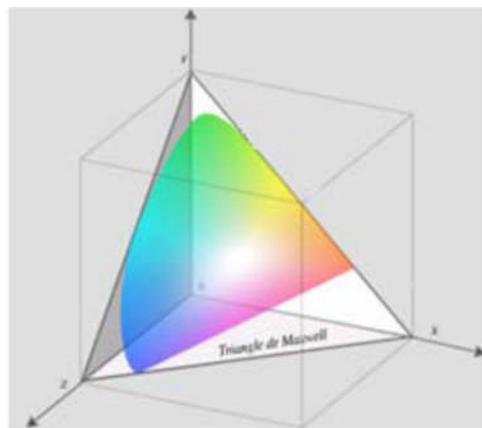


Figure 7 : Représentation du modèle CIE XYZ par le triangle de Maxwell

F) LE MODÈLE LAB

Le modèle chromatique Lab (Figure 8) a été développé par la CIE en 1976 dans le but d'améliorer la représentation des couleurs. Il s'agit du modèle chromatique le plus complet utilisé conventionnellement pour décrire l'ensemble des couleurs visibles pour l'œil humain. Il s'agit d'un espace chromatique tridimensionnel dans lequel les différences de couleurs perçues comme étant d'importance égale sont également séparées par des distances égales.

Chaque couleur peut être précisément désignée à l'aide de ses valeurs a et b spécifiques et de sa luminosité L. [27]

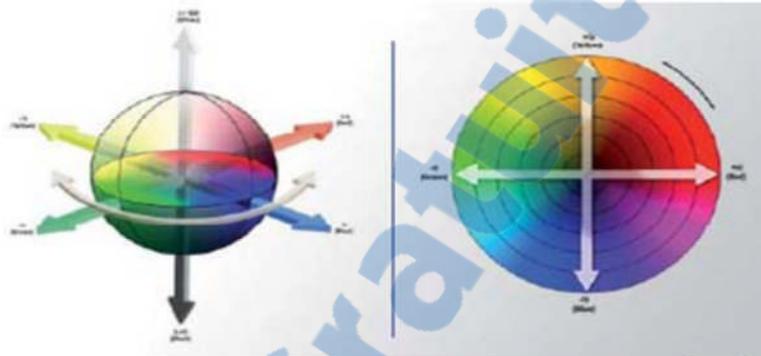


Figure 8 : Le modèle CIE LAB.

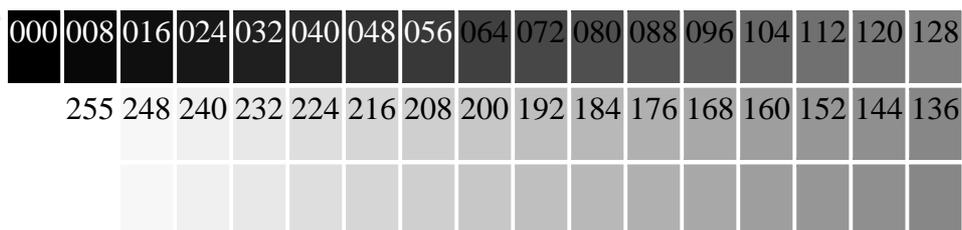
Les trois paramètres du modèle représentent la luminance de la couleur, L (le niveau L le plus faible donne du noir), sa position entre le rouge et le vert, a (le niveau a le plus faible donne du vert) et sa position entre le jaune et le bleu, b (le niveau b le plus faible donne du bleu), en fonction d'un point blanc de référence. Cependant, l'avantage de cet espace chromatique est son indépendance par rapport au matériel, et l'objectivité qui en résulte. La même combinaison de a, b et L décrit toujours exactement la même couleur. [27]

2.1.2.2. Les types d'images :

On distingue généralement les différents types d'images suivants :

A) Images en teintes (ou niveaux) de gris

On ne code ici plus que le niveau de l'intensité lumineuse, généralement sur un octet (256 valeurs). Par convention, la valeur zéro représente le noir (intensité lumineuse nulle) et la valeur 255 le blanc (intensité lumineuse maximale) :



Ce procédé est fréquemment utilisé pour reproduire des photos en noir et blanc ou du texte dans certaines conditions (avec utilisation d'un filtre pour adoucir les contours afin d'obtenir des caractères plus lisses).

B) Images à palettes, images en 256 couleurs (8 bits)

Pour réduire la place occupée par l'information de couleur, on utilise une *palette de couleurs* « attachée » à l'image. On parle alors de couleurs indexées : la valeur associée à un pixel ne véhicule plus la couleur effective du pixel, mais renvoie à l'entrée correspondant à cette valeur dans une table (ou palette) de couleurs appelée *look-up table* ou LUT en anglais, dans laquelle on dispose de la représentation complète de la couleur considérée.

Selon le nombre de couleurs présentes dans l'image, on peut ainsi gagner une place non négligeable : on considère en pratique que 256 couleurs parmi les 16 millions de couleurs 24 bits sont suffisantes. Pour les coder, on aura donc une palette occupant 24 bits x 256 entrées, soit 3 x 256 octets, et les pixels de l'image seront associés à des index codés sur un octet. L'occupation d'une telle image est donc de 1 octet par pixel plus la LUT, ce qui représente un peu plus du tiers de la place occupée par une image en couleurs 24 bits (plus l'image contient de pixels, plus le gain de place est important, la limite étant le tiers de la place occupée par l'image en couleurs vraies).

A) Images 24 bits (ou « couleurs vraies »)

Il s'agit d'une appellation trompeuse car le monde numérique (fini, limité) ne peut pas rendre compte intégralement de la réalité (infinie). Le codage de la couleur est réalisé sur trois octets, chaque octet représentant la valeur d'une composante couleur par un entier de 0 à 255. Ces trois valeurs codent généralement la couleur dans l'espace RVB. Le nombre de couleurs différentes pouvant être ainsi représenté est de 256 x 256 x 256 possibilités, soit près de 16 millions de couleurs. On considère commodément que ce système permet une restitution exacte des couleurs, c'est pourquoi on parle de « couleurs vraies ».

R	V	B	Couleur
0	0	0	noir
0	0	1	nuance de noir
255	0	0	rouge
0	255	0	vert
0	0	255	bleu
128	128	128	gris
255	255	255	blanc

Les images bitmap basées sur cette représentation peuvent rapidement occuper un espace de stockage considérable, chaque pixel nécessitant trois octets pour coder sa couleur.

2.2. Formats d'image

Un format d'image est une représentation informatique de l'image, associée à des informations sur la façon dont l'image est codée et fournissant éventuellement des indications sur la manière de la décoder et de la manipuler.

La plupart des formats sont composés d'un en-tête contenant des attributs (dimensions de l'image, type de codage, LUT, etc.), suivi des données (l'image proprement dite). La structuration des attributs et des données diffère pour chaque format d'image.

De plus, les formats actuels intègrent souvent une zone de métadonnées (*metadata* en anglais) servant à préciser les informations concernant l'image comme :

- la date, l'heure et le lieu de la prise de vue.
- les caractéristiques physiques de la photographie (sensibilité ISO, vitesse d'obturation, usage du flash...)

Ces métadonnées sont par exemple largement utilisées dans le format Exif (extension du format JPEG), qui est le format le plus utilisé dans les appareils photo numériques.

2.2.1. Précautions à prendre

Quelques précautions à prendre concernant les formats d'images :

- les formats dits « propriétaires », peuvent différer selon le logiciel qui les manipule. De plus, leur pérennité n'est pas garantie : réaliser de nouveaux programmes pour les lire peut s'avérer difficile (surtout si leurs spécifications n'ont pas été rendues publiques), cela peut même s'avérer illégal si les algorithmes utilisés sont protégés par des brevets ;
- il faut prêter attention aux différentes versions que peut recouvrir un format particulier, notamment pour le format TIFF qui varie selon les versions ; certaines d'entre elles ne sont pas reconnues par certains logiciels.

2.2.2. Tableau comparatif de différents formats d'images

	Type (matriciel/vectériel)	Compression des données	Nombre de couleurs supportées	Affichage progressif	Animation	Transparence
JPEG	matriciel	Oui, réglable (avec perte)	16 millions	Oui	Non	Non
JPEG2000	matriciel	Oui, avec ou sans perte	32 millions	Oui	Oui	Oui
GIF	matriciel	Oui, Sans perte	256 maxi (palette)	Oui	Oui	Oui
PNG	matriciel	Oui, sans perte	Palettisé (256 couleurs ou moins) ou 16 millions	Oui	Non	Oui (couche Alpha)
TIFF	matriciel	Compression ou pas avec ou sans pertes	de monochrome à 16 millions	Non	Non	Oui (couche Alpha)
SVG	vectériel	compression possible	16 millions	* ne s'applique pas *	Oui	Oui (par nature)

[s1]

3. Caractéristiques d'images

L'image est un ensemble structuré d'informations caractérisé par les paramètres suivants:[s2]

3.1. Pixel

Le pixel, abrégé px, est une unité de surface permettant de définir la base d'une image numérique. Son nom provient de la locution anglaise picture element, qui signifie, « élément d'image » ou « point élémentaire ».

3.2. La texture

Une texture est 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 est composée de texels, l'équivalent des pixels.

3.3. Dimension

C'est la taille de l'image. Cette dernière se présente sous forme de matrice dont les éléments sont des valeurs numériques représentatives des intensités lumineuses (pixels). Le nombre de lignes de cette matrice multiplié par le nombre de colonnes nous donne le nombre total de pixels dans une image.

3.4. Résolution

C'est la clarté ou la finesse de détails atteinte par un moniteur ou une imprimante dans la production d'images. Sur les moniteurs d'ordinateurs, la résolution est exprimée en nombre de pixels par unité de mesure (pouce ou centimètre). On utilise aussi le mot résolution pour désigner le nombre total de pixels affichables horizontalement ou verticalement sur un moniteur; plus grand est ce nombre, meilleure est la résolution.

3.5. Bruit

Un bruit (parasite) dans une image est considéré comme un phénomène de brusque variation de l'intensité d'un pixel par rapport à ses voisins, il provient de l'éclairage des dispositifs optiques et électroniques du capteur.

3.6. Histogramme

L'histogramme des niveaux de gris ou des couleurs d'une image est une fonction qui donne la fréquence d'apparition de chaque niveau de gris (couleur) dans l'image. Pour diminuer l'erreur de quantification, pour comparer deux images obtenues sous des éclairages différents, ou encore pour mesurer certaines propriétés sur une image, on modifie souvent l'histogramme correspondant.

Il permet de donner un grand nombre d'information sur la distribution des niveaux de gris (couleur) et de voir entre quelles bornes est répartie la majorité des niveaux de gris (couleur) dans les cas d'une image trop claire ou d'une image trop foncée.

Il peut être utilisé pour améliorer la qualité d'une image (Rehaussement d'image) en introduisant quelques modifications, pour pouvoir extraire les informations utiles de celle-ci.

3.7. Contours et textures

Les contours représentent la frontière entre les objets de l'image, ou la limite entre deux pixels dont les niveaux de gris représentent une différence significative. Les textures décrivent la structure de ceux-ci. L'extraction de contour consiste à identifier dans l'image les points qui séparent deux textures différentes.

3.8. Luminance

C'est le degré de luminosité des points de l'image. Elle est définie aussi comme étant le quotient de l'intensité lumineuse d'une surface par l'aire apparente de cette surface, pour un observateur lointain, le mot luminance est substitué au mot brillance, qui correspond à l'éclat d'un objet.

3.9. Contraste

C'est l'opposition marquée entre deux régions d'une image, plus précisément entre les régions sombres et les régions claires de cette image. Le contraste est défini en fonction des luminances de deux zones d'images.

Si L_1 et L_2 sont les degrés de luminosité respectivement de deux zones voisines A_1 et A_2 d'une image, le contraste C est défini par le rapport :

$$C = \frac{L_1 - L_2}{L_1 + L_2}$$

4. Système de traitement d'image

Dans le contexte de la vision artificielle, le traitement d'images se place après les étapes d'acquisition et de numérisation, assurant les transformations d'images et la partie de calcul permettant d'aller vers une interprétation des images traitées. Cette phase d'interprétation est d'ailleurs de plus en plus intégrée dans le traitement d'images, en faisant appel notamment à l'intelligence artificielle pour manipuler des connaissances, principalement sur les informations dont on dispose à propos de ce que représentent les images traitées (connaissance du domaine).

* Acquisition et numérisation :

C'est le mécanisme qui permet l'obtention d'une image numérique (représentée par une matrice) à deux dimensions à partir d'une scène à trois dimensions, en passant par un système optique, l'image continue $f(x,y)$ est approximée par des échantillons qui sont obtenus par discrétisation des coordonnées (x,y) (ce qu'on appelle l'échantillonnage) et la discrétisation des amplitudes de ces points là (c'est la quantification).

La représentation obtenue ne peut être parfaite à cause du bruit introduit dans l'image lors de son acquisition. [4]

*Visualisation :

Un dispositif de visualisation permet l'affichage de l'image. Les reconstituteurs permettent de transformer le signal numérique qui est la matrice image en un signal analogique visible à l'oeil humain pour cela on dispose d'une multitude de supports qui sont employés (moniteur vidéo, impression sur papier...). [4]

5. Filtrage

On peut scinder les filtres en deux grandes catégories :

5.1. Filtrés linéaires

Les premières et les plus simples méthodes de filtrage sont basées sur le filtrage linéaire, chacun de ses opérateurs est caractérisé par sa réponse impulsionnelle $h(x, y)$, l'expression qui relie les entrées avec les sorties est donnée par la relation suivante :

$$S(i, j) = \sum_u \sum_v (E(i, j) \cdot h(i-u, j-v)) = h(i, j) * E(i, j).$$

Où : u, v varient de moins l'infini à plus l'infini.

Le filtrage linéaire est un produit de convolution c à d une combinaison linéaire du voisinage du pixel concerné.

Les différents types des filtres linéaires sont :

5.1.1. FILTRE moyenneur (lissage):

L'intensité du pixel considéré est remplacée par la moyenne des pixels de son voisinage, la taille de la zone (fenêtre) entourant le pixel est un paramètre important, plus cette dimension est grande, plus Sa sensibilité au bruit diminue, et le lissage devient important (le flou s'accroît).

Le filtre moyenneur est un filtre passe-bas c à d qu'il laisse passer les basses fréquences (les faibles changements d'intensité de l'image) et atténue les hautes fréquences (variations rapides). [4]

$$\frac{1}{9} * \begin{array}{|c|c|c|} \hline I & I & I \\ \hline I & I & I \\ \hline I & I & I \\ \hline \end{array}$$

Masque du filtre.

Exemple:



Image originale.



Image filtrée.

Figure 9 : Application du filtre moyen (lissage)

5.1.2. Filtre gaussien

L'expression gaussienne en deux dimensions est donnée par :

$$G_{\sigma}(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right)$$

L'intérêt de ce filtre est que l'on contrôle facilement le degré de filtrage à travers le paramètre σ . Le filtre gaussien donne plus d'importance aux pixels proches du pixel central, et diminue cette importance au fur et à mesure que l'on s'éloigne de celui-ci, il a les mêmes inconvénients que le filtre moyenneur c à d il dégrade les contours. [4]

La discrétisation de ce filtre pour un σ égale à 0.6 donne le masque suivant :

$1/16 \times$	1	2	1
	2	4	2
	1	2	1

Masque du filtre.

