الجمهورية الجزائرية الديمقراطية الشعبية
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
وزارة التعليم العالي و البحث العلمي
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
جامعة أبي بكر بلقايد – تلمسان –
Université Aboubakr Belkaïd – Tlemcen –
Faculté de TECHNOLOGIE

MEMOIRE

Présenté pour l'obtention du **diplôme** de **MASTER**

En : Télécommunication

Spécialité : Réseaux et Télécommunications

Par : DICKO ALLASANE DOUIDI ASSIA

Sujet

Mise en place d'un système d'indexation et de recherche d'image par le contenu

Soutenu publiquement, le 23 / 06 / 2022 , devant le jury composé de :

M BOUSAHLA Miloud	M.C.B	Université de Tlemcen	Président
Mme BENLALDJ Lamia	M.A.A	Université de Tlemcen	Examinatrice
M BOUABDALLAH Réda	M.A.A	Université de Tlemcen	Encadreur

Année universitaire :2021/2022

Remerciement

On remercie dieu le tout puissant de nous avoir donné la santé et la volonté d'entamer et de terminer ce mémoire.

Tout d'abord, ce travail ne serait pas aussi riche et n'aurait pas pu avoir le jour sans l 'aide et l'encadrement de Mr BOUABDALLAH. On le remercie pour la qualité de son encadrement exceptionnel, pour sa patience, sa rigueur et sa disponibilité durant notre préparation de ce mémoire

Nos profonds remerciements à Monsieur BOUSAHLA Miloud qui nous a fait l'honneur de présider le jury de ce travail

Nos remerciements s'adressent aussi à Madame BENLALDJ Lamia qui nous on fait l'honneur d'examiner et juger ce modeste travail.

Nos sincères remerciements et gratitudes a notre chef de département Mr IRID , et Mr HADJILA et l'ensemble des enseignants du Département des télécommunication pour leurs générosité et la, grande patience dont Ils ont su faire preuve malgré leurs charges académiques et professionnelles.

Dédicace

A mon très cher père, Abdelmadjid

l'homme de ma vie, mon exemple éternel, mon soutien moral et source de joie et de bonheur, celui qui s'est toujours sacrifié pour me voir réussir, que dieu te garde dans son vaste paradis, à toi mon père.

A ma très cher mère, Nacera

si Dieu a mis le paradis sous les pieds des mères, ce n'est pas pour rien. Tu représentes pour moi le symbole de la bonté par excellence, la source de tendresse et l'exemple du dévouement qui n'a pas cessé de m'encourager et de prier pour moi. Ta prière et ta bénédiction m'ont été d'un grand secours pour mener à bien mes études. Aucune dédicace ne saurait être assez éloquente pour exprimer ce que tu mérites pour tous les sacrifices que tu n'as cessé de me donner depuis ma naissance, durant mon enfance et même à l'âge adulte. Tu as fait plus qu'une mère puisse faire pour que ses enfants suivent le bon chemin dans leur vie et leurs études. Je te dédie témoignage de mon profond amour. Puisse Dieu, le tout puissant, te préserver et t'accorder santé, longue vie et bonheur

A Mon chère frère et sœurs, Oussama et Nihel

pour leur compréhension et leur grande tendresse, qui en plus de m'avoir encouragé tout le long de mes études, m'ont consacré beaucoup de temps et disponibilité, et qui par leur soutien, leurs conseils et leur amour, m'ont permis d'arriver jusqu'à ici car ils ont toujours cru en moi, Merci d'avoir toujours soutenu et merci pour tous les bons moments passé ensemble, et ce n'est pas fini.

A ma famille et toutes les personnes que j'aime

A Mon fiancé Abderrahmane et ma futur famille

A mon binôme dicko avec lequel j'ai partagé cette expérience.

A tous mes amies surtout wahiba, nihed,amina, kawter, et masheke qui m'ont toujours aidé et encouragé, qui étaient toujours à mes côtés, et qui m'ont accompagnaient durant mon chemin d'études supérieures.

ASSIA DOUIDI

Dédicace

A mon très cher père, Lamine

Tu as toujours été mon plus grand soutient dans toute vie éducative. Toi qui as su m'enseigner toutes les grandes valeurs qui ont fait de moi ce que je suis devenue. Je le dirai peut être pas ouvertement, mais ton sérieux et ta persévérance dans ton métier ont toujours été une très grande source d'inspiration pour moi. Je te suis reconnaissant pour tous les sacrifices dont tu as fait preuve pendant tant d'années.

A ma très cher mère, Niamoye

Toi qui est, et qui sera probablement pour toujours mon plus grand soutient. Toi qui s'est montré si patiente face à mon manque de tact émotionnelle. J'ai observé et j'admire, le dévouement avec lequel tu t'es appliquer nous rendre la vie plus confortable. Ton amour inconditionné, ta conviction morale et religieuse ont fait de moi une personne meilleur.

Grace à toi, je me suis senti aimé pendant tant d'années.

A ma tente, Moulher

Toi, qui m'a fait souhaiter avoir une grande sœur par tous les bon moment qu'on aura passé. Qui a pris soin de moi pendant de nombreuse années.

Merci pour tout ce que tu as pu faire pour moi.

A mes frères et ma sœur,

Je souhaite à chacun d'entre vous un future radieux.

Pour tous les bon moments qu'on aura passé ensemble, Merci.

Il serait injuste de finir cette dédicace, sans remercier tous les amis qui m'ont soutenue durant toute mon parcours universitaire.

Je remercie mon binôme Assia et mon encadreur BOUABDALLAH.R avec qui j'ai effectué ce laborieux travail.

Je remercie également toute personnes qui aura souhaité ma réussite pendant ces dix-sept long années d'étude.

Puisse Dieu vous donné santé, bonheur, courage et réussite.

DICKO ALLASSANE

Résumé

Avec les avancées technologiques, l'importance des medias numérique a évoluée de manière exponentielle. Ainsi la nécessité de mise en place des systèmes d'indexation et de recherche de media par le contenue (CBIR en anglais content based image retrieval) s'est imposé. Ces systèmes auront pour objectif de donner une 'Vision' à l'ordinateur afin de permettre une détection automatique du contenue des images. Pour atteindre ce but, il faut développer des technique de traitement d'image numérique de plus en plus avancées.

Dans ce mémoire, nous allons étudier les images numériques, les descripteurs d'image, le fonctionnement des systèmes CBIR, les méthodes de classifications et nous allons implémenter un système d'indexation et de recherche d'image par le contenu. Dans notre implémentation nous utiliseront les réseaux de neurone convolutif comme méthode récente de classification d'image basée sur l'apprentissage profond (deep learning). Nous utiliserons le réseau de neurone convolutif MobileNet et mettront en avant ses avantages pour la classification d'image. MobileNet combiné a un algorithme de détection d'objet appelé SSD (Single Shoot Multi Box Detector), nous permettra une classification d'image efficace et rapide. Nous utiliserons également comme descripteur d'image des histogrammes de couleurs dans l'espace colorimétrique HSV.