5.1.3. Filtre rehausseur de CONTOURS

C'est un filtre passe haut c à d il met en évidence les changements rapides de l'intensité de l'image (les hautes fréquences) et laisse les zones uniformes inchangées (basses fréquences). [4]

-1	-1	-1
-1	9	-1
-1	-1	-1

Masque du filtre.

Exemple:



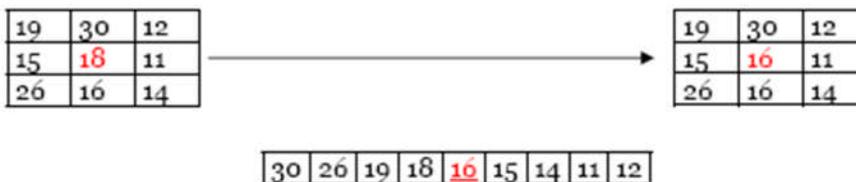
Figure 10 : Rehaussement de contour

5.2. Filtres non linéaires

Ils sont conçus pour régler les problèmes des filtres linéaires, sur tout pour ce qui concerne la mauvaise conservation des contours. Leur principe est le même que celui des filtres linéaires, il s'agit toujours de remplacer la valeur de chaque pixel par la valeur d'une fonction calculée dans son voisinage, la seule différence c'est que cette fonction n'est plus linéaire mais une fonction quelconque (elle peut inclure des opérateurs de comparaisons).

5.2.1. Filtre médian

Sur un voisinage à huit, le nouveau niveau de gris du pixel centre est choisi comme étant la valeur médiane de tous les pixels de la fenêtre d'analyse centrée sur ce dernier. Son avantage est qu'il garde la netteté des éléments qui constituent l'image sans étaler les transitions. [4]



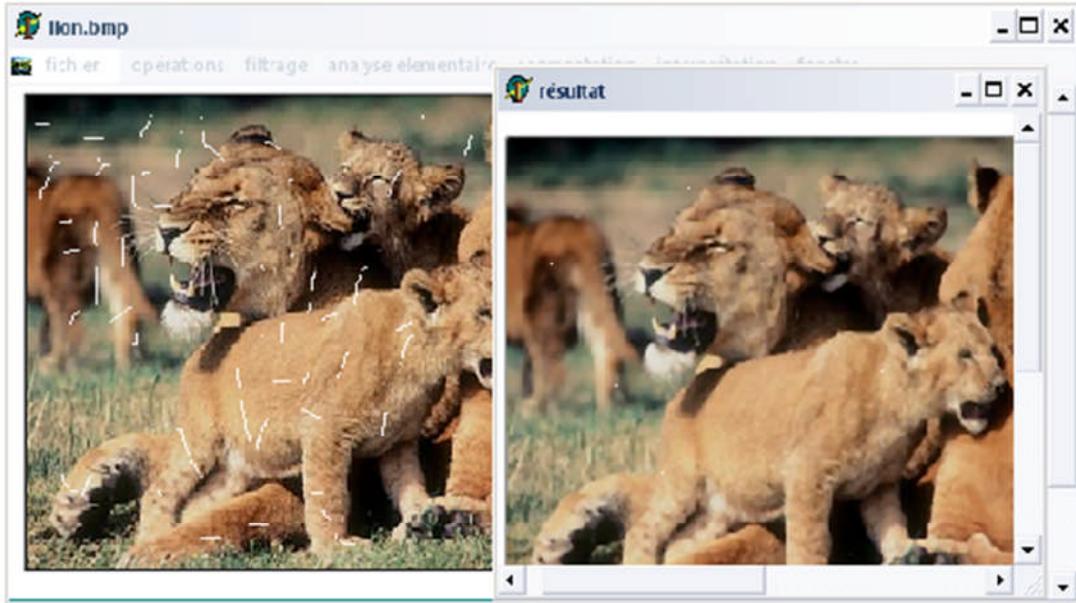


Figure 11 : Principe du filtre médian

5.2.2. Filtre maximum :

Même principe mais la valeur choisie est la valeur maximale.

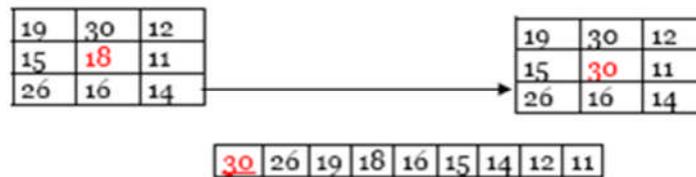


Figure 12 : Principe du filtre maximum

5.2.3. Filtre minimum :

Même chose, sauf que la valeur choisie est la valeur minimale.

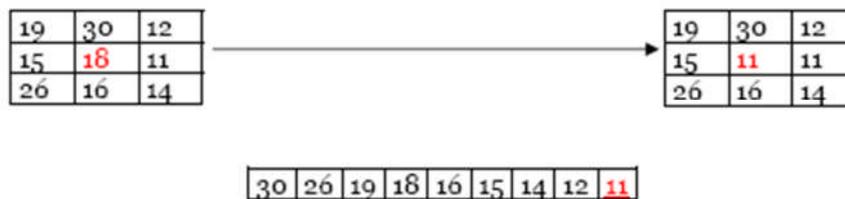


Figure 13 : Principe du filtre minimum

6. Segmentation :

La **segmentation d'image** est une opération de traitement d'images qui a pour but de rassembler des pixels entre eux suivant des critères prédéfinis. Les pixels sont ainsi regroupés en régions, qui constituent un pavage ou une partition de l'image. Il peut s'agir par exemple de

séparer les objets du fond. Si le nombre de classes est égal à deux, elle est appelée aussi binarisation.

Si l'homme sait naturellement séparer des objets dans une image c'est grâce à des connaissances de haut niveau (compréhension des objets et de la scène). Mettre au point des algorithmes de segmentation de haut niveau (chaque région est un objet sémantique) est encore un des thèmes de recherche les plus courants en traitement d'images.

La segmentation est une étape primordiale en traitement d'image. À ce jour, il existe de nombreuses méthodes de segmentation, que l'on peut regrouper en quatre principales classes :

1. Segmentation fondée sur les régions (en anglais : region-based segmentation). On y trouve par exemple : la croissance de région (en anglais : region-growing), décomposition/fusion (en anglais : split and merge).
2. Segmentation fondée sur les contours (en anglais : edge-based segmentation).
3. Segmentation fondée sur classification ou le seuillage des pixels en fonction de leur intensité (en anglais : classification ou thresholding).
4. Segmentation fondée sur la coopération entre les trois premières segmentations.

7. Quelques applications concrètes de traitement d'images

- Contrôle de présence/absence : Sur des chaînes de production, on vérifie en bout de chaîne avec une caméra vidéo la présence d'une pièce dans un ensemble plus complexe. Pour cela, bien souvent il suffit de faire un simple seuillage dans une région spécifique.
- Contrôle du niveau de maturation des fruits sur une chaîne de conditionnement. Il s'agit de reconnaître à la couleur et à la texture du fruit son degré de maturité et donc la catégorie sous laquelle il sera emballé puis vendu.
- Construction et correction de cartes géographiques d'après des images satellites ou des images aériennes. On recalcule d'après des informations topographiques les images reçues, puis on les met sur la carte en correspondance avec les informations trouvées dans l'image : voies de communication, voies et plans d'eau, parcelles agricoles...
- Surveillance et évaluation de la production agricole. Il est possible de déterminer le degré de maturation des cultures, la quantité d'eau nécessaire pour l'irrigation, le rendement moyen... On peut ainsi établir des prévisions à large échelle de la récolte à venir.
- Reconnaissance de l'écriture: La reconnaissance de l'écriture manuscrite progresse de jour en jour. Elle est suffisamment opérationnelle pour que la majorité des adresses, même manuscrites, soient reconnues automatiquement sur le courrier postal.

- Recherche d'image par le contenu: L'objectif de cette technique est de rechercher, parmi une base de données d'images, les images similaires à une image exemple, ou ayant certaines caractéristiques, par exemple rechercher toutes les images comportant un vélo.

8. Conclusion

La représentation des images fixes est un des éléments essentiels des applications multimédias, comme dans la plupart des systèmes de communication.

La manipulation des images pose cependant des problèmes beaucoup plus complexes que celle du texte. En effet, l'image est un objet à deux dimensions, censé représenter un espace à trois dimensions, ce qui a deux conséquences majeures :

- Le volume des données à traiter est beaucoup plus important.
- La structure de ces données est nettement plus complexe.

Il en résulte que la manipulation, le stockage et la représentation de ces données se heurtent à certaines limitations.

Grâce au traitement d'image, ces contraintes sont levées ou contournées. En effet, ce domaine cherche à détecter la présence de certaines formes, certains contours ou certaines textures d'un modèle connu, c'est le problème de la détection.

Un autre aspect de traitement d'image concerne l'analyse et la compréhension de l'image dans le but d'en extraire des informations utiles. Il cherche aussi parfois à comprimer l'image, afin de gagner en vitesse lors de la transmission de l'information, et en capacité de stockage, tout en dégradant le moins possible les images considérées.

Dans un CBIR il est mieux de faire un prétraitement avant de passer à l'indexation des images de base et ensuite la recherche d'image par contenu pour éliminer le bruit présent dans ces images.

Dans le chapitre suivant nous présentons les notions de bases sur les systèmes de recherche par contenu(CBIR).

CHAPITRE 2 : SYSTEMES DE RECHERCHE D'IMAGE PAR CONTENU (CBIR)



1. Introduction.

L'expression « recherche d'images par le contenu » (« Content-Based Image Retrieval, CBIR, en Anglais) remonte aux travaux de Kato en 1992. Son système, ART MUSEUM, permet de retrouver des images d'art par couleurs et contours. Le terme s'est étendu par la suite à tout procédé permettant de rechercher des images selon des descripteurs, pouvant être de type « signal », comme la couleur et la forme, mais également symboliques. Comme le remarquent les auteurs d'un rapport important sur les systèmes de recherche par le contenu [44], retrouver des images indexées manuellement par des mots clefs n'est pas de la recherche par le contenu au sens où le terme est généralement compris, même si ces mots clefs décrivent le contenu effectif de l'image.

Les applications des systèmes de recherche d'images existants (et donc les collections d'images) sont variées. Elles incluent des applications judiciaires : les services de police possèdent de grandes collections d'indices visuels (visages, empreintes) exploitables par des systèmes de recherche d'images.

Les applications militaires, bien que peu connues du grand public, sont sans doute les plus développées [44] : reconnaissance d'engins ennemis via images radars, systèmes de guidage, identification de cibles via images satellites en sont des exemples connus. Le journalisme et la publicité sont également d'excellentes applications. Les agences de journalisme ou de publicité maintiennent en effet de grosses bases d'images afin d'illustrer leurs articles ou supports publicitaires. Cette communauté rassemble le plus grand nombre d'utilisateurs de recherche par le contenu (davantage pour les vidéos) mais l'aide apportée par ces systèmes n'est absolument pas à la hauteur des espoirs initiaux ([44]). D'autres applications incluent: le diagnostic médical, les systèmes d'information géographique, la gestion d'œuvres d'art, les moteurs de recherche d'images sur Internet et la gestion de photos personnelles.

Concevoir un système permettant d'assister des utilisateurs dans leurs tâches de recherche d'images pose des problèmes variés. Dans [44] les difficultés suivantes sont identifiées :

1. Comprendre les besoins des utilisateurs d'images et leurs comportements : de quoi les utilisateurs ont-ils besoin ?
2. Identifier une manière « convenable » de décrire le contenu d'une image. C'est une tâche rendue difficile par la subjectivité intrinsèque aux images.
3. Extraire des « descripteurs » des images brutes.
4. Pouvoir stocker de manière compacte un grand nombre d'images.
5. Comparer requêtes et images stockées de manière à refléter les jugements de similarité humains.
6. Accéder efficacement aux images par leur contenu.

7. Fournir des interfaces utilisables.

Il convient donc d'abord de montrer quelles sont les approches existantes ainsi que leurs limitations. Nous commençons par rappeler les composants d'un système de recherche d'images par le contenu, puis nous présentons une taxonomie des systèmes selon leur niveau d'abstraction en donnant systématiquement des exemples.

2. Composants d'un CBIR.

Nous décrivons brièvement ici les caractéristiques communes à la plupart des étapes : le traitement de la base d'images, les requêtes puis la mise en correspondance et la présentation des résultats. La Figure 14 illustre l'ordonnancement de ces étapes : [26]

Dans un premier temps (2), des descripteurs sont calculés à partir de chaque image de la collection (1), ils peuvent être de type signal ou/et symbolique (le vocabulaire d'indexation). Les données extraites (à présent représentatives du contenu de l'image du point de vue du système) constituent la base d'index (3). Les requêtes de l'utilisateur (4) sont alors transformées afin d'être comparables avec la base d'index (5) ; une mise en correspondance (6) entre la requête transformée et la base d'index permet ensuite de produire le résultat de la requête (7). Il se peut également que le système possède des composantes liées à la personnalisation, comme par exemple l'extraction, le stockage et l'utilisation d'un profil d'utilisateur.

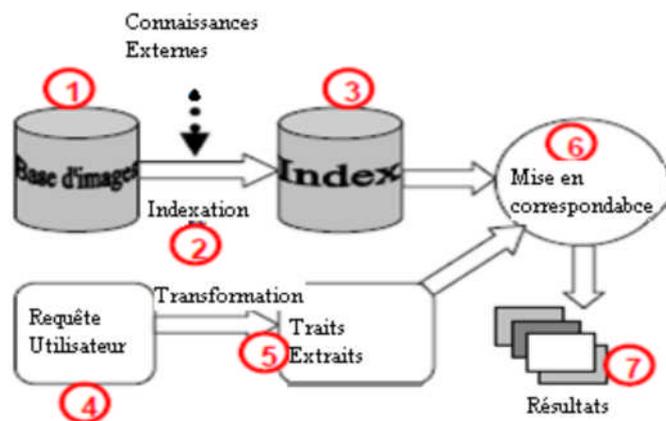


Figure 14 : Principaux composants d'un Système de Recherche par le Contenu

2.1 La base d'image.

La collection (ou base) d'images est la donnée principale du système. Les bases d'images varient d'abord par leur taille : la majorité des systèmes est conçue pour des bases de quelques centaines ou milliers d'images. Ce nombre peut cependant s'approcher du milliard (880 millions d'images) lorsque la base est constituée par les images collectées par des robots sur Internet. La taille de la base d'image impose des contraintes sur la complexité des traitements effectués sur chaque image. Il en résulte que la comparaison qualitative entre des systèmes travaillant sur des bases de tailles très différentes est peu pertinente.

Le type d'image composant la base varie également : des portraits en noir et blanc [28], des peintures chinoises anciennes, des images personnelles [42], des images de tissus humains [10], etc. Le type d'image influe fortement sur la conception globale du système,

particulièrement sur les descripteurs de bas niveau calculés. D'une manière générale, plus la variabilité intra et inter images est importante, plus le système doit être riche et précis (et plus le problème d'indexer/rechercher ces images est difficile).

Finalement, les collections diffèrent par leur stabilité, c'est-à-dire le taux de changements (ajouts d'images, retraits, etc.) en fonction du temps. Faible pour une collection d'images représentant les œuvres d'un peintre ne créant plus, elle peut être très forte lorsque, par exemple, on s'intéresse aux images de la Toile ou à l'actualité.

2.1.1 Exemples de bases d'images existant

Ils existent plusieurs bases d'images, on peut citer les plus utilisées dans le domaine de recherche d'image par le contenu:

2.1.1.1 La base de Wang

La base d'images de Wang est un sous-ensemble de la base d'images Corel.

Cette base d'images contient 1000 images naturelles en couleurs. Ces images ont été divisées en 10 classes, chaque classe contient 100 images. L'avantage de cette base est de pouvoir évaluer les résultats. Cette base d'images a été utilisée pour faire des expériences de classification. Un exemple de chaque classe peut être vu sur la figure 15 [9]. Cette base d'images a été créée par le groupe du professeur Wang de l'université Pennsylvania State et est disponible à l'adresse : <http://wang.ist.psu.edu/>. Chaque image dans cette base d'images a une taille de 384×256 pixels ou 256×384 pixels.



Figure 15 :10 classes de la base de Wang

2.1.1.2 COIL (Columbia Object Image Library)

Cette base d'images est très connue pour la reconnaissance des objets. Il y a deux bases d'images COIL : COIL-20 qui contient des images en niveaux de gris prises à partir de 20 objets différents et COIL-100 qui contient des images en couleurs prises à partir de 100 objets différents. Les deux bases d'images consistent en des images prises à partir des objets 3D avec des positions différentes. La base COIL-100 a 7200 images en couleurs (100 objets x 72 images/objet). Chaque image a une taille de 128×128 pixels. Tous les objets utilisés dans COIL-100 sont montrés dans la figure 18 [9]. La base COIL-20 a 1440 images en niveaux de gris (20 objets x 72 images/objet). Chaque image a une taille 128×128 pixels. Tous les objets

utilisés dans COIL-20 sont dépeints sur la figure 16 [9] . Ces bases d'images sont disponibles à l'adresse :

<http://www1.cs.columbia.edu/CAVE/research/softlib/>



Figure 16 : Les objets utilisés dans COIL-100

2.1.1.3 Pollen

La base de 347 images de grains de pollen (31 classes au total) de l'INRIA Sophia-Antipolis et provenant du projet Européen ASTHMA1. Quelques exemples de cette base d'images sont montrés dans la figure 17.

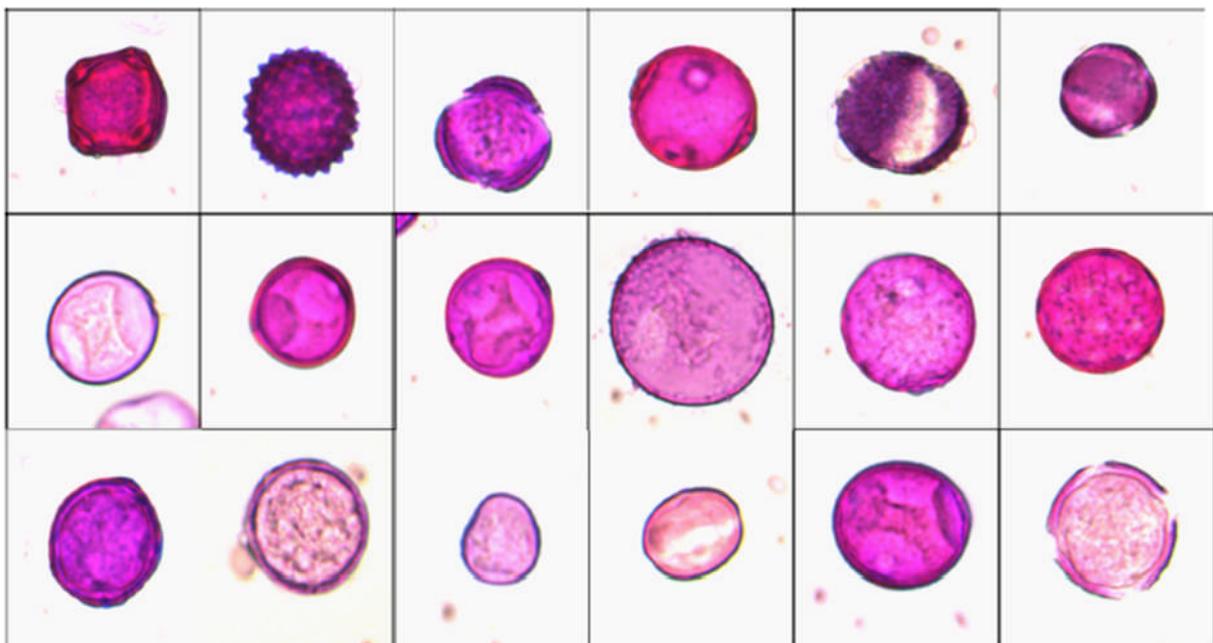


Figure 17 : Quelques images exemples de la base de Pollens

2.1.1.4 CURET (Columbia Utrecht Reflectance and Texture Database)

Cette base a plus de 14000 images de textures (70 textures x ~200 images/texture) des Universités Columbia et d'Utrecht (CURET). Ces images sont disponibles à l'adresse : <http://www1.cs.columbia.edu/CAVE/curet>.

Quelques exemples de cette base d'images sont montrés dans la figure 18.

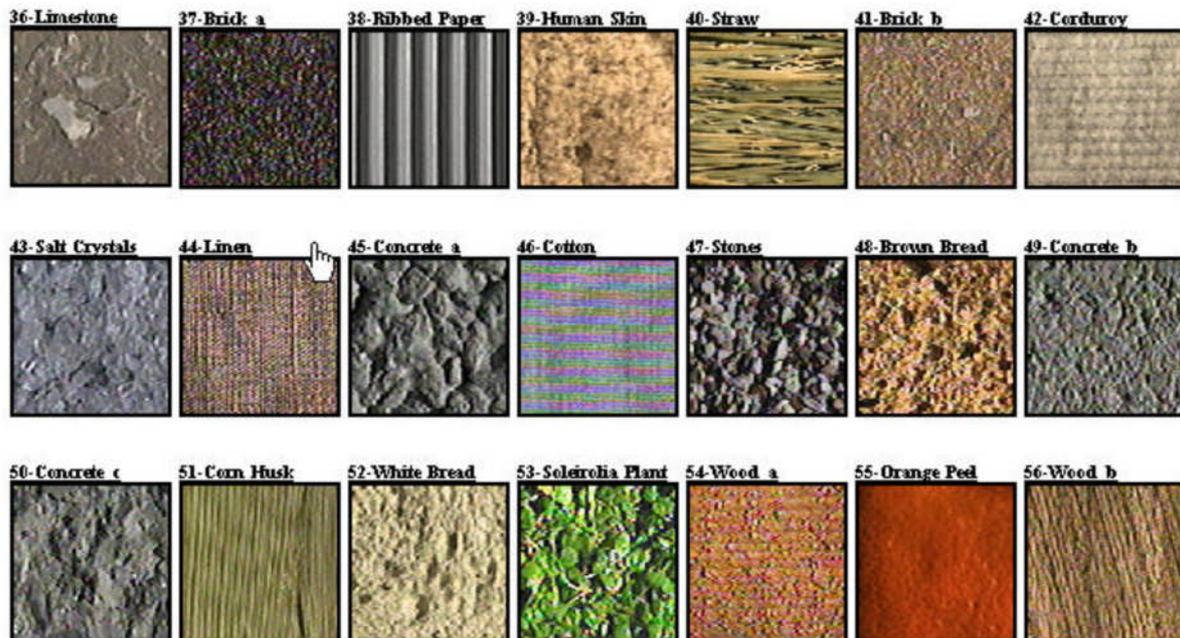


Figure 18 : Quelques images exemples dans la base de CURET(CURET)

2.1.1.5 La base de FeiFei

Cette base contient des images de 101 objets collectées par Fei-Fei Li, Marco Adreetto et Marc Aurolio Ranzato. Avec chaque objet, de 40 à 800 images ont été prises. Chaque image a une taille de 300×200 pixels. Ces images sont disponibles à l'adresse :

<http://www.vision.caltech.edu/feifeili/Datasets.htm>. La figure 19 montre quelques exemples de cette base :



Figure 19 : Quelques images exemples dans la base de Fei-Fei

2.2 L'indexation.

L'indexation est l'ensemble des processus aboutissant à la construction d'un index de l'image. Contrairement à d'autres types de données, comme le texte, il n'est pas utile d'utiliser les images directement dans un CBIR vue la taille des images actuelles. Il faut caractériser les images par des informations à la fois discriminantes et invariables à certains paramètres (comme la taille de l'image, l'angle de la prise de vue, etc.). L'indexation peut être fixe: les descripteurs calculés sont toujours les mêmes. L'indexation peut aussi être évolutive [33] : les

descripteurs s'adaptent à l'utilisateur ou au contexte dans le temps, ce qui permet de renforcer l'adéquation système/utilisateur.

L'indexation peut être générique (indexation de photographies diverses dans [11]), pouvant caractériser des collections hétérogènes, ou spécifique (indexation de peintures chinoises dans [29]), adaptée à un type d'image particulier. Une collection hétérogène est par exemple constituée de photographies personnelles, mettant en scène diverses entités physiques dans des conditions de prise de vue variables. Indexer une telle collection impose l'usage de descripteurs suffisamment génériques (la couleur par exemple), c'est-à-dire qui caractérisent une propriété discriminante applicable à la plupart des entités physiques. A l'inverse, indexer une collection d'images très spécifiques (des empreintes digitales par exemple) requiert l'utilisation de descripteurs également très spécifiques qui, par ailleurs, ne conviendraient probablement pas à une collection hétérogène.

La phase d'indexation peut inclure une étape de segmentation, afin de caractériser des régions homogènes de l'image [40] ou bien indexer l'image dans sa globalité [42]. La segmentation de l'image précède généralement l'indexation individuelle des régions de l'image et cela permet, outre le fait d'accéder à des parties de l'image, de calculer des descripteurs de « forme ».

Enfin, l'indexation varie d'un système à l'autre par son niveau d'abstraction : extraire des histogrammes de couleurs est une opération directe, alors que reconnaître des personnes ou des objets est beaucoup plus complexe et requiert un apprentissage préalable.

2.3 La gestion des index.

Elle concerne la manière dont sont gérés les index des images : stockage et accès. La gestion des index, anecdotique pour une collection de taille modeste, devient une préoccupation essentielle lorsque l'on travaille sur une base de taille conséquente. La manière la plus basique de stocker les index est la liste séquentielle, que ce soit en mémoire ou dans un fichier. Cependant, lorsque le nombre d'images augmente, le temps d'accès à une image augmente linéairement et il est souvent nécessaire d'organiser les index de manière hiérarchique, sous forme d'arbres (organisés selon les descripteurs), ou de tables de « hash-code » par exemple, afin d'accélérer l'accès à l'information.

2.4 Les requêtes.

Le type de requête proposé découle de choix fait en amont, au niveau de l'indexation. Dans des systèmes où seuls des descripteurs de bas niveau sont extraits, les requêtes ne peuvent être que de bas niveau : requête par « image exemple », par croquis ou par manipulation directe des traits de bas niveau. Dans ces systèmes, des descripteurs sont extraits à partir de la requête (une image, un croquis...) et sont comparés aux descripteurs calculés à partir des images de la base (les index des images).

A l'opposé, dans des systèmes proposant plus d'abstraction, les requêtes peuvent être sémantiques (textuelles par exemple). Par exemple dans [46], les images sont indexées par des

« catégories sémantiques visuelles », ce qui permet à un utilisateur de formuler des requêtes sémantiques (« Je veux des images prises à l'extérieur. »).

2.4.1 Les types des requêtes:

2.4.1.1 Requête par l'exemple.

Dans ce cas le système a besoin de comparer un exemple de même type (image) avec la base pour produire les documents similaires. Cette méthode est simple naturelle et ne nécessite pas de connaissance approfondies pour manipuler le système. Elle est donc bien adaptée à un utilisateur non spécialiste.

2.4.1.2 Requête par crayonnage (Sketch).

L'utilisateur spécifie un exemple par un dessin pour lequel il cherche des documents similaires. Outre la forme des contours le dessin peut inclure la couleur et la texture. Cette modalité entraîne plus d'interaction avec l'utilisateur.

2.4.1.3 Requête par caractéristique.

L'utilisateur indique la ou les caractéristiques qu'il veut utiliser pour trouver les images similaires, par exemple trouver les images contenant 25% de rouge et 30% de jaune. Ces caractéristiques sont répertoriées dans un vocabulaire compilé en outils de traitement.

2.4.1.4 Requête exemple et texte.

Cette méthode consiste à renforcer l'image requête en lui associant du texte afin d'accéder à un niveau d'abstraction supérieur. Les images sont organisées et indexés en groupes de pertinence. Chaque groupe définit un type ou un domaine dont l'image relève, par exemple végétation ou animale, etc. Cette approche permet une sémantique accrue mais elle est plus biaisée voire moins générale que les précédentes en ce que la désignation des groupes de pertinence est une tâche à fort caractère subjectif ou expert.

2.5 Analyse de la requête.

Cette étape a pour but de transformer la requête utilisateur pour la rendre comparable avec les index de la base d'images ; elle consiste donc généralement à extraire les mêmes types de descripteurs que ceux extraits de la base d'image lors de l'indexation.

2.6 Mise en correspondance requête / base.

Il s'agit d'estimer dans quelle mesure une image (son index) satisfait une requête donnée. Dans le contexte de la recherche d'images, cela se ramène souvent à calculer la similarité entre les caractéristiques extraites de la requête et les caractéristiques de chaque image dans la base. Cela aboutit généralement à une valeur de correspondance qui caractérise la pertinence (du point de vue du système) d'une image par rapport à la requête. Cette mise en correspondance peut être simple (comparaison d'histogrammes) ou complexe (comme dans [40] par exemple, avec une mise en correspondance qui tient compte de l'arrangement spatial des régions).

La phase de mise en correspondance peut également inclure une pondération des descripteurs (comme dans [11] où chaque descripteur est pondéré par rapport à son pouvoir discriminant

dans la base). Pondérer les descripteurs permet d'éliminer une partie du bruit dans la mesure où les descripteurs les moins pertinents voient leur influence diminuer dans l'évaluation de la similarité requête/image.

La mise en correspondance peut également inclure un bouclage de pertinence. Le but est également d'éliminer le bruit (augmenter la précision) en tentant de converger vers une précision maximale.

2.7 La présentation des résultats.

Dans la grande majorité des systèmes disponibles [45], le résultat d'une requête est présenté sous la forme d'une liste d'images (réduites à des vignettes) ordonnées par pertinence décroissante. Parfois cette présentation prend d'autres formes, comme par exemple l'œil de poisson (FishEye View) [16]. L'avantage des images par rapport aux documents textuels est qu'il est possible de visionner d'un coup d'œil l'intégralité du document, ce qui permet de visualiser un grand nombre de résultats et de les comparer plus rapidement. Comme indiqué plus haut, la présentation des résultats est souvent couplée avec une possibilité d'interaction, qui permet par exemple de raffiner une requête en indiquant au système les résultats pertinents et ceux qui ne le sont pas (bouclage de pertinence), et de permettre ainsi une reformulation automatique de la requête.

3. Représentation des images dans un CBIR.

Dans la majeure partie des systèmes existant, les images sont représentées avec des descripteurs de bas niveau, i.e., en termes de couleur, texture, formes (voir le chapitre suivant) ou par des descripteurs de haut niveaux.