Le système implémenté sera testé sur la célèbre base d'image Pascal VOC (Visual Object Classes) contenant 17 250 images et 20 classes d'images.

Abstract

With advances in technology, the importance of digital media has grown exponentially. Thus the need to set up indexing and image search systems by content has become essential. These systems will aim to give a 'Vision' to the computer in order to allow automatic detection of the content of the images. To achieve this goals; increasingly advanced digital image processing techniques must be developed.

In this thesis, we will study digital images, image descriptors, the functioning of CBIR systems, classification methods and we will implement a system of indexing and image search by CBIR content. In our implementation we will use convolutional neural networks, which are recent methods of image classification based on deep learning. We will use MobileNet's convolutional neural network and highlight its advantages for image classification. MobileNet combined has an object detection algorithm called SSD (Single Shoot Multi Box Detector). We will also use colour histograms in the HSV color space as an image descriptor.

The implemented system will be tested on the famous Pascal VOC (Visual Object Classes) image database containing 17,250 images.

ملخص

مع التقدم التكنولوجي، نمت أهمية الوسائط الرقمية بشكل كبير. وبالتالي فإن الحاجة إلى إنشاء أنظمة فهرسة الوسائط والبحث فيها عن طريق المحتوى (CBIR و معناها استرجاع الصور على أساس المحتوى) أصبحت ضرورية.

تهدف هذه الأنظمة إلى إعطاء رؤية للكمبيوتر للسماح بالكشف التلقائي عن محتوى الصور. لتحقيق هذا الهدف ،يجب تطوير المزيد والمزيد من تقنيات معالجة الصور الرقمية المتقدمة

في هذه الرسالة ، سوف ندرس الصور الرقمية ، وصفات الصور ، وعمل أنظمة CBIR ، وطرق التصنيف ، وسنقوم بتنفيذ نظام الفهرسة والبحث عن الصور حسب المحتوى. في تطبيقنا سوف نستخدم الشبكات العصبية التلافيفية كطريقة حديثة لتصنيف الصور على أساس التعلم العميق. سنستخدم الشبكة العصبية التلافيفية MobileNet ونسلط الضوء على مزاياها في تصنيف الصور. MobileNet مع خوارزمية لاكتشاف SSD (كاشف اللقطة الواحدة متعدد الصناديق) ، سيتيح لنا تصنيفًا سريعًا وفعالاً للصور. سنستخدم أيضًا الرسوم البيانية الملونة في مساحة ألوان HSV كواصف للصورة.

سيتم اختبار النظام المنفذ على قاعدة بيانات صور Pascal VOC الشهيرة(Visual Object Classes) التي تحتوي على 17250 صورة و 20 فئة صور

Table de matière

Liste des tableaux

Liste des figures

Introduction générale

Avec les avancées technologiques, l'importance des médias numérique a évoluée de manière exponentielle. Les données numériques constituent une écrasante majorité des données stockées sur internet. Le besoin de rechercher des images c'est alors imposé sur le web. Les systèmes de recherche d'image étaient tout d'abord textuels. Les métas data qui sont données donnant une description global des images ont alors été les seul outils utilisé pour la recherche d'image. Ainsi avec des mots clés, il était possible d'effectuer des recherches sur une base d'image. Cette méthode de recherche d'image, bien que parfois efficace, est très limité. Ainsi la nécessités de meilleur système de recherche, la nécessité d'un système d'indexation automatique (due à une quantité trop importante d'image pour une indexation manuelle), la nécessité d'exploiter les informations contenu dans les médias de manière automatique ont permis le développement de système d'indexation et de recherche d'image par le contenue. Ces systèmes auront pour objectif de donner une 'Vision' à l'ordinateur afin de permettre une détection automatique du contenue des images. Cela permettra la mise en place de système de reconnaissance faciale, de recherche d'image, la détection d'objet sur une vidéo, et même de réguler le contenue des médias que les utilisateurs peuvent poster sur une base de données.

Pour atteindre ces buts ; il faut développer des techniques de traitement d'images de plus en plus avancées. Ces techniques nous permettrons d'extraire les valeurs caractéristiques des pixels des images numériques. Ces valeurs caractéristiques également appelé descripteurs d'images permettent de mesurer la similarité entre les images. On peut les répartir en plusieurs catégories de descripteurs : les descripteurs de couleur, les descripteurs de texture, les descripteurs de forme. Ainsi, plus les descripteurs sont efficaces dans leur rôle de caractérisation des images, plus le système dans lequel ils sont utilisés peut-être efficace.

Récemment, avec l'évolution de l'intelligence artificielle et des systèmes d'apprentissage automatique, on voie apparaitre des algorithmes d'apprentissage automatique améliorant encore mieux l'efficacité des systèmes de recherche. Ces techniques d'apprentissage automatique peuvent servir à extraire de nouvelles variables d'image plus abstraites et plus stables. On peut citer parmi ces techniques, le Clustering, les réseaux de neurones et c'est d'ailleurs ce dernier que nous tenterons d'implémenter par la suite. L'apparition des réseaux de neurones convolutifs et leur utilisation dans la classification d'image permet des classifications de haute précision.

Dans ce mémoire, nous allons implémenter un système de recherche d'image par le contenue en utilisant un model pré-entrainer de Mobilenet. Ces choix sont justifiés grandement par la rapidité de MobileNet et son accessibilité aux systèmes de faible performance tout en conservant une précision véritablement correcte. Notre implémentation peut ainsi être utilisé autant sur des ordinateurs de faible performance, des serveurs web/mobile, des Smartphone et autre systèmes similaires.

Ce mémoire se compose de cinq chapitres. Dans le premier chapitre, nous expliqueront les bases des traitements d'image indispensable à tous systèmes CBIR. Dans le deuxième chapitre nous nous attarderont sur le fonctionnement, l'architecture, ainsi que des informations générales sur les systèmes CBIR. Dans le troisième chapitre, nous présenteront quelque catégorie descripteur d'image (couleur, forme, texture), ainsi que des modèles de

caractérisation permettant l'extraction de ces catégories de descripteur. Dans le quatrième chapitre, nous parlerons des quelques méthodes de classification des images. Notamment les algorithmes de classification et les approches de la classification d'image par les réseaux de neurone convolutif. Le cinquième chapitre est en fin consacré à l'implémentation et à la présentation de notre logiciel de cherche d'image basé sur le contenue. Dans ce chapitre, nous justifieront nos diffèrent choix et présenteront les résultats obtenus par notre système.

Chapitre I : Généralités sur le Traitement d'Images numérique.

1 Introduction

Avec l'omniprésence et l'utilisation des technologies telles que les Smartphones, les ordinateurs et les télévisions numériques, l'utilisation et le stockage des media numériques atteint un seuil jamais atteint au par avant. De ce fait, dans le but d'établir un système de recherche d'image basé sur le contenue, il est indispensable d'étudier les images numériques.

Dans ce chapitre, nous nous intéresseront aux notions de bases concernant les images numérique. Nous parlerons des différents types d'image, des différents formats d'image, ainsi que des caractéristiques d'une image numérique.

2 Définition d'image numérique

Une image numérique est une représentation d'une image réelle sous la forme d'un ensemble de nombres pouvant être stockés et manipulés par un ordinateur numérique. Afin de traduire l'image en chiffres, elle est divisée en petites zones appelées pixels (éléments d'image). Pour chaque pixel, le dispositif d'imagerie enregistre un nombre, ou un petit ensemble de nombres, qui décrivent une propriété de ce pixel, comme sa luminosité (l'intensité de la lumière) ou sa couleur. Les nombres sont disposés dans un tableau de lignes et de colonnes qui correspondent aux positions verticale et horizontale des pixels dans l'image.

3 Types d'images

Il y a plusieurs types d'image, parmi elles, les images noir et blanc, les images en teinte de gris, les images en couleurs. Les images noir et blanc n'enregistrent que l'intensité de la lumière tombant sur les pixels et c'est le type d'image que l'on utilise pour scanner du texte quand celui-ci est composé d'une seule couleur. Une image en niveaux de gris renferme 256 teintes de gris, allant de 0 à 255. Cet intervalle de valeur signifie que chaque pixel est codé sur 8 bits. 256 niveaux de gris suffisent pour la reconnaissance de la plus part des objets d'une scène. Une image couleur peut avoir trois couleurs RVB (Rouge, Vert, Bleu) ou quatre couleurs, CMJN (Cyan, Magenta, Jaune, noir).

Les images sont divisées généralement en deux types :

3.1 Image vectorielle

Dans une image vectorielle les données sont représentées par des formes géométriques simples qui sont décrites d'un point de vue mathématique. Par exemple, un cercle est décrit par une information du type (cercle, position du centre, rayon). Ces images sont essentiellement utilisées pour réaliser des schémas ou des plans.

3.2 Image matricielle

Une image matricielle est formée d'un tableau de points ou pixels. Plus la densité de point est élevée, plus le nombre d'informations est grand et plus la résolution de l'image est élevée. Corrélativement la place occupée en mémoire et la durée de traitement seront d'autant plus grandes. Les images vues sur un écran de télévision ou une photographie sont des images matricielles. On obtient également des images matricielles à l'aide d'un appareil photo numérique, d'une caméra vidéo numérique ou d'un scanner **[Site 1]**.

4 Formats d'image

Un format d'image est une représentation informatique liée à la manière dont l'image est encodée et décoder et manipuler. La plupart des formats incluent un en-tête qui contient les propriétés (dimensions de l'image, type d'encodage, etc.) et les données d'image. La structure de chaque propriété et donnée est différente Format d'image. On peut citer parmi les formats d'image:

BMP (Bitmap) : Le format BMP est le format par défaut du logiciel Windows. C'est un format matriciel. Les images ne sont pas compressées. Son logiciel d'origine.

EPS : matriciel n'est pas très différent du EPS vectoriel. En fait seules les données contenues dans le fichier sont différentes. Ainsi un logiciel de retouche de photos tel que Photoshop permet l'importation, la modification et l'exportation de fichiers en format EPS.

GIF (GraphicalInterchange Format) : Le format GIF est un format qui a ouvert la voie àl'image sur le World Wide Web. C'est un format de compression qui n'accepte que les images en couleurs indexés codé sur 8 bits, C'est un format qui perd beaucoup de son marché suite à une bataille juridique concernant les droits d'utilisation sur Internet.

JPEG (Joint Photographique Experts Group) : Les images JPEG sont des images de 24 bits. C'est-à dire qu'elles peuvent afficher un spectre de 16 millions de couleurs. C'est la meilleure qualité d'images disponible.

PCX : Le format PCX est utilisé par le logiciel Paintbrush sous Windows. C'est un format matriciel.

TIFF (Tagged Image File Format): Le format TIFF, conçu à l'origine par la compagnie Aldus est un format matriciel. Conçu au départ pour n'accepter que les images en RGB, ce format permet de coder des images CYMK.

PBM (Potable BitMap) : Ce format permet de stocker des images en noir et blanc

PGM (Potable GrayMap) : Le format pgm permet de représenter des images en niveaux de gris dont les pixels ont des valeurs entières comprises entre 0 (noir) et 255 (blanc). La valeur de chacun des pixels est enregistrée dans le fichier au format ASCII.

PPM (Potable PixMap): Le format ppm concerne les images couleurs. Chaque pixel a pour valeur un triple (R, G, B) composé d'une composante rouge, verte et bleue.

5 Caractéristiques d'images

Les images numériques sont également définies par des propriétés appelée les caractéristiques globale. C'est propriété permettent généralement de caractériser l'aspect visuel des images.

Les images numériques ont plusieurs caractéristiques de base, se sont :

5.1 Dimension

La dimension d'une image est par définition la taille de l'image. Les dimensions d'une image numérique se représente sous forme de matrice dont les éléments sont des valeurs numériques représentatives des intensités lumineuse . La multiplication du nombre de ligne par le nombre de colonnes de la matrice nous donne le nombre total de pixels dans une image

5.2 Pixel

Le pixel est une valeur numérique représentative des intensités lumineuses. Son nom provient de la locution anglaise picture élément, qui signifie, « élément d'image » ou « point élémentaire ».

Chaque pixel a une couleur et l'ensemble des pixels forment l'image. La couleur d'un pixel peut être stockée en utilisant une combinaison de rouge , de vert et de bleu (RVB), mais d'autres combinaison sont également possibles, comme le cyan , le magenta , le jaune et le noir (CMJN).

5.3 Texture

Une texture est un champ de l'image qui apparait comme un domaine cohérent et homogène, c'est -à-dire formant un tout pour un observateur. C'est cette propriété de cohérence par l'œil humain qui sera recherchée le plus souvent par le traiteur des images, dans le but d'isoler les textures, soit pour segmenter l'image, soit pour reconnaitre des régions. **[2]**

5.4 Résolution

La résolution de l'image est le nombre de pixels par unités de longueur. Elle s'exprime en ppp (pixel par pouce) ou dpi (dot per inch).Un pouce mesure 2.54cm. La résolution permet ainsi d'établir le rapport entre la définition en pixels d'une image et la dimension réelle de sa représentation sur un support physique (affichage écran, impression papier…) **[3].**

5.5 Bruit

Un bruit 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 **[4].**

5.6 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. Il permet de donner un grand nombre d'information sur la distribution des niveaux de gris (couleur) et de voir entre quelles bornes est repartie la majorité des niveaux de gris (couleur) dans le 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 **[5].**

5.7 Luminance

C'est le degré de luminosité des points de l'image, pour un observateur lointain, le mot luminance est substitué au mot brillance. Elle correspond à la sensation visuelle de luminosité d'une surface. C'est l'énergie lumineuse émise d'une surface, dans une direction particulière.