4. Mesures pour évaluer un système

Avant l'exécution d'un système de recherche d'informations, une évaluation qui permet de mesurer la performance de ce système est nécessaire. Les mesures les plus courantes pour évaluer un système sont le temps de réponse et l'espace utilisé. Plus le temps de réponse est court, plus l'espace utilisé est petit, et plus le système est considéré bon. Mais avec des systèmes qui ont été faits pour la recherche d'informations, en plus de ces deux mesures, on s'intéresse à d'autres mesures. Dans le système de recherche d'informations, l'utilisateur s'intéresse aux réponses pertinentes du système. Donc les systèmes de recherche d'informations exigent l'évaluation de la précision de la réponse. Ce type d'évaluation est considéré comme l'évaluation des performances de recherche. Le système d'indexation et de recherche d'images est un système de recherche d'informations. Dans les systèmes de recherche d'images, les auteurs ont souvent utilisé les mesures d'évaluation pour évaluer des systèmes de recherche d'informations.

Dans cette section, nous allons décrire les deux mesures les plus courantes: le rappel et la précision. Ces mesures sont reliées entre elles. Donc on décrit souvent cette relation par une courbe de rappel et précision. Ensuite nous présentons d'autres mesures que l'on utilise aussi pour évaluer des systèmes de recherche d'informations.

4.1. Rappel et précision (en anglais : Recall and Precision)

2 valeurs pour évaluer le système de recherche par contenu :

4.1.1. Le rappel:

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.

$$\text{Rappel} = \frac{|Ra|}{|R|}$$

4.1.2. La précision :

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

$$\text{Précision} = \frac{|Ra|}{|A|}$$

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.

Des définitions sont montrées dans la figure 20 :

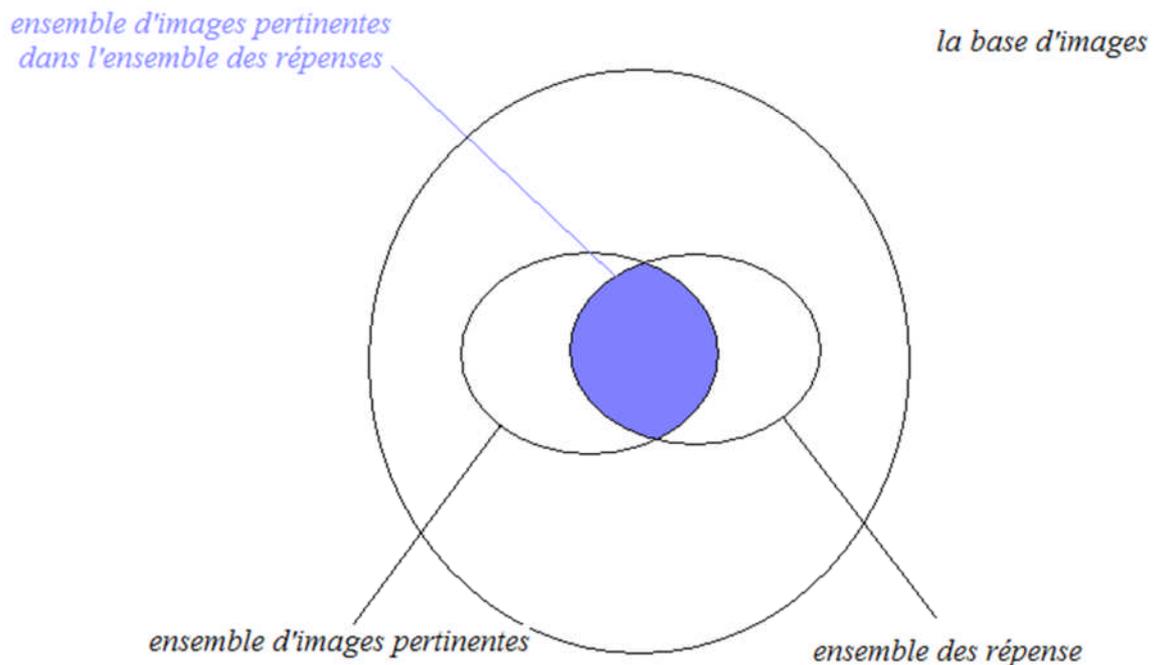


Figure 20 : Le rappel et la précision pour une requête

Dans les systèmes de recherche d'informations, afin de définir si une information est pertinente ou non, on a besoin d'experts dans le domaine. Dans les systèmes de recherche d'images, une image est pertinente pour une requête si les deux images sont dans la même classe. C'est pourquoi dans l'étape de préparation de la base d'images pour évaluer, on doit faire des annotations. L'annotation est un processus qui permet aux utilisateurs de choisir des mots clés correspondants à chaque image. Après l'annotation, on va classer les images en classes appropriées. Si des images ne contiennent pas beaucoup d'objets, c'est facile de les classer dans ces classes. Mais si les images contiennent beaucoup d'objets, la tâche de classification devient de plus en plus difficile. Dans ce cas là, chaque image appartient à plusieurs classes.

4.2. La courbe de rappel et précision

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. La figure 21 donne un exemple de courbe de rappel et précision. Pour dessiner cette courbe, on doit calculer plusieurs paires de rappel et précision et les interpoler.

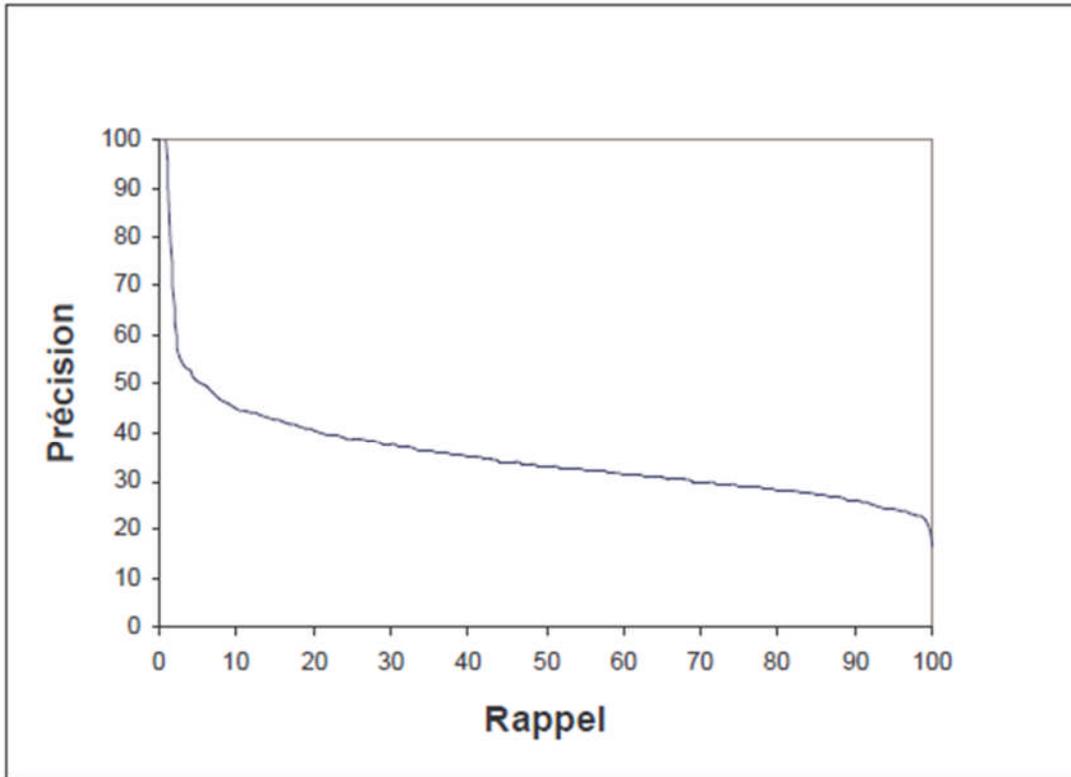


Figure 21 : Une courbe de rappel et précision

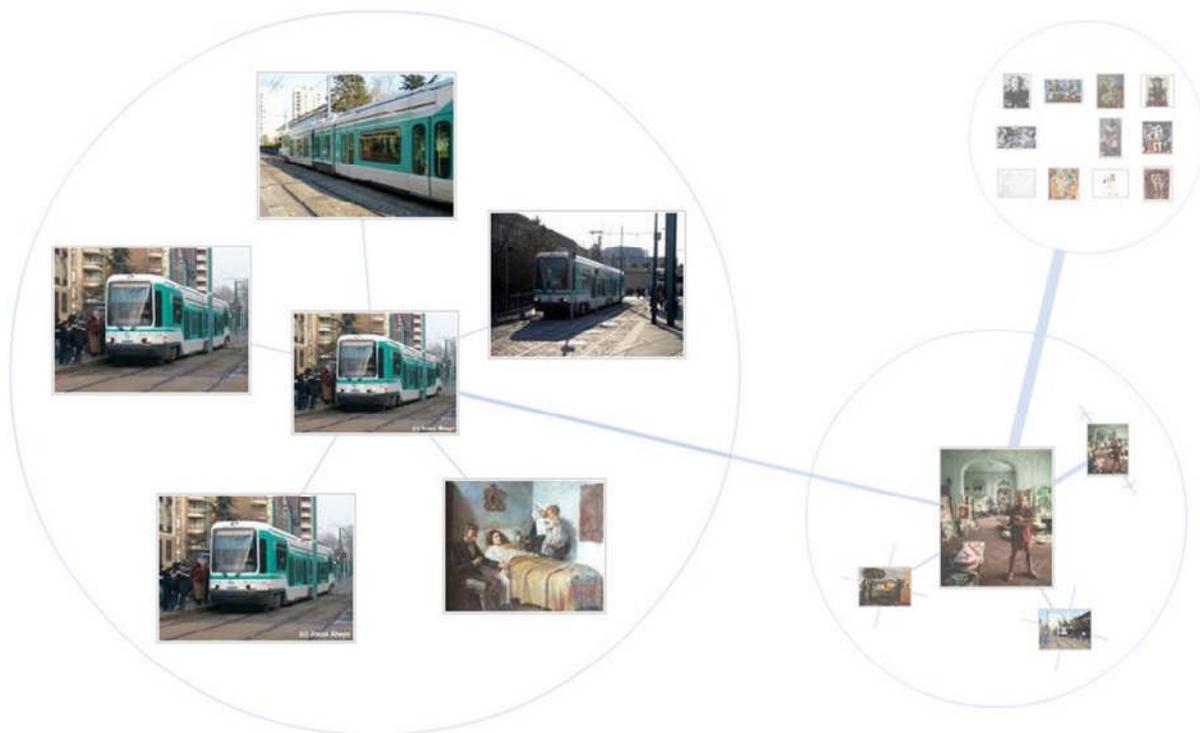
5. Conclusion

Dans ce chapitre nous avons essayé de décrire les concepts de base pour la construction d'un système de recherche d'images par le contenu. Nous avons abordé dans un premier lieu les différents composants d'un CBIR. Par la suite nous avons parlé Représentation des images dans un CBIR, les mesures pour évaluer un CBIR, difficulté de l'indexation des images.

À la fin de ce chapitre, nous avons présenté les domaines d'application de la recherche d'images par le contenu. Comme nous pouvons le constater, la recherche d'images par le contenu, s'est imposée dans tous les domaines de notre vie quotidienne. Par conséquent elle attire beaucoup d'attention et devient un axe de recherche très actif avec tous ses aspects.

Dans le prochain chapitre, nous présenterons les différents descripteurs extraient a partir d'une image (de couleurs, de texture et de forme) et les mesures de similarité entre ces descripteurs.

CHAPITRE 3 : DESCRIPTEURS D'IMAGES & MESURES DE SIMILARITÉ



1. Introduction

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. En effet l'image et la vidéo numérique sont partie intégrante de tels systèmes par la densité et la richesse de leur contenu. La même image peut présenter plusieurs significations à différents niveaux : analyse, description, reconnaissance et interprétation.

La recherche d'information couvre le traitement de documents numériques impliquant la structure, l'analyse, le stockage et l'organisation des données. Dans le passé, le terme recherche d'information était lié au concept de l'information textuelle. Actuellement « RI » est associé à tout type d'information, textuelle, visuelle ou autre. Cependant dû aux limitations des méthodes textuelles, le développement des méthodes basées sur le contenu visuel est devenu primordial. Ceci explique l'activité de recherche intense consacrée au système CBIR ces dernières années. Le « RIC » est souvent confronté au problème de pertinence de la recherche, et au temps de recherche.

L'objectif de n'importe quel système CBIR est de satisfaire la requête d'un utilisateur par la pertinence des résultats. Comme l'accès à un document via sa pure sémantique est impossible, les systèmes CBIR traditionnels s'appuient sur un paradigme de représentation de bas niveau du contenu de l'image, par la couleur, la texture, la forme, etc..., et d'autres par une combinaison de celles-ci. La recherche d'images se fait ainsi par comparaison des descripteurs.

L'analyse et la représentation du contenu des données sources mises sous forme de vecteur caractéristique. L'information obtenue dans cette étape est une sorte de résumé des images de la base (segmentation en régions, couleur, texture, relations spatiales,...). La transformation est généralement gourmande en temps de calcul.

Dans la suite de ce chapitre, nous présentons les différents attributs utilisés dans les systèmes de recherche d'image par contenu et ensuite les mesures de similarité entre les images après la définition de leurs descripteurs.

2. Descripteurs d'image

2.1. Descripteurs de couleur

La couleur est l'information visuelle la plus utilisée dans les systèmes de recherche par le contenu. Ces valeurs tridimensionnelles font que son potentiel discriminatoire soit supérieur à la valeur en niveaux de gris des images. Avant de sélectionner le descripteur de couleur approprié, la couleur doit être déterminée d'abord par son modèle (voir le chapitre 1).

2.1.1. *L'histogramme*

Une technique très utilisée pour la couleur est l'intersection d'histogrammes [48]. 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. Premièrement, 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. Deuxièmement, ils ne possèdent pas d'informations spatiales sur les positions des couleurs. Troisièmement, ils sont

sensibles à de petits changements de luminosité, ce qui est problématique pour comparer des images similaires, mais acquises dans des conditions différentes. Et quatrièmement, ils sont inutilisables pour la comparaison partielle des images (objet particulier dans une image), puisque calculés globalement sur toute l'image.

2.1.2. Les moments de couleur

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 :

$$\mu_i = \frac{1}{N} \sum_{j=1}^N f_{ij}$$

$$\sigma_i = \left(\frac{1}{N} \sum_{j=1}^N (f_{ij} - \mu_i)^2 \right)^{1/2}$$

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

Où f_{ij} est la valeur de la i ème composante chromatique du pixel j , et N le nombre de pixels de l'image.