5.8 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 L1 et L2 sont les degrés de luminosité respectivement de deux zones voisines A1 et A2 d'une image, le contraste C'est le rapport entre L1 et L2

6 Système de traitement d'image

Le traitement d'image est l'ensemble d'opérations qui permettent l'amélioration (filtrage, rehaussement de contraste…), la modification (rotation, symétrie …) et l'extraction des l'information à partir des images numériques.

Le traitement d'image peut être considérer comme un système qui permet d'améliorer la qualité d'image par des logiciel tel que Photoshop[6]

7 Filtrage

On peut scinder les filtres en deux grandes catégories :

7.1 Filtres 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),

. Le filtrage linéaire est un produit de convolution c'est à dire une combinaison linéaire du voisinage du pixel concerné. Les différents types des filtres linéaires sont :

7.1.1 Filtre moyenné (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'accentue). Le filtre moyen est un filtre passe-bas, il laisse passer les basses fréquences (les faibles changements d'intensité de l'image) et atténue les hautes fréquences (variations rapides). **[7]**

1	1	1
1	1	1
1	1	1

x1/9

Avant filtrage Après filtrage

Figure 1: Application du filtre moyen (lissage)

7.1.2 Filtre gaussien

L'expression gaussienne en deux dimensions est donnée par :

$$G_0(x,y) = \frac{1}{2\pi\sigma^2}\, e^{-\frac{x^2+y^2}{2\sigma^2}}$$

L'avantage du filtre gaussien est de faciliter le contrôle du degré de filtrage grâce au paramètre σ. Ce type de filtrage priorise grandement les pixels rapproché du pixel central, et priorise proportionnellement moins les pixel éloigné du pixel central. Cependant, comme le filtre moyenner, le filtre gaussien a le désavantage de dégrade les contours. La discrétisation de ce filtre pour un σ = 0.6 donne le masque suivant :

1	2	1
2	4	2
1	2	1

x1/16

7.1.3 Filtre rehausseur de contours

Les filtres rehausseur de contours sont des filtre passe haut. C'est filtre mettent donc en évidence les changement rapide de l'intensité de l'image (changement qui arrive en général au niveaux des contours) et laissera les zone uniformes inchangées (généralement les zone entre les contours ayant de basse fréquence).

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

Figure 2: Exemple d'utilisation de filtre rehausseur de contours [7]

7.2 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).

7.2.1 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. **[7]**

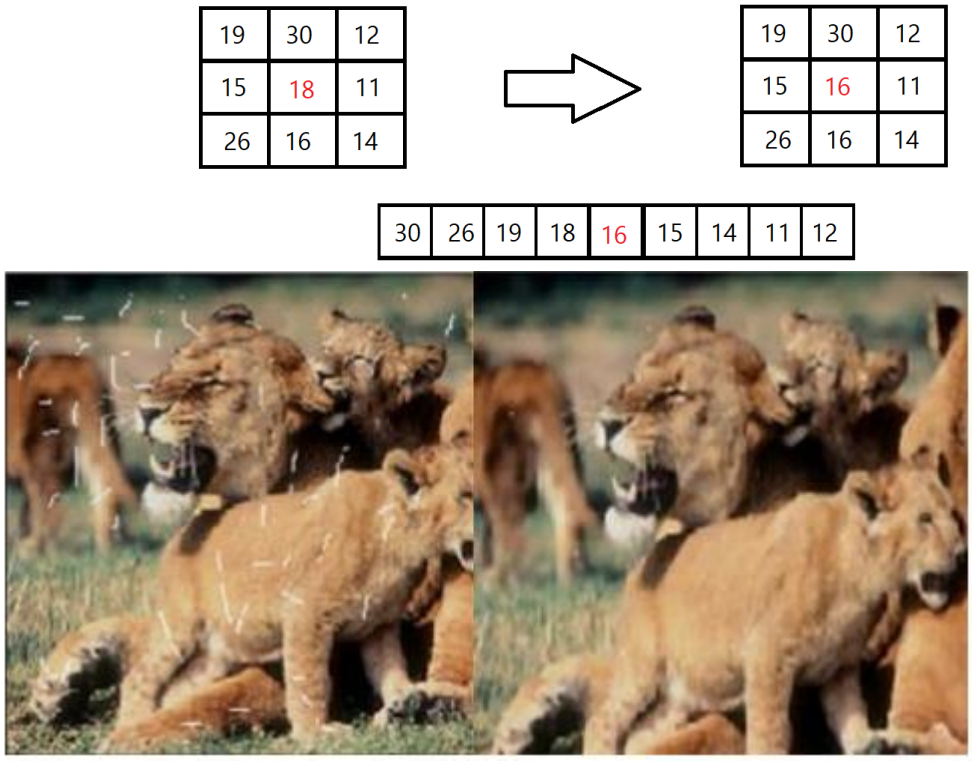
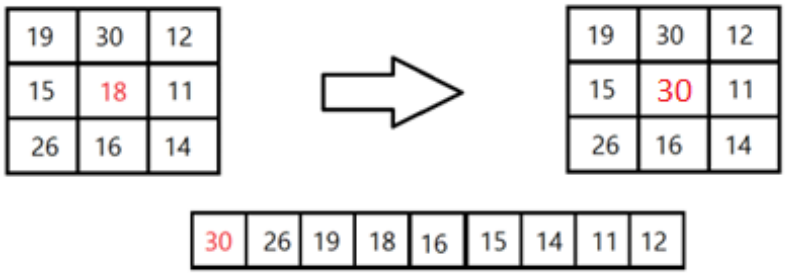


Figure 3: Exemple du filtre médian

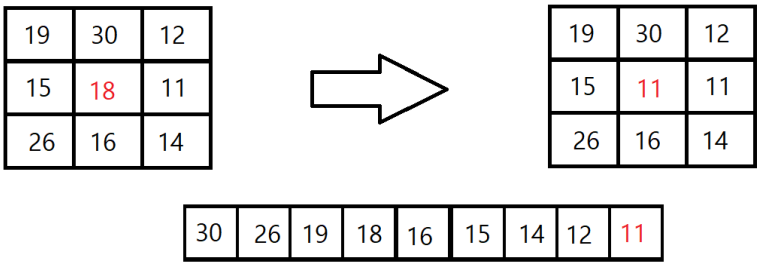
7.2.2 Filtre maximum

Ce filtre utilise le même principe que le filtre médian ; mais la valeur choisie est la valeur maximale.