Les moments de couleur est une représentation compacte comparée aux autres descripteurs de couleur. Car seulement 9 valeurs (3 pour chaque composante chromatique) sont utilisées pour représenter le contenu d'une image. Pour cette raison ils peuvent diminuer le pouvoir de discrimination (description). [27]

2.1.3. 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 :

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

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

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

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

$$SCR(c) = \frac{H_1(c)}{H(c)}$$

Où H représente l'histogramme 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. [26]

2.1.4. Couleurs dominantes.

L'utilisation d'histogrammes pour représenter la distribution de couleur présente quelques inconvénients. Du point de vue de l'espace mémoire, les histogrammes à plusieurs dimensions sont ((creux)), c'est-à-dire que la majorité des cellules ne comptent aucun pixel. Une grande partie de l'espace mémoire est utilisée inutilement. De plus, toutes les classes ont la même taille, alors qu'il serait plus intéressant d'avoir des classes plus petites dans les régions contenant des couleurs très fréquentes, et de grandes classes pour les couleurs moins répandues. Du point de vue des mesures de similarité employées, les mesures traditionnelles effectuent uniquement une comparaison cellule à cellule. Même si les histogrammes sont ordonnés, le voisinage des cellules n'est pas pris en compte quand elles ont des valeurs différentes.

Les signatures par couleurs dominantes, proposées dans [47], permettent de résoudre ces différents problèmes. La signature $s = \{s_i = (m_i, w_i)\}$ est un ensemble de nuages de points. Chaque nuage est représenté par son mode m_i (le mode d'un nuage de point correspond à un maximum local de sa densité de probabilité), et le nombre w_i de pixels qui appartiennent au nuage.

Contrairement aux histogrammes, ces signatures ne stockent que les couleurs qui appartiennent à l'image, elles ne stockent pas les cellules vides.

La norme MPEG-7 définit un descripteur pour les couleurs dominantes, appelée DCD (Dominant Color Descriptor). D'après [35], ce descripteur est défini par :

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

Où N est le nombre de couleurs dominantes (inférieur ou égal à huit). Le terme c_i est un vecteur qui représente la $i^{\text{ème}}$ couleur dominante, attribuée à un pourcentage p_i de pixels dans l'image (ou dans la région de l'image). Le terme v_i représente un paramètre optionnel, la variance des couleurs des pixels qui sont associés à la couleur dominante c_i . Le terme s est une valeur qui représente l'homogénéité spatiale des couleurs dominantes dans l'image (en termes de connexité de régions). Le nombre d'occurrences de couleurs dominantes est

variable selon les spécifications MPEG-7. Etant donné que le calcul de distance est assez complexe et pour des raisons de normalisation des calculs dans notre implémentation qui vise justement à réduire le temps de calcul, ce descripteur est peu intéressant dans un contexte d'indexation temps réel.

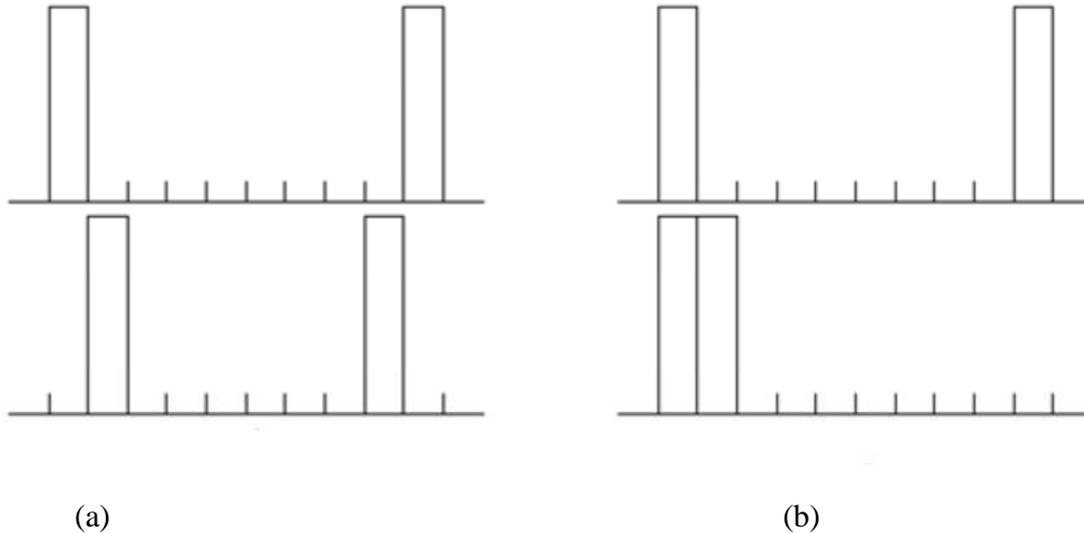


Figure 22 : Comparaison entre histogrammes

Avec une comparaison cellule à cellule, les deux histogrammes dans la situation (a) ont une intersection nulle, alors qu'ils sont très proches. Dans la situation (b), les deux histogrammes ont une intersection de 50% alors qu'ils ne sont pas visuellement plus proches que dans le cas (a), lorsque les cellules sont ordonnées par proximité de couleurs.

La mesure de similarité la plus utilisée pour comparer ces signatures est l'Earth Mover Distance, ou EMD. La distance entre deux distributions S_1 et S_2 est donnée par le coût minimum de travail nécessaire pour transformer S_1 en S_2 . L'EMD peut être définie comme la solution d'un problème de transport qui peut être résolu par une optimisation linéaire.

Si on note d_{ij} la distance entre le mode m_i de s_1 et le mode m_j de s_2 , et f_{ij} le point optimal entre les deux distributions, la solution de ce problème est donnée par :

$$EMD(s_1, s_2) = \frac{\sum_{i=1}^m \sum_{j=1}^m d_{ij} f_{ij}}{\sum_{i=1}^m \sum_{j=1}^m f_{ij}}$$

Où m et n représentent respectivement le nombre de classes dans s_1 et s_2 . Pour avoir plus de détails sur les problèmes d'optimisation de l'EMD, le lecteur peut se référer à [47].

Il existe plusieurs approches pour trouver les différents nuages de points. La méthode la plus simple consiste, à partir d'un histogramme, à prendre toutes les classes qui ont un effectif non nul.

Cette approche a l'inconvénient de fournir un nombre assez élevée de classes. Une autre approche consiste à ne garder que les couleurs dominantes, et à remplacer chaque couleur de l'image par la couleur conservée la plus proche. Il existe plusieurs algorithmes pour détecter les couleurs dominantes : citons notamment la segmentation d'image, qui consiste à regrouper tous les pixels ayant une couleur proche, ou encore l'extracteur de couleurs dominantes proposée pour le descripteur de MPEG-7 par [8]. Ce dernier effectue une classification des couleurs de l'image dans un espace perceptuellement uniforme (habituellement le système $L^*u^*v^*$), à l'aide de l'algorithme de Lloyd généralisé.

Même si l'EMD présente de bons résultats en comparaison d'autres mesures de similarité [1], son utilisation est limitée par la complexité de calcul. Même s'il existe des implémentations optimisées assez rapides, notamment dans la bibliothèque Open CV [Open CV], l'exécution demeure toujours plus lente qu'avec les distances traditionnelles appliquées sur les histogrammes. [26]

Conclusion

L'information relative aux couleurs est particulièrement importante dans la caractérisation d'une image. Plusieurs études ont été menées pour trouver un critère de choix des descripteurs de couleurs pour l'indexation des images, mais aucune n'a abouti. Ceci peut s'expliquer par le manque de subjectivité de cette information, les descripteurs couleur ne suffisent pas à indexer efficacement une image, ni à la chercher.

2.2. Descripteurs des 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.

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.

D'après [7], une définition formelle de la texture est quasiment impossible.

D'une manière générale, la texture se traduit par un arrangement spatial des pixels que l'intensité ou la couleur seules ne suffisent pas à décrire. Elles peuvent consister en un placement structuré d'éléments mais peuvent aussi n'avoir aucun élément répétitif.

De nombreuses définitions ont été proposées, mais aucune ne convient parfaitement aux différents types de textures rencontrées. Dans une définition couramment citée [36], la texture est présentée comme une structure disposant de certaines propriétés spatiales homogènes et invariantes par translation. Cette définition stipule que la texture donne la même impression à l'observateur quelle que soit la position spatiale de la fenêtre à travers laquelle il observe cette texture. Par contre l'échelle d'observation doit être précisée. On peut le faire par exemple en précisant la taille de la fenêtre d'observation.

La notion de texture est liée à trois concepts principaux:

- 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: les textures structurées (macrotextures) et les textures aléatoires (microtextures).

Une texture qualifiée de structurée est constituée par la répétition d'une primitive à intervalle régulier. On peut différencier dans cette classe les textures parfaitement périodiques (carrelage, damier, etc.), les textures dont la primitives subit des déformations ou des changements d'orientation (mur de briques, grains de café, etc.).

Les textures qualifiées d'aléatoires se distinguent en général par un aspect plus fin (sable, herbe, etc.). Contrairement aux textures de type structurel, les textures aléatoires ne comportent ni primitive isolable, ni fréquence de répétition. On ne peut donc pas extraire de ces textures une primitive qui se répète dans l'image mais plutôt un vecteur de paramètres statistiques homogènes à chaque texture.

Dans tous les cas, ces objectifs nécessitent l'extraction d'un ou de plusieurs paramètres caractéristiques de cette texture. Nous désignerons ces paramètres sous le terme d'attributs texturaux (textural features) et l'ensemble qu'ils constituent sous le terme de descripteur de texture.

Certains de ces paramètres correspondent à une propriété visuelle de la texture (comme la directionnalité ou la rugosité). D'autres correspondent à des propriétés purement mathématiques auxquelles il est difficile d'associer une qualification perceptive.

Un recensement ainsi qu'une classification des termes de description des textures employés par les principaux auteurs pourront être trouvés dans [38] et [39].

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 [37] [15], les intégrales curvilignes [3], l'application de filtres [12] [5] [24] [43] ou sur la théorie fractale [Sar92] [Fio95] peuvent également être employées pour obtenir des attributs texturaux.



Figure 23 : Des textures différentes

2.2.1. Les matrices de co-occurrences

En 1973, Haralick [18] a proposé une méthode en se basant sur les matrices de co occurrences 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 co-occurrences contient les fréquences spatiales relatives d'apparition des niveaux de gris selon quatre directions

$$(\theta = 0, \theta = \pi/4, \theta = \pi/2, \theta = 3\pi/4).$$

Une matrice de co-occurrences est définie au moyen d'une relation géométrique π entre deux pixels (x_1, y_1) et (x_2, y_2) .

La matrice de co-occurrences $P_{d,\theta}(i, j)$, q 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 co occurrences sont donc les niveaux de gris de la texture étudiée.

On définit la matrice de co-occurrences $P_{d,\theta}$ 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 a présenté 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.

2.2.2. Transformée en ondelettes

La transformée en ondelettes est à la base de nombreuses analyses de texture, telles que les filtres de Haar [25]. La description de texture à base d'ondelettes est utilisée pour la recherche d'images.

Pour avoir plus d'information sur les fondements mathématiques de la transformée en ondelettes, le lecteur peut se référer au livre [42]. Comme pour la transformée de Fourier, une présentation plus pédagogique et plus historique des ondelettes peut également être trouvée dans [23].

L'approche continue des ondelettes pour un signal 2D est trop complexe pour être applicable rapidement sur des images. Pour résoudre ce problème, Mallat [32] 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 (Discrete Wavelet Transform) 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.

Une des transformées en ondelettes les plus couramment employées en analyse d'images est la transformée de Haar, mais d'autres ondelettes sont aussi largement exploitées [42]. Les filtres de Haar sont fréquemment employés en apprentissage pour obtenir la description d'un objet (comme un visage ou une personne).

Conclusion

Les attributs texturaux sont des attributs très importants pour la description de l'image et la reconnaissance des objets, cependant elles ne suffisent pas pour une bonne représentation du contenu de l'image, un autre attribut essentiel est la forme.

Dans la suite nous allons introduire cet attribut et les différentes approches utilisées pour l'extraire.

2.3. Descripteurs de Formes

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

Nous présentons dans ce qui suit la méthode utilisé permettant de reconnaître une forme donnée dans une image. Itérée pour toutes les formes d'une image, cette méthode permet finalement de relever toutes les formes communes à deux images.

2.3.1. Les attributs géométriques de région

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 \mathcal{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 :

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

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

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

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

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

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.

2.3.2. Les moments géométriques

Les moments géométriques [SHB99] 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 :

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

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 :

$$m_{2,0}^s = m_{2,0} - m_{0,0}x_c^2 \quad m_{1,1}^s = m_{1,1} - m_{0,0}x_c y_c \quad m_{0,2}^s = m_{0,2} - m_{0,0}y_c^2$$

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

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

À partir des moments géométriques, Hu [22] a introduit sept invariants aux translations, rotations et changement d'échelle, appelés moments de Hu.

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

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

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

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

$$M_5 = (\mu_{30} - 3\mu_{12})(\mu_{30} + \mu_{12}) \left[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2 \right] + \\ (3\mu_{21} - \mu_{03})(\mu_{21} + \mu_{03}) \left[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2 \right].$$

$$M_6 = (\mu_{20} - \mu_{02}) \left[(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2 \right] + 4\mu_{11}(\mu_{30} + \mu_{12})(\mu_{03} + \mu_{21}).$$

$$M_7 = (3\mu_{21} - \mu_{03})(\mu_{30} + \mu_{12}) \left[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2 \right] - \\ (\mu_{30} - 3\mu_{21})(\mu_{12} + \mu_{03}) \left[3(\mu_{30} + \mu_{12})^2 - (\mu_{12} + \mu_{03})^2 \right].$$

Les moments de Hu offrent d'excellents attributs 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.

2.3.3. Transformée de Hough

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

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

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

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

$$F = \left\{ \left\{ x : f(x, a) = 0, x \in \mathfrak{R}^n \right\}, a \in \Omega \right\}.$$

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

Conclusion

Les formes représentent un descripteur puissant pour décrire les objets contenu dans l'image.

3. 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. Au lieu d'un appariement exact, la recherche d'images par le contenu calcule des similarités visuelles entre une image requête et les images de la base d'images. En conséquent, le résultat d'une recherche n'est pas une seule image mais une liste d'images ordonnées selon leur degré de similitude avec l'image requête. Plusieurs mesures de similarités ont été proposées dans la littérature. Les différentes mesures de similarité influencent les performances de recherche des systèmes de recherche par le contenu.

Plusieurs mesures de similarité sont basées sur la distance L_p entre deux points. pour deux points donnés (x,y) dans \mathbb{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)

Au plus bas niveau d'abstraction, les images sont tout simplement des agrégations de pixels. La comparaison entre les images, est réalisée pixel par pixel, et les mesures de similarité couramment utilisées comprennent : le coefficient de corrélation, la somme des valeurs absolue des différences (SVAD), la distance des moindres carrés, et l'information mutuelle. La comparaison au niveau des pixels est très spécifique et, par conséquent, n'est utilisée que lorsque des appariements relativement précis sont nécessaires.

B. Attributs visuels

Les attributs visuels sont des valeurs numériques de données extraites des images ou des objets dans les images, tels que la couleur, la forme et la texture. Plusieurs mesures de similarité sont couramment utilisées pour la comparaison d'attributs: la distance euclidienne, la distance de Minkowsky et la distance d'intersection.

[27]

3.1. Les méthodes de calcul

Ci-après les distances les plus utilisées pour comparer des images considérées comme vecteurs ou comme distributions statistiques. [26]

3.1.1. Distance de Mahalanobis

Cette distance prend en compte la corrélation entre les distributions des classes. Elle est ainsi définie par

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

sous la forme : $D_M = \sqrt{(f_1 - f_2)^T C^{-1} (f_1 - f_2)}$.