7.2.3 Filtre minimum

Ce filtre utiliser le même principe que le filtre médian et le filtre maximum ; mais la valeur choisie est la valeur minimal.



8 Segmentation

Dans le traitement d'image, les opérations segmentation d'image sont des opérations constants à regrouper des pixel celons des critère prédéfinis. Ses ensembles de pixel forment ainsi des régions et ces régions forment une partition de l'image. L'opération de segmentation peut par exemple servir à retirer le fond d'un image. Quand il y a deux classes sur l'image, on appel cet opérations une binarisation.

L'homme, grâce à son cerveau (plus précisément ses réseaux de neurone) peut distinguer les objets sur une image grâce à des connaissance de haut niveau (compréhensions des différences entre objets et scènes).

Mettre au point des algorithme de segmentation de haut niveau est encore un des thèmes de recherche couramment traité dans le traitement d'image. La segmentation étant une importante étape dans le processus de traitement d'image, de nombreuse méthode de segmentation efficace ont été développer :

- Segmentation basée sur la région : Elles comprennent la croissance des régions, la division et la fusion split et marge.
- Segmentation basée sur les bords.
- Segmentation basée sur la classification ou le seuillage des pixels en fonction de leur intensité.
- Segmentation basée sur la coopération entre les trois premières segmentations

9 Conclusion

Le traitement des images numériques est un pilier principal pour la compréhension de tout systèmes CBIR. Les caractéristiques extraits des images grâce au traitement d'image nous serviront grandement à différencier les image.

Dans ce chapitre, nous avons présenté quelque notions de base du traitement des images numériques. Nous avons défini les composantes et caractéristiques des images, et nous avons développé une connaissance générale sur le filtrage et la segmentation.

Dans le chapitre suivant, nous nous intéresserons aux systèmes CBIR. Nous verrons quelque types de requêtes CBIR, et quelque notion de base sur la recherche d'image en particulier.

Chapitre II : Systèmes de recherche d'image par le contenu CBIR

1 Introduction

Dans ce chapitre, nous développerons les notions d'utilisation des images numériques dans les systèmes CBIR. Nous parlerons de l'architecture générale des systèmes CBIR, des bases de données populaires et reconnue pour les tests de systèmes CBIR. Nous nous attarderons sur les différents types de requêtes. Nous expliquerons les méthodes de mesure de performance dans les systèmes CBIR.

2 Recherche d'Images par le Contenu

La recherche d'image est un ensemble de techniques qui visent à fournir une description plus complète des images que de simples descripteurs de bas niveau. Ces techniques se concentrent sur l'identification du contenu sémantique des images (présence d'objets, de personnes, de certains concepts) plutôt que sur leur aspect visuel. La description sémantique des images implique l'utilisation de mots pour décrire les images à la place ou en plus des descripteurs de bas niveau [15]. Raisonner au niveau sémantique signifie que l'analyse de l'image se fait en termes d'objets, de contenu et de structure, et pas seulement en termes de statistiques sur les couleurs, les textures ou d'autres caractéristiques de base de l'image. Cela nécessite une certaine quantité d'informations supplémentaires pour la méthode, puisque par définition, seules les caractéristiques de base sont directement disponibles dans l'image. La sémantique elle-même n'est pas inscrite dans l'image, mais se trouve ailleurs. Nous devons donc chercher ces sources externes qui nous donnent accès aux clés du décodage sémantique de l'image. Nous soutenons que la sémantique exprimée dans une image dépend de deux éléments:

- Le niveau de connaissance et de perception que l'observateur a de cette image.
- le but que l'utilisateur de cette image a en tête lorsqu'il la regarde. Ces sémantiques doivent être trouvées selon deux approches :
- Une approche basée sur les méthodes pour comprendre le but de l'utilisateur, le sens de sa requête.
- Une approche basée sur les moyens pour connecter (ou lier) la connaissance sémantique humaine et l'apparence de l'image[16].

3 Architecture des systèmes d'indexation par contenu

Architecture générale d'un système d'indexation et de recherche d'images par le contenu Deux aspects indissociables coexistent dans les systèmes de recherche d'images par le contenu, l'indexation et la recherche. [18]

- **La phase d'indexation (hors-Ligne) :** Dans cette phase, des caractéristiques sont automatiquement extraites à partir de l'image et stockées dans un vecteur numérique appelé descripteur visuel. Grâce aux techniques de la base de données, on peut stocker ces caractéristiques et les récupérer rapidement et efficacement.

- **La phase recherche (On-line) :** Dans cette étape, le système analyse une ou plusieurs requêtes émises par l'utilisateur et lui donne un résultat correspondant en une liste d'images ordonnées, en fonction de la similarité entre leur descripteur visuel et celui de l'image requête en utilisant une mesure de distance. La figure 4 schématise le fonctionnant d'un système de recherche et d'indexation d'images.