Où C'est la matrice de covariance. Dans les cas où les dimensions des caractéristiques sont indépendantes, C ne comporte que des les 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.

3.1.2. Intersection d'histogrammes

Cette mesure est l'une des premières distances utilisée 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 .

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

Deux images présentant une intersection normalisée d'histogrammes proche de 1 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 [49] .

3.1.3. Earth Mover Distance (EMD)

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 dissimilarité 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 :

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

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

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

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

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

EMD prétend également mimer la vision humaine.

3.1.4. Distance de Minkowski

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

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

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 ($p=1$).

3.1.5. Distance quadratique

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 donné par : $D_Q = \sqrt{(f_1 - f_2)^T A (f_1 - f_2)}$ où $A = [a_{ij}]$ est la matrice de similarité. Représente la distance entre deux éléments des vecteurs f_1 et f_2 . Hafner et al [HSE95] 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.

3.1.6. Distance de Bhattacharya

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.

3.1.7. Distance de Kullback Leiber (KL)

La divergence de Kullback Leiber exprime l'entropie relative de deux distributions :

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

3.1.8. Divergence de Jeffrey (JD)

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

3.1.9. Distance de Kolmogorov Smirnov

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

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

3.1.10. Distance de Cramer Von Mises :

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

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

4. Conclusion

Le choix des descripteurs pour un système de recherche d'images par contenu est important, dans le sens où, ce choix influe sur les résultats attendus. Cependant, d'une part il n'y a pas d'attributs universels, et d'autre part le choix des descripteurs dépend fortement de la base d'image à utiliser et des connaissances à priori qu'on peut avoir sur la base.

CHAPITRE 4 : LE TRAVAIL EFFECTUÉ



1. Introduction

Après avoir étudié le domaine de CBIR, l'implémentation d'une application d'un système de recherche d'images devient une nécessité afin d'avoir une vue plus claire. Dans ce chapitre nous allons exposer les différentes étapes par lesquelles nous sommes passés pour la réalisation de notre application.

2. Les caractéristiques visuelles choisies

Nous avons choisi les caractéristiques suivantes

2.1. Les modèles de couleur utilisés

Dans notre application, nous avons intégré seulement l'espace RGB. Il est composé des trois couleurs primaires : rouge, vert et bleu.

2.2. Préparation des indexes

Dans l'étape de préparation des indexes nous avons choisie comme descripteurs d'images les histogrammes de couleur, le nombre d'objets d'image et les moments statistiques.

2.2.1 Les histogrammes de couleur

Nous avons choisi les histogrammes de couleur comme une technique de base pour la description de la couleur dans notre application.

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.

Les histogrammes de couleur vont être calculés pour les différentes images et vont être utilisés pour la recherche des images.

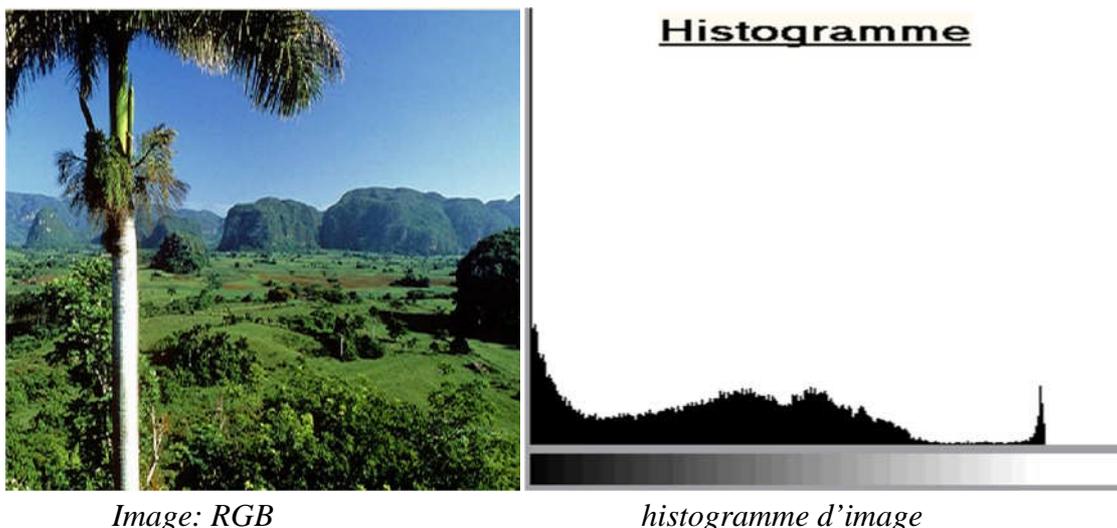


Figure 24: Exemple d'histogramme

La méthode d'histogramme utilise la distribution complète de la couleur. On doit stocker de nombreuses données.

Au lieu de calculer la distribution complète, nous avons essayé de réduire les données de l'histogramme et nous avons tiré que les informations pertinentes après des essais et des tests sur la précision des résultats et aussi sur le temps de calcul.

Les informations que nous avons tirées sont : la moyenne et le maximum de couleur pour chaque ensemble de niveaux de couleurs successifs (après des tests on a fixé un pas de 20) et maximum globale de tout l'histogramme.

2.2.2 Le nombre d'objets d'image

Nous avons ajouté une information qui est le nombre d'objets aux descripteurs d'images. Pour chaque succession de 20 niveaux on a calculé le nombre de pixels et on a fixé un seuil après des essais (seuil=5000). Si ce nombre est supérieur au seuil donc, nous considérons que c'est un objet.

2.2.3 Les moments statistiques

Pour enrichir les index afin d'améliorer les résultats de recherche de notre application, nous avons ajouté les moments statistiques. Pour cela, on calcule seulement des caractéristiques dominantes de couleur à savoir l'espérance et la variance.

Les recherches prouvent que les méthodes des moments statistiques utilisées marchent plus vite et donnent des résultats meilleurs que les méthodes d'histogrammes. [41]

L'espérance, la variance, les moments d'ordre 3 peuvent également être calculés sur chaque composante couleur par la formule suivante:

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

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

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

Où E_i est l'espérance, δ_i la variance et s_i est le troisième moment qui correspond à la composante couleur i .

Descripteurs d'image utilisés (Le nom de l'image, moyenne de couleur pour chaque ensemble de pixel, maximum de couleur, valeur maximum de l'histogramme d'image, nombre d'objet de l'image, espérance, variance)

Dans le calcul de similarité entre le descripteur d'image requête et les descripteurs des images de la base on a utilisé des coefficients associés à certains champs de descripteurs par exemple pour le nombre d'objets on a utilisé un coefficient de 10.

2.3. Mesures de distance utilisées

Mesure de distance qu'on a utilisé c'est la distance euclidienne entre les différents champs est définie par :

$$D(v1, v2) = \sqrt{\sum_{i=1}^n (v1(i) - v2(i))^2}$$

Où $v1$ et $v2$ deux vecteur de descripteur d'image et i le nombre de champs.

2.4. Les mesures de performance étudiées

Nous avons introduit deux mesures de performance à savoir : la précision et le rappel.

Dans notre projet on s'intéresse à la précision. La précision représente la chance d'avoir des images similaires à l'image requête parmi l'ensemble des images retournées. Dans un moteur de recherche on souhaiterait avoir des images similaires à notre image requête parmi les premières images retournées, tandis que dans le cas d'un système de reconnaissance de visages par exemple le rappel est très important car on ne voudrait pas rater le visage d'un criminel par exemple.

3. Architecture de l'application

Nous avons adopté le Borland Delphi comme langage de programmation, ce choix est justifié par plusieurs facteurs :

- Possibilité d'intégrer l'orienté objet dans notre logiciel.
- L'interface graphique (conviviale) offerte par cet environnement.
- Puissance du langage en matière de gestion de la mémoire central (Pointeurs, listes, tableaux...).
- Facilité de la manipulation des bases de données.

Dans un système d'extraction d'images par le contenu il existe deux différents types de traitements :

Traitement offline :

Ce type de traitement représente la phase de la construction de la base d'attributs. Cette opération est réalisée durant la construction du système.

Traitement online :

Ce type de traitement est effectué lors de l'introduction de la requête de l'utilisateur. Les attributs visuels sont extraits de l'image requête puis comparés à ceux de la base d'attributs qui a été déjà construite au préalable.

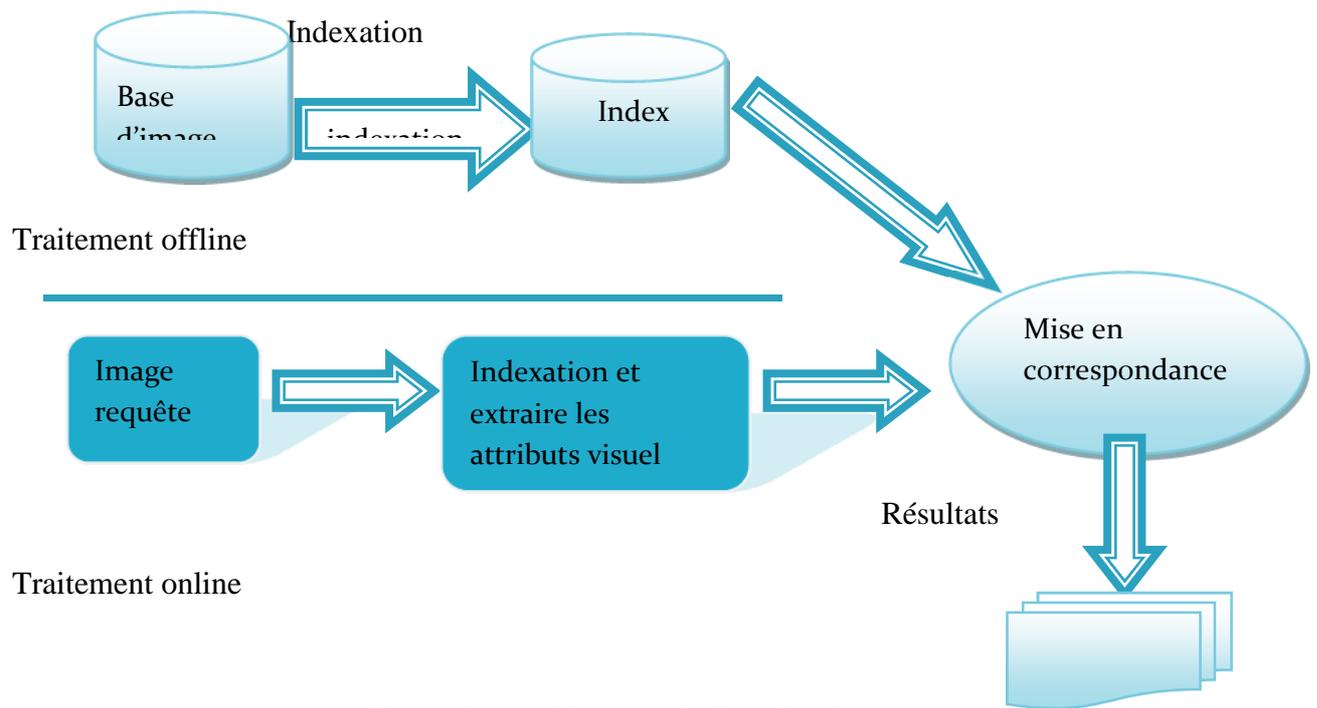


Figure 25: architecture de l'application

3.1 L'interface utilisateur

Ce composant permet à l'utilisateur d'introduire son image requête (Figure 26). La requête par exemple est la technique que nous avons choisie pour la réalisation de l'application.

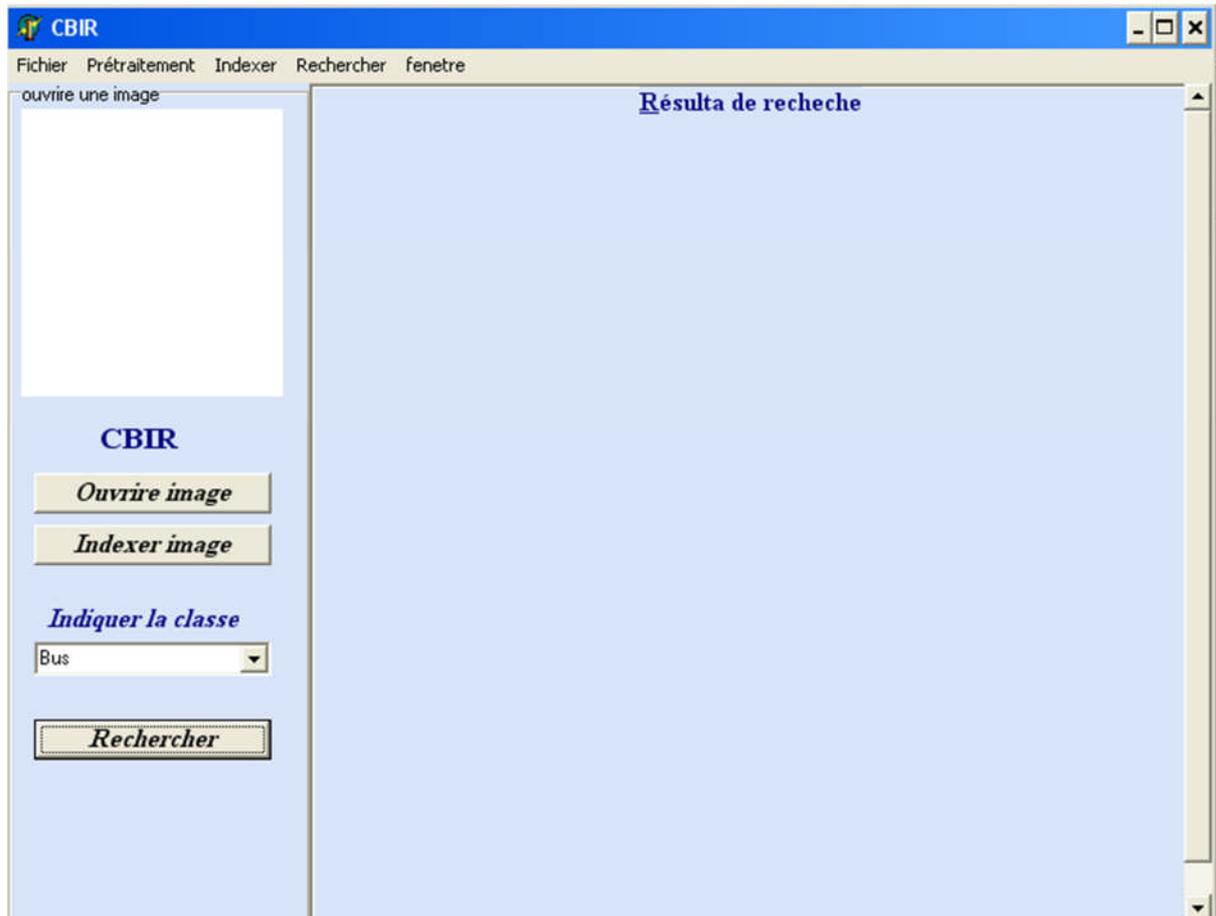


Figure 26: L'interface utilisateur.

Ce composant permet aussi de visualiser les résultats de la recherche (les images retournées après un processus de recherche).

Dans un premier lieu, l'utilisateur commence par cliquer sur le bouton 'ouvrir image' pour choisir une image requête, si l'extension de cette image est jpeg alors une petite transformation pour pouvoir l'utiliser et extraire les informations à partir de cette image. Dans le cas où l'extension de l'image est bpm, aucune transformation n'est nécessaire.

Ensuite, l'utilisateur doit choisir une classe parmi les classes d'images proposées, dans cette étape on fixe des seuils pour chaque classe.

(Figure 27).

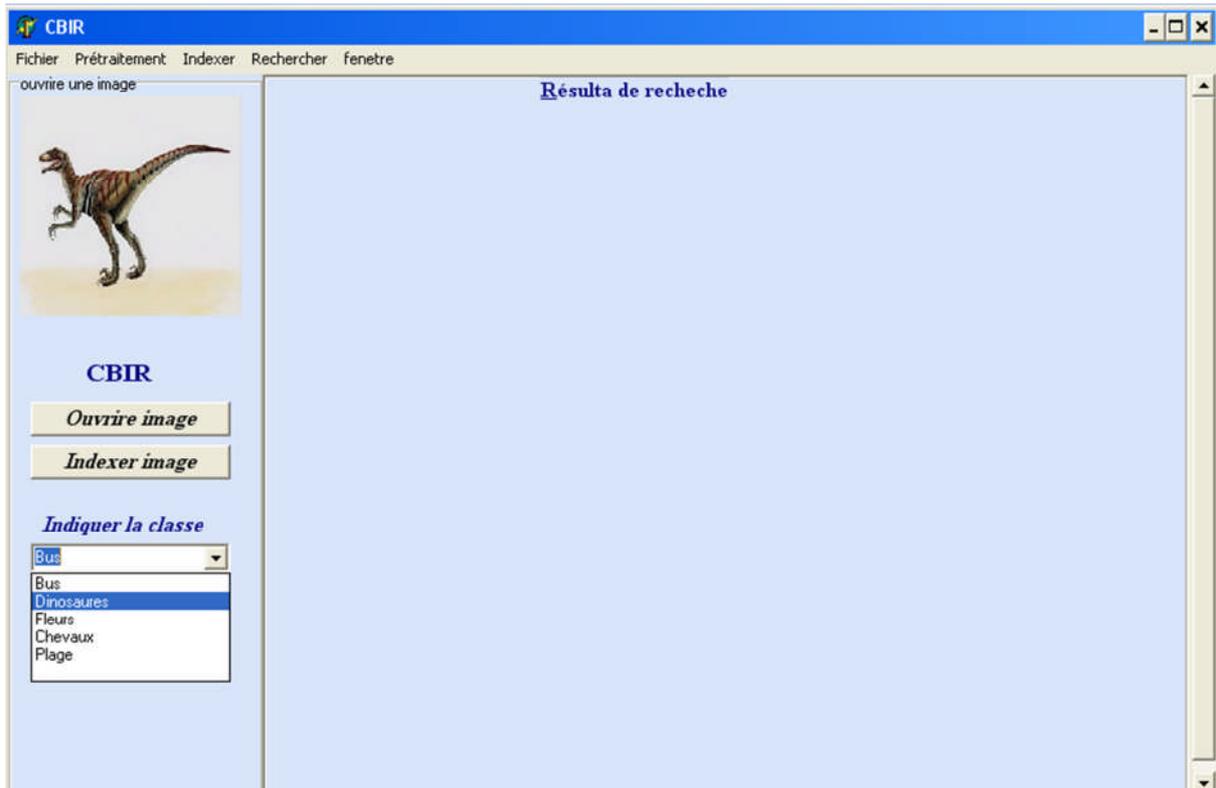


Figure 27: Le choix de l'image requête et de la classe.

Après avoir choisi l'image requête l'utilisateur indexe cette image, pendant cette étape les descripteurs d'image sont calculer et ajouter à la table d'image requête.

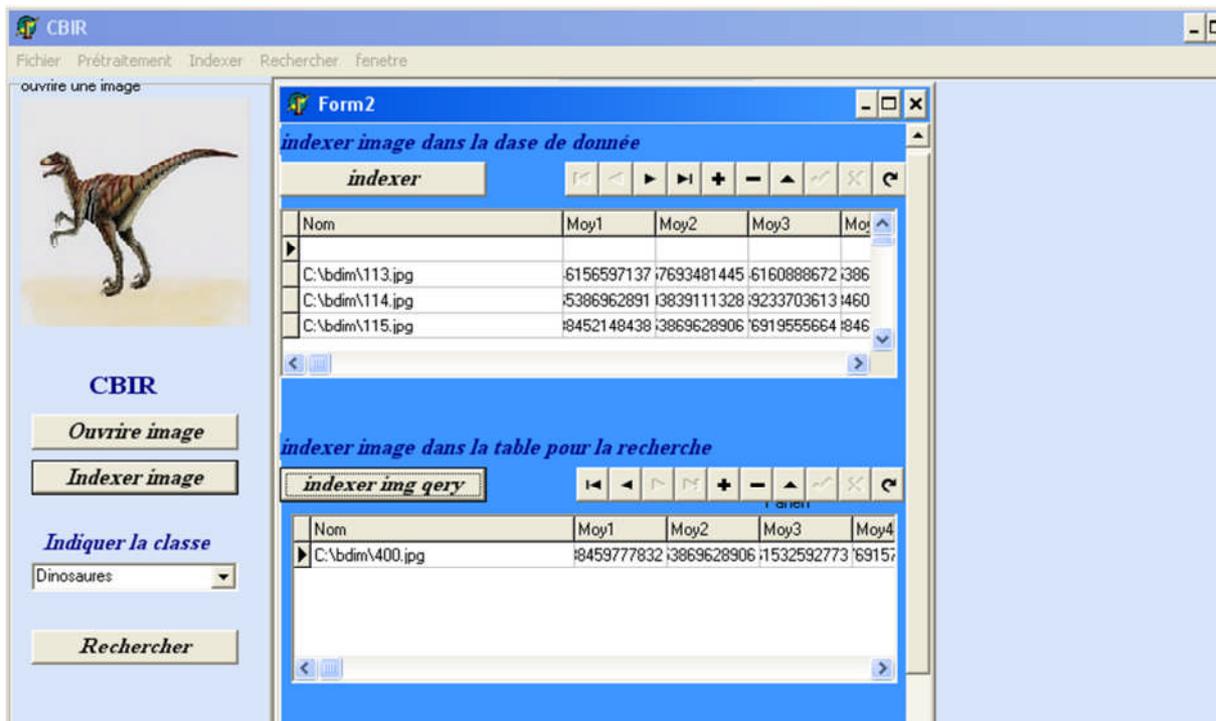


Figure 28: indexation de l'image requête

Après l'indexation l'utilisateur peut lancer la recherche en appuyant sur le bouton 'rechercher' et un ensemble d'images sont retournées est affiché (Figure 29).

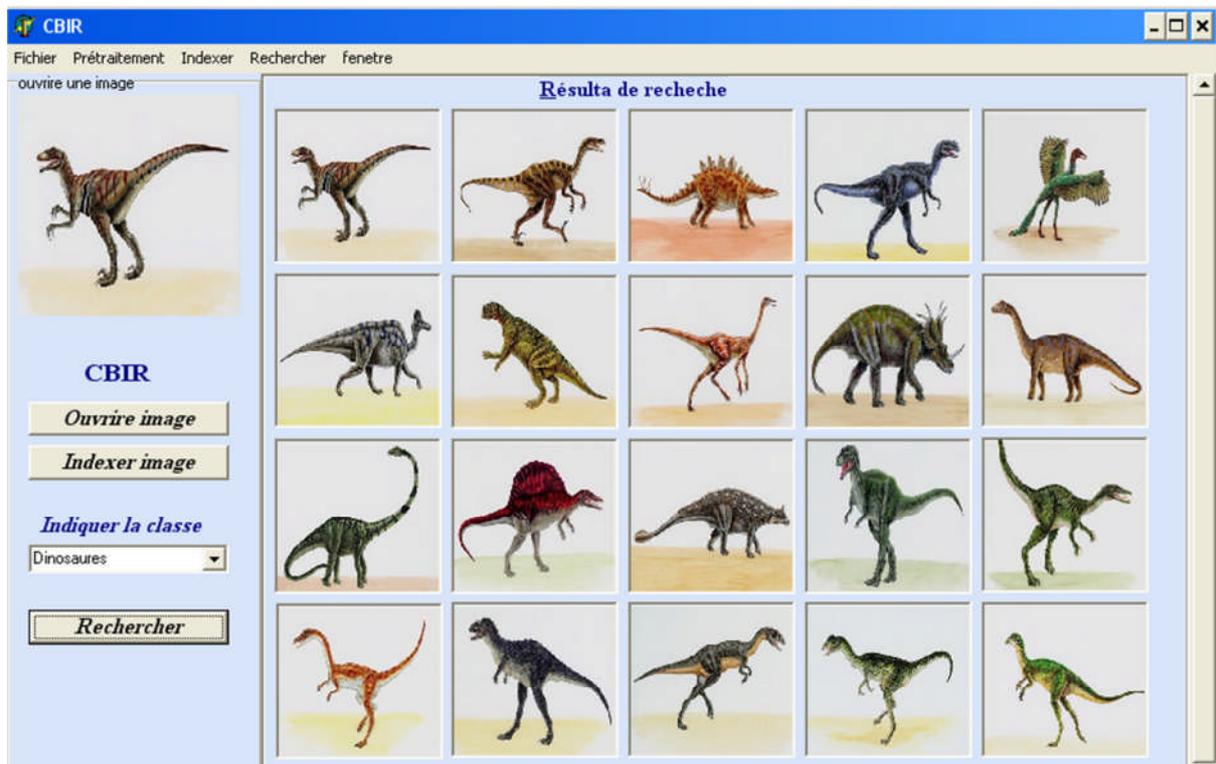


Figure 29: Un exemple d'une recherche effectuée.

3.2 Le moteur de recherche

Ce composant extrait les attributs visuels de l'image requête et effectue une recherche en se basant sur les caractéristiques utilisées. Il compare les indexes de l'image requête avec ceux des images de la base et retourne des images similaires à l'image requête

3.3 La base d'attributs

La base d'attributs est constituée d'une table contenant les histogrammes de couleur réduits et les moments statistiques des images de la base. Elle est constituée de 16 champs :

- ❖ Champ1 : Le chemin de l'image
- ❖ Champ2 au Champ6 : moyenne de couleur pour chaque ensemble de pixel.
- ❖ Champ7 au Champ11 : maximum de couleur pour chaque ensemble de pixel.
- ❖ Champ12 : valeur maximum de l'histogramme d'image.
- ❖ Champ13 : nombre d'objet de l'image.
- ❖ Champ14 : la valeur d'espérance de l'image.
- ❖ Champ15 : la valeur de variance de l'image.
- ❖ Champ16 : réserver pour récupérer l'image (DB image).

4. Evaluation de l'application

Nous avons choisi les histogrammes de couleur et les moments statistiques comme une technique de base pour la description de l'image dans notre application.

Nous avons testé la performance de ces techniques et dans cette section nous avons présenté les différents résultats.

4.1 La base d'image utilisée

La base d'images utilisée comporte 500 images couleur. Elle a été téléchargée depuis <http://wang1.ist.psu.edu/>. La base d'images originale comporte 1000 images divisées en 10 classes. Nous avons opté pour l'utilisation de 500 images et 5 classes pour des raisons de temps de calcul. Chaque classe représente un thème défini :

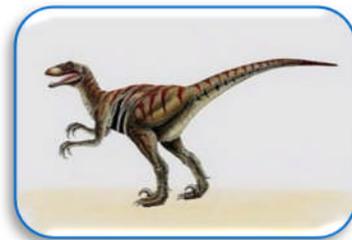
Plage, bus, dinosaures, Fleurs et Cheveaux. Un échantillon de la base est Présenté par la figure suivante (Figure 30) :



Plage



Bus



Dinosaures



Fleurs



Cheveaux

Figure 30: Un échantillon de la base d'images utilisée.

4.2 Processus suivi

Dans cette partie nous avons évalué l'application avec le calcul de précision pour chaque image requête. Le principe de fonctionnement de ce processus d'évaluation peut être décrit par la figure 32. Les images requêtes sont sélectionnées de manière aléatoires (sans répétition de la même image requête). La figure 31 représente un exemple d'un ensemble d'images requête.

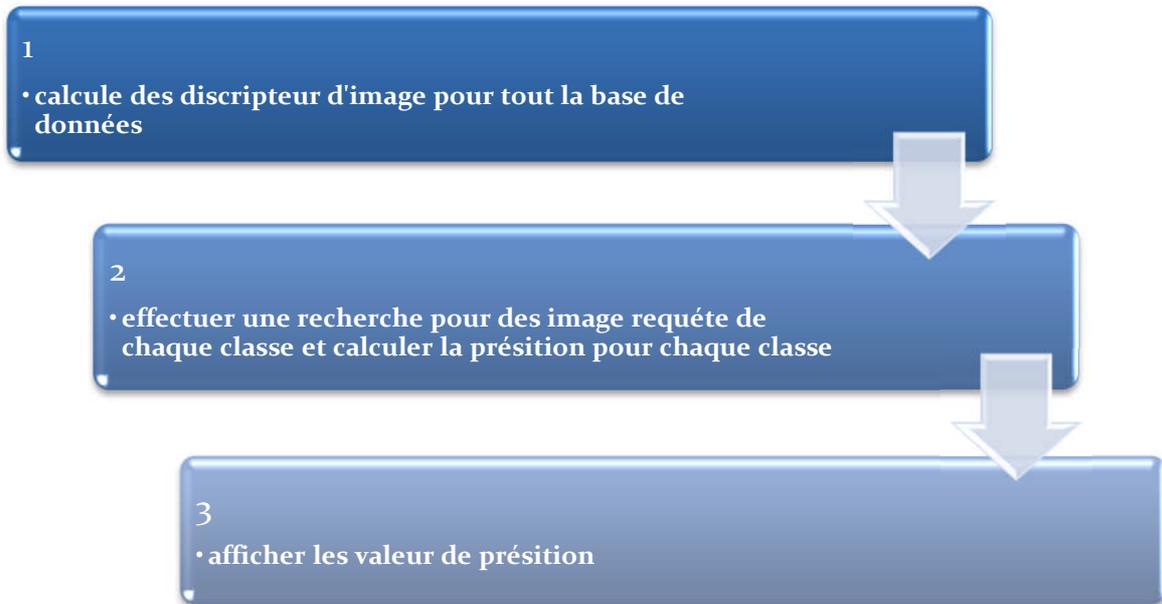


Figure 31: schémas représentative du processus

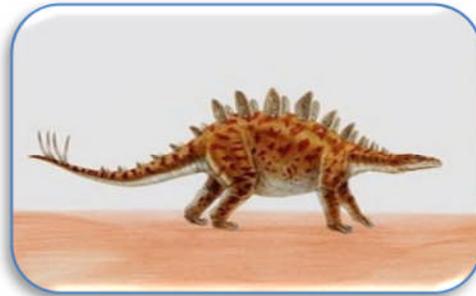




Figure 32: un exemple d'un ensemble d'image requête.

Nous avons effectué un ensemble de tests sur 50 images requête (10 images de chaque classe). Ces images ont été choisies aléatoirement de la base d'images. Les descripteurs de chaque image requête sont comparés à ceux de la base et les images les plus similaires vont être retournées.