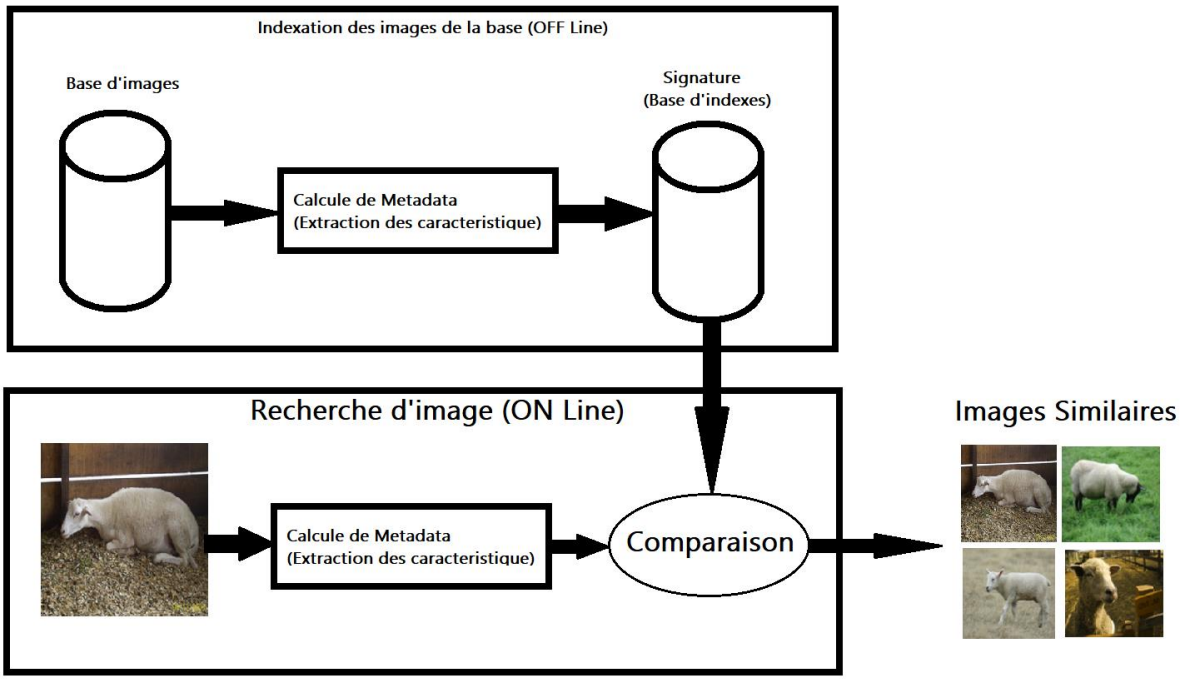


Figure 4: Le fonctionnant d'un système de recherche et d'indexation d'images.

4 Bases d'images

La gestion de bases de données désigne la branche de l'informatique qui étudie le stockage et l'interrogation des données numériques. Une base de données informatique est donc un ensemble d'informations numériques stockées selon un modèle dans le but de les conserver, de les enrichir et de les interroger avec la garantie de l'intégrité de ces données [10].

Une base d'images est une collection d'images stockées sur un dispositif informatique sous la forme d'une série de bits. Elle peut contenir des dizaines, des centaines ou même des millions d'images bien structurées. Le dispositif comporte un système de gestion de base qui régit la collecte, le stockage, le retraitement et l'utilisation d'images. Généralement, les systèmes de recherche d'images incluent deux bases d'images différentes, la première collecte les images brutes, c'est-à-dire les images non traitées, et la deuxième gère les images indexées.

Ils existent plusieurs bases d'images sur Internet dont la plupart sont librement utilisées. Différentes par leur contenu, chaque base possède des images groupées en plusieurs classes bien définies où chaque image n'appartient qu'à une seule classe. Généralement, les développeurs utilisent ces bases pour tester les algorithmes de recherche d'images mis en œuvre afin d'évaluer et valider leurs systèmes [8].

4.1 Base inria

La base inria est un ensemble d'images photos personnelles de vacances. Cette base contient aussi des versions d'images changées à travers la rotation, le point de vue et d'illumination, flou, etc.

L'ensemble de données contient 500 groupes d'images, qui représentent une scène ou un objet distinct. La première image de chaque groupe est l'image de requête et les résultats corrects de récupération sont les autres images du groupe. La taille totale du corpus est : 1491 images au total : 500 questions et le reste ce sont les images de l'index . Les images de cette base ont trois résolutions différentes[9].

4.2 Base wang

La base d'images de Wang est un sous-ensemble de la base d'images Corel. 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/ La base originale contient 1000 images naturelles en couleurs, divisées en 10 classes, chaque classe contient 100 images. L'avantage de cette base est de pouvoir évaluer les résultats [10]

(La figure 5) présente un exemple de chaque classe

Figure 5 : 10 classes de la base de Wang [10]

4.3 Base COIL (Columbia Object Image Library)

4.3.1 Contexte

COIL-100 a été recueilli par le Centre de recherche sur les systèmes intelligents du Département d'informatique de l'Université de Columbia. La base de données contient des images en couleur de 100 objets. Les objets ont été placés sur un plateau tournant motorisé sur fond noir et les images ont été prises à des poses internes de 5 degrés. Cet ensemble de données a été utilisé dans un système de reconnaissance de 100 objets en temps réel dans lequel un capteur du système pouvait identifier l'objet et afficher sa pose angulaire[11].

Figure 6: Exemple d'image de la base COIL-100 [11]

4.3.2 Teneur

Il y a 7 200 images de 100 objets. Chaque objet a été tourné sur 360 degrés pour faire varier la pose de l'objet par rapport à une caméra couleur fixe. Les images des objets ont été prises à des intervalles de pose de 5 degrés. Cela correspond à 72 poses par objet. Ces images ont ensuite été normalisées en taille. Les objets ont une grande variété de caractéristiques géométriques complexes et de réflectance [4].

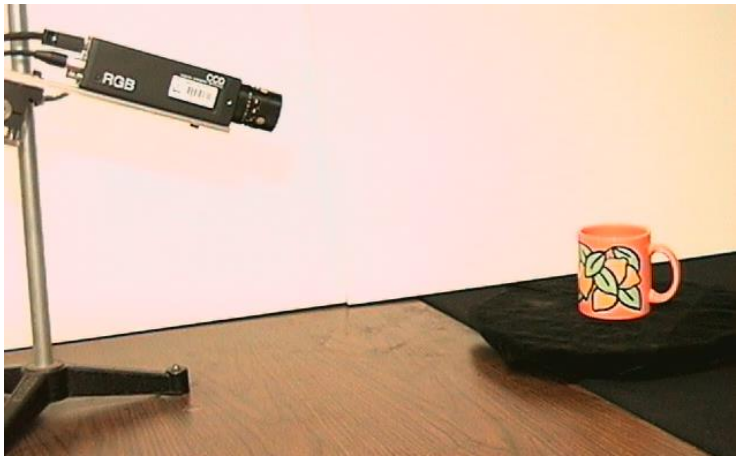


Figure 7: Méthode de cartographiassions des objets sur COIL-100[11]

4.4 La base Pascal VOC (Visual Object Classes) 2012

Cette base d'image contient environ 20 classe d'objet : les véhicules, les ménages, les animaux et autres : avion, vélo, bateau, bus, voiture, moto, train, bouteille, chaise, table à manger, plante en pot, canapé, TV/moniteur, oiseau, chat, vache, chien, cheval, mouton et personne. Chaque image de cet ensemble de données est annoter a des classes d'objets, des segmentions sémantique et des segmentation de classification. Cet ensemble de données est l'une des plus populaire pour les taches de détection d'objet, de segmentation sémantiques et de segmentation de classification. L'ensemble de données PASCAL VOC est divisé en trois sous-ensembles : 1 464 images pour la formation, 1 449 images pour la validation et un ensemble de test privé.

5 Indexation

L'indexation a pour but de substituer à une image un représentant (ou descripteur) moins encombrant qui la caractérise le mieux possible et de ne travailler que sur ce modèle lors de la recherche. Cela permettra une meilleure organisation des données, de limiter la quantité de données examinées durant une recherche, d'y accéder rapidement et de confiner la recherche au maximum [18]. Un système d'indexation comprend généralement deux phases de traitement :

5.1 Indexation logique

L'indexation logique consiste à extraire et à modéliser les caractéristiques de l'image qui sont principalement la forme, la couleur et la texture. Chacune de ces caractéristiques pouvant être considérée pour une image entière ou pour une région de l'image.

5.2 Indexation physique

L'indexation physique consiste à déterminer une structure efficace d'accès aux données pour trouver rapidement une information. De nombreuses techniques basées sur des arbres (arbre-B, arbre-R, arbre quaternaire,…) ont été proposées. Pour qu'un système de recherche d'images soit performant, il faut que l'indexation logique soit pertinente et que l'indexation physique permette un accès rapide aux documents recherchés.

6 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 [17].

7 Requêtes

La plupart des systèmes de recherche d'images dispose d'un interface graphique permettant aux utilisateurs d'accéder au moteur de recherche par le biais d'une requête. Mais cela pose le problème de la facilité pour l'utilisateur de définir précisément ses besoins à travers cette interface. Selon le cas, l'utilisateur peut spécifier directement les attributs de bas niveau de l'image cible dans sa requête, interroger le système en esquissant un croquis, ou bien en présentant au système une image exemple de ce qu'il recherche [12].

7.1 Requête par mots clés

La plupart des systèmes de recherche d'images développés utilisent des mots clés ou des descripteurs textuelles pour caractériser chaque image de la base (ex : recherche d'images sur Internet). Ces images sont recherchées suivant un ou plusieurs critères, par exemple trouver les images contenant 80% de rouge. Donc, le système se base sur l'annotation manuelle et textuelle d'images. Beaucoup de moteurs de recherche d'images tels que Google, Yahoo…utilisent cette façon. Cette méthode n'est pas parfaite parce que quelques mots qui n'expriment pas le sens d'une image.

(La figure 8) donne un exemple, l'utilisateur veut trouver des images contiennent une (des) voiture(s) avec le ciel cependant les premières images résultats ne sont pas pertinentes car ils n'ont ni les voitures ni le ciel. De plus l'indexation de ces images représente une tâche longue et répétitive pour l'humain, surtout avec les bases d'images qui deviennent aujourd'hui de plus en plus grandes. Elle est subjective à la culture, à la connaissance et aux sentiments de chaque personne[13].

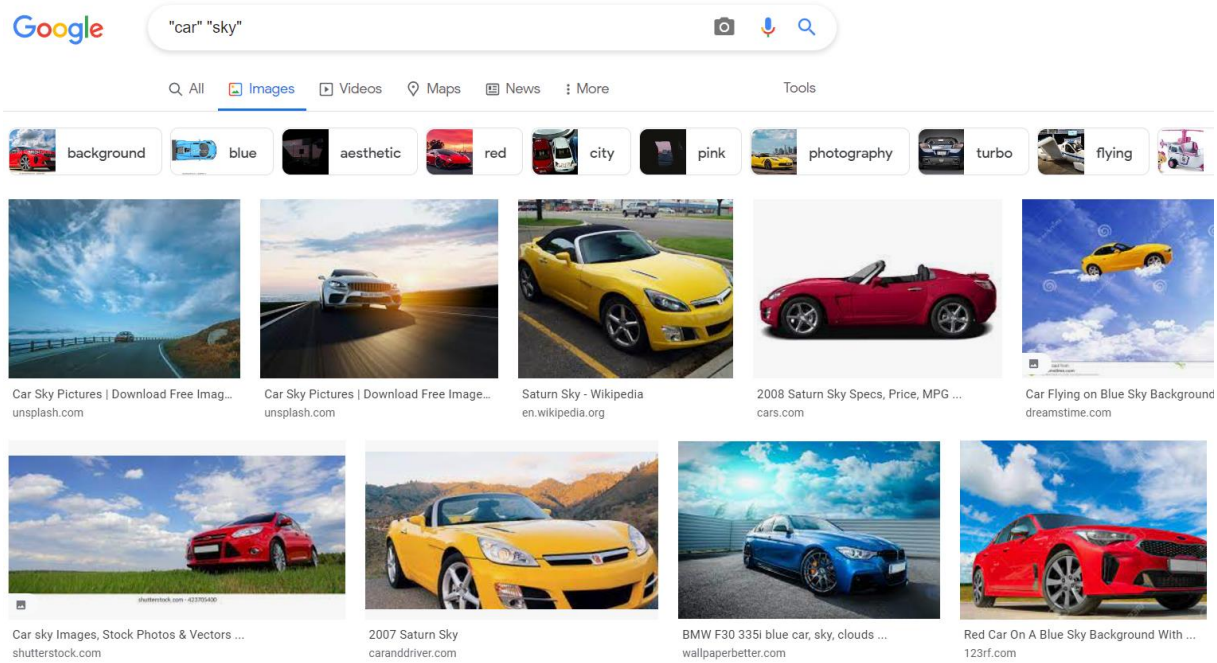


Figure 8: Un exemple de recherche d'images par mot clé dans Google

Puisque la requête par mots clés a des désavantages, les systèmes de recherche d'images par le contenu utilisent souvent les trois autres méthodes de requêtes.

7.2 Requête par esquisse

Puisque les deux méthodes précédentes ont des désavantages, les chercheurs développent d'autres approches pour renforcer les moteurs de recherche d'images tel que la requête par esquisse ou bien par crayonnage. Dans cette approche, le système propose à l'utilisateur un ensemble d'outils de dessin et une palette de couleurs qui lui permet de spécifier sa requête en forme d'un sketch. Après, le système calcule les ressemblances entre le dessin et les images de sa base [14].

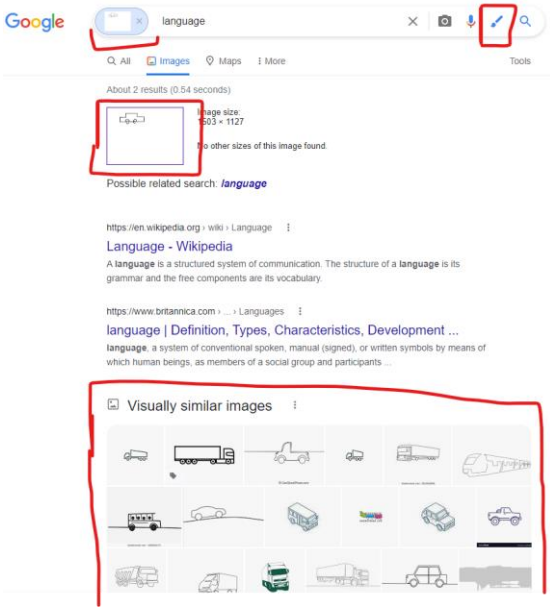


Figure 9: Un exemple de recherche d'image par esquisse (Recherche par esquisse)

7.3 Requête par le contenu

Cette technique se base sur l'identification des images à partir de leur contenu, et pas du texte associé aux images. L'indexation des images, et leurs paramètres qui sont la couleur, la texture, l'intensité ou, encore, les formes contenues dans l'image doivent être extraits au préalable.et ils fournissent une "signature" [19].