Les précisions est calculée pour chaque classe de la base c'est la moyenne de la précision des images requêtes de chaque classe. Ainsi on obtienne 5 précision moyenne chacune pour une classe .pour une évaluation globale, la précision moyenne de toutes les classe est calculer.

Les résultats sont décrits par le tableau des résultats (Figure 33)

Résultats obtenus :

	Plage	Bus	Dinosaures	fleurs	Chevaux
Image requête 1	1	1	1	0.35	0.57
Image requête 2	0.7	1	0.7	0.5	0.37
Image requête 3	0.9	0.9	1	0.72	0.61
Image requête 4	1	1	1	0.55	0.7
Image requête 5	1	1	1	0.25	0.3
Image requête 6	1	1	0.95	0.2	0.5
Image requête 7	0.3	0.85	1	0.1	0.4
Image requête 8	0.55	0.95	1	0.87	0.45
Image requête 9	0.95	0.85	0.9	0.6	0.6
Image requête 10	1	1	1	0.45	0.5
Précision moyenne	0.84	0.95	0.95	0.45	0.5

Figure 33: les précisions moyennes obtenues

Précision moyenne globale des 5 classes est de 73%

Ces précisions sont considérées comme acceptables en prenant en considération la base d'images utilisée. Malgré que les images de chaque classe représentent le même objet, elles présentent une certaine hétérogénéité dans la distribution de la couleur.

La classe « Dinosaures » présente la plus grande précision (95%), pour tous les modèles de couleur. Les images de cette classe ne sont pas des photographies mais des images synthétisées. 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 (5%) sont des images qui contiennent une grande quantité de couleur blanche.

Classe « Bus » présentent une bonne précision pour la même raison.

En conclusion, les histogrammes de couleur et les moments statistiques fournissent de bons résultats pour la recherche d'images par le contenu.

5. 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.
- aussi le choix du modèle de couleur et la mesure de distance à utiliser.
- Et aussi les extensions des images. Dans un premier lieu nous avons travaillé qu'avec les images bmp après, nous avons pu utiliser des images jpg après une petite transformation.

6. Perspective

Comme perspectives, Nous proposons :

- d'utiliser les descripteurs de hauts niveaux (sémantique).
- D'ajouter d'autres types de descripteurs de bas niveau comme la texture et la forme.
- De Travailler avec tous les extensions des images existant.
- D'essayer autres techniques de mesures de similarité entre les descripteurs d'images.
- Tester notre application avec d'autres bases de données.

7. Conclusion

Nous avons présenté dans ce chapitre les caractéristiques visuelles choisies (modèle de couleur, préparation de l'index, les mesures de distance utilisées 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

L'indexation et la recherche d'images par le contenu sont des problèmes complexes et incontournables étant donnée la place que l'image numérique occupe à présent dans notre quotidien. Internet en est la meilleure illustration. Dans notre travail nous avons intéressés d'une part à l'indexation des images par le contenu et d'autre part à la recherche par le contenu.

Du point de vue de l'indexation, nous avons proposé un traitement automatique pour le calcul des indexes des images. Ce traitement se base sur l'analyse de la couleur, où nous avons porté une attention à l'utilité des histogrammes de couleurs et les moments statistiques.

Du point de vue de recherche, nous avons utilisé le principe de la recherche par similarité. Cette dernière prend la forme d'une image exemple. Durant ce travail nous avons constaté deux points essentiels et cruciaux. Le premier, est que la couleur est une caractéristique discriminante d'une image, mais l'utilisation de la couleur tout seul dans un CBIR ne suffit pas, il faut rajouter autre descripteurs de texture et de forme pour la performance d'un CBIR.

Le domaine de recherche d'image par le contenu est un domaine très riche et exhibe une variance dans les techniques utilisées. Il n'existe pas une loi qui impose le choix d'une technique particulière pour l'extraction des images en se basant sur le contenu.

On espère qu'on a atteint notre but par ce mémoire qui ne représente en réalité qu'une goutte d'eau dans la mer du domaine.

BIBLIOGRAPHIE

- [1]: P. Agouris, J. Carswell et A. Stefanidis, "An environment for content-based image retrieval from large spatial databases", *Journal de photogrammetry and remote sensing*, Vol. 54, No. 4, pp. 263-272, 1999.
- [2]: D. H. Ballard: Parameter networks, Towards a theory of low level vision. Proc 7th IJCAI, Vancouver, pp. 1068-1078, 1978.
- [3]: D. Barba et J. Ronsin, Image segmentation using new measure of texture feature. *journal de Digital Signal Processing* 84, pp. 749-753, 1984.
- [4]: Hadjila Feth Allah et Bouabdallah Réda, Reconnaissance des visages en utilisant les réseaux de neurones. Mémoire d'ingénieur. Université de Tlemcen. 2003.
- [5]: A. C. Bovik, M. Clark et W. S. Geisler, Multichannel texture analysis using localized spatial filters; *journal IEEE Trans. PAMI*, vol. 12 pp. 55-73, 1990.
- [6]: C. Coroyer, Apport des corrélations d'ordre élevé à l'analyse de textures non gaussiennes. Thèse de Doctorat, Univ. Cergy-Pontoise, France, 1996.
- [7]: A. Del Bimbo, Visual Information Retrieval. Morgan Kaufmann Publishers, 1999.
- [8]: Y. Deng et B. Manjunath, Unsupervised segmentation of color texture regions in images and video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23:800–810, 2001.
- [9]: Deselaers T, Features for Image Retrieval, Diploma Thesis, RWTH Aachen University. Aachen, Germany, December 2003.
- [10]: Francis R. Bach et Michael I. Jordan. Learning Spectral Clustering, EECS Department, University of California, Berkeley. Technical Report No. UCB/CSD-03-1249. June 2003.
- [11]: M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele et P. Yanker, Query by Image and Video Content: The QBIC System, *IEEE Computer*, vol. 28, no. 9, pp. 23-32, 1995.
- [12]: I. Fogel et D. Sagi. Gabor, filters as texture discrimination. *Bio. Cybern.*, Vol. 61, pp. 103-113, 1989.

- [13]: A. P. Gagalowicz, Vers un modèle de texture. Thèse d'état, université Pierre et Marie Curie, Paris VI, 1983.
- [14]: Gong Y., Proietti G. et Faloutsos C, Image Indexing and Retrieval Based on Human Perceptual Color Clustering. Proceeding Of International Conference on Computer Vision and Pattern Recognition (CVPR), 1998.
- [15]: C. Gratin, J. Vitria, F. Moreso et D. Seron, Texture classification using neural networks and local granulometries. Kluwer Academic Publishers, J.Serra & P. Soille Ed., pp. 309-317, 1994.
- [16]: A.Guérin-Dugué, S. Ayache et C.Berrut Image retrieval a first step for a human centered approach, in Fourth Pacific-Rim Conference on Multimedia, Singapore, 15-1, 2003.
- [17]: T. E. Hall et G. B. Giannakis, Image modeling using inverse filtering criteria with application to textures. IEEE Trans. on Image Processing, Vol. 5, No. 6, pp. 938-949, 1996.
- [18]: R.M.Haralik, K. Shanmugam et I. Dinstein, Textural features for images Classification. IEEE Transaction on System , Man, Cybernetics, 3,610-621, 1973.
- [19]: R. O. DUDA et P. E. HART, Use of the Hough Transformation to Detect Lines and Curves in Pictures. Communication of the ACM, Vol. 15, No. 1, pp. 11-15, 1972.
- [20]: P. V. C. HOUGH, Method et means for recognizing complex patterns. United States Patent, No. 3, 069, 654, 1962.
- [21]: P.S. Hiremath, S. Shivashankar, et J. Pujari , Wavelet based features for color texture classification with application to CBIR. International Journal of Computer Science and Network Security (IJCSNS), 6(9) :124–133, 2006.
- [22]: M. K. Hu, Visual pattern recognition by moments invariants. Computer methods in image analysis. Transactions on Information Theory, 8, 1962.
- [23]: B. Burke Hubbard, Ondes et ondelettes. Sciences d'Avenir.Belin, 1995.
- [24]: A.K.Jain et F.Farrokhnia, Unsupervised texture segmentation using Gabor filters. Pattern recognition, Vol. 24, No. 12, pp. 1167-1186, 1991.

- [25]: G.W. Jiji et L. Ganesan, Comparative analysis of colour models for colour textures based on feature extraction. *International Journal of Soft Computing*, p 361–366, 2007.
- [26]: Kamel Houari, Recherche d'images par le contenu Doctorat kin, Université Mentouri de Constantine, 2010.
- [27]: Khouloud Meskaldji, Extraction et traitement de l'information : Un prototype d'un système de recherche d'images couleurs par le contenu magistère, Université Mentouri de Constantine, 2009.
- [28]: Michael S. Lew, D. P. Huijsmans et Dee Denteneer, Content based image retrieval KLT, projections, or templates. In A. W. M. Smeulders et R. Jain: editors. *Image Databases and Multi-Media Search, proceedings of the First International Workshop IDB-MMS'96*, Amsterdam: The Netherlands. Amsterdam University Press, pages 27-34. , 1996.
- [29]: Jia Li et James Z. Wang, Studying digital imagery of ancient paintings by mixtures of stochastic models, *IEEE Transactions on Image Processing*.
- [30]: G. Loum ,Segmentation pyramidale de textures par décomposition en ondelettes. Thèse de Doctorat de l'université de Paris XII, Val de Marne, 1996.
- [31]: C. S. Lu, P. C. Chung et C. F. Chen, Unsupervised texture segmentation via wavelet transform. *Pattern Recogniton*, Vol. 30, No. 5, pp. 729-742, 1997.
- [32]:S.G.Mallat, A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 11, pp. 674-693, 1989.
- [33]: T. P. Minka, An Image Database Browser that Learns from User Interaction , Master of Engineering Thesis, 1996.
- [34]: Virginia E. Ogle et Michael Stonebraker. Chabot, Retrieval from a relational database of images. *IEEE Computer*, Vol. 28, No. 9, pp. 40-48, 1995.
- [35]: P.Wu, B.S. Manjunath, S.Newman et H.D. Shin, A texture descriptor for browsing and similarity retrieval, *Signal processing: Image communication*, vol.16, no.1, 2, pp :33-43,2000.
- [36]: Fabio Policarpo, *The Computer Image*, ACM Press. Pages 298-308. 1998.

- [37]: F. Preteux et M. Schmitt : Boolean texture analysis and synthesis. Academic Press, J. Serra Ed., Vol. 2, pp 377-400, 1988.
- [38]: A. R. Rao et G. L. Lohse: Towards a texture naming system , Identifying relevant dimension of texture. IBM Research report, RC 19140 (83352), pp.29 , 1993.
- [39]: A. R. Rao et G. L. Lohse, Identifying high level features of texture perception. Computer Vision, Graphics and Image Processing , Graphic Models and Image Processing, Vol. 55, pp. 218-233, 1993.
- [40]: J. R. Smith et S.-F: Chang, Querying by color regions using the VisualSEEk content-based visual query system, In Intelligent Multimedia Information Retrieval. IJCAI, pp 159-173, 1996.
- [41]: M.A. Stricker et M. Orengo, Similarity of color images . In SPIE, Storage and Retrieval for image Video Databases, pages 381-392, 1995.
- [42]: Aditya Vailaya, Mário Figueiredo, Anil Jain et HongJiang Zhang. A: Bayesian Framework for Semantic Classification of Outdoor Vacation Images, IEEE Trans. Image Processing, Vol. 10, No. 1, pp. 157-172, 2001.
- [43]: J. Kovacevic et M. Vetterli, Wavelets et Subband Coding. Prentice Hall, 1995.
- [44]: J. Wang, Champs Markoviens multi-échelles applications à la segmentation d'images texturées et à la fusion multi-film. Thèse de Doctorat, Université Paris XI, Centre d'Orsay France, 1994.
- [45]: John P .Eakins et Margaret E .Graham, Content-based Image Retrieval. A report to the JISC Technology Applications Programme. Institute for Image Data Research, University of Northumbria at Newcastle. January 1999.
- [46]: Remco C. Veltkamp, Mirela Tanase, "Content-Based Image Retrieval Systems, A Survey", Department of Computing Science, Utrecht University, Technical Report UU-CS-2000-34, March 8, 2001.
- [47]: T. S. Lai, "CHROMA, a photographic image retrieval system", PhD thesis, School of computing, engineering and technology, University of Sunderland, UK, 2000.

[48]: Y. Rubner. Perceptual metrics for image database navigation. Rapport Technique CS-TR-99-1621, Stanford University, 1999.

[49]: Y. Rubner. Perceptual metrics for image database navigation. Rapport Technique CS-TR-99-1621, Stanford University, 1999. Swain M.J., Ballard D.H.(1991), "Color indexing". International Journal of Computer Vision, vol. 7, no. 1, pp. 11-22, 1991.

SITOGRAFIE

[S1]: http://fr.wikipedia.org/wiki/Image_num%C3%A9rique

[S2]: <http://www.kaddour.com/chap1/chap1.htm>

Résumé

Résumé

L'objectif principal de ce mémoire est d'étudier les méthodes déjà existantes à l'heure actuelle dans le domaine de la recherche d'images par le contenu et de construire une application de recherche d'images par le contenu en se basant sur ces connaissances. Nous nous intéressons dans notre cas aux images couleurs. Il existe une large variété de sources de recherches d'images couleurs. Nous avons choisi les histogrammes de couleur comme une source d'information de base pour la réalisation de l'application et quelques descripteurs de couleurs.

Nous étudions d'abord l'effet de la réduction d'histogramme et nous essayons de tirer quelques conclusions de cette étude, aussi les moments statistiques tels que l'espérance et la variance. Par la suite nous exploitons ces paramètres pour l'implémentation de notre application.

Notre application reçoit une image requête, effectue une recherche, et retourne les images similaires à l'image requête en se basant sur les caractéristiques colorimétriques (de couleur) des images.

ملخص

الهدف الرئيسي من هذه المذكرة هو دراسة الأساليب و الطرق الموجودة حاليا في مجال البحث عن الصور حسب المحتوى وإنشاء نموذج لنظام بحث عن الصور حسب المحتوى اعتمادا على هذه المعارف. ونحن نهتم في مثالنا هذا خاصة على الصور الملونة. مع العلم انه يوجد مجموعة كبيرة و متنوعة من تقنيات البحث على الصور الملونة. اخترنا رسوم بيانية ملونة كمصدر للمعلومات الأساسية لإنشاء النموذج و بعض واصفات الألوان. لقد درسنا أولا تأثيرات تقليص الرسم البياني، و لقد استخلصنا بعض المعلومات من هذه الدراسة. لقد استخدمنا أيضا لحظات لإحصائية مثل المتوسط والتباين، بعد ذلك نستخدم هذه المعلومات في تطبيق النموذج. النموذج يستقبل صورة ، ويؤدي عملية البحث، تم يأتي بصور مماثلة للصورة الأولى معتمدا على الخصائص اللونية للصورة.

Abstract

The main objective of this paper is to study the existing methods currently in the field of image retrieval by content and build an application based image retrieval content based on these knowledge. We are interested in our case with color images. A wide variety of research sources of color images. We chose the color histograms as a source of basic information for making the application and some descriptors of color. We study first the effect of the reduction of histogram and we try to draw some conclusions from this study also the statistical moments such as mean and variance. Subsequently we use these parameters for the implementation of our application. Our application from an image query, searches, and returns the images similar to the query image based on color characteristics (color) images.