Exemple :

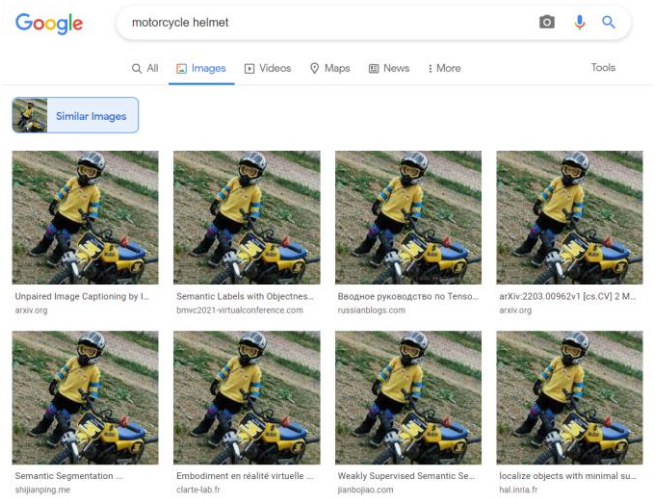


Figure 10: Exemple de recherche d'image par une image requête

(Remarque : les mots clés ont été automatiquement remplie après la recherche)

7.4 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 [17].

8 Analyse de la requête

Le but de cette étape est de convertir la demande de l'utilisateur afin qu'elle soit comparable au index de la base d'images ; par conséquent, il s'agit généralement d'extraire le même type de descripteurs que les descripteurs extraits de la base d'images lors de l'indexation.

9 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 [8] 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 [17].

10 La présentation des résultats

Les résultats de la requête sont affichés sous forme de listes d'images (réduite en vignettes), et ces listes sont classées par ordre décroissant de pertinence. Par rapport aux documents textuels, l'avantage des images est que vous pouvez voir l'intégralité du document en un coup d'œil, ce qui vous permet de visualiser un grand nombre de résultats et d'effectuer des comparaisons plus rapidement. Comme mentionné ci-dessus, la présentation des résultats est généralement associée à la possibilité d'interaction, ce qui permet ,par exemple, d'optimiser la requête en indiquant des résultats pertinents et non pertinents(cycles pertinents) au système, permettant ainsi une reformulation automatique de la requête. [19]

11 Mesures de performance de système CBIR

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[19].

Le rappel est défini par le nombre d'image pertinents retrouvés au regard du nombre d'image pertinents que possède la base d'image. Le rappel est calculé comme suit :

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

La précision est le nombre d'image pertinents retrouvés rapporté au nombre d'image total proposé pour une donnée. L'expression de la précision :

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

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 11):

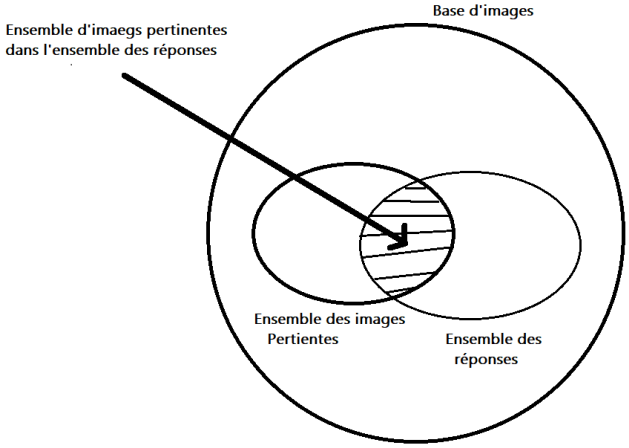


Figure 11: Le rappel et la précision pour une requête (Yates, 1999)

L'exemple présenté dans (la figure 11) illustre le calcul de rappel et précision.
Si la liste des images pertinentes présente dans la base de donné pour la recherche d'un avion est:

Calcul de Rappel et précision pour cette recherche :

Figure 12: Exemple de calcul rappel et précision

- Rappel=4/7= 57,14% (4 image pertinent afficher, 7 présent dans la base d'image)
- Précision =4/6= 66,66% (4 image pertinent afficher, 6 image afficher)

12 Représentation des images dans un CBIR

Généralement, les systèmes de recherche d'images par contenu utilisent des descripteurs de bas niveau pour représenter les images traitées. Chaque image en état brute contient une énorme quantité d'informations, pour pouvoir l'utiliser dans un CBIR, il faut réduire cette quantité en gardant seule les petits détails qui seront utilisable.

Un descripteur de bas niveau est un ensemble de valeurs extraites directement et qui caractérisent l'image. L'extraction de descripteurs de bas niveau représente une première abstraction par rapport à l'image brute. Elle constitue la perception du système, dans la mesure où les descripteurs extraits sont la seule information conservée. Décrire une image par des descripteurs de bas niveau est un problème difficile. La sémantique « contenue dans les pixels de l'image » n'est absolument pas accessible directement par une machine alors qu'elle apparaît de manière évidente à un humain qui voit simultanément l'ensemble des pixels. Associer une description à un ensemble de pixels requiert en effet de nombreux processus et surtout, une quantité énorme de connaissances [12].

Les trois modules les plus utilisés pour la construction d'un descripteur de bas niveau sont la couleur, la texture et la forme. D'autres systèmes ajoutent même les points d'intérêt comme un paramètre de bas niveau.

Nous parlerons de cela plus en détail dans le chapitre 3

13 Conclusion

Nous avons vu dans ce chapitre les notions générales des systèmes CBIR. Nous avons présenté leur architecture, les types de requête utilisé, ainsi que les bases de données généralement utilisé pour mesurer les performances de tous systèmes CBIR.

Dans le prochain chapitre nous parlerons plus en détail de méthodes utilisées par les systèmes CBIR pour caractériser deux images. Des méthodes d'extraction et de comparaison de ses caractéristiques.

Chapitre III : Méthodes de caractérisation d'un système CBIR et mesure de similarité

1 Introduction

Les caractéristiques visuelles de bas niveaux tel que la couleur, forme, texture, disposition spatial… sont appelé dans le CBIR des « Descripteur d'image ». Ces caractéristiques de bas niveau sont extraites et mise en correspondance selon la similarité avec des caractéristiques requêtes. Query-By-Image (QBIC) et SIMPlicity sont des modèles populaire de récupération d'image basé sur les caractéristiques de bas niveau. Dans cette section du chapitre, nous parlerons des différents descripteurs de bas niveaux avec des exemples et des informations utiles.

2 Descripteur de Couleur

La couleur est l'une des caractéristiques les plus importants pour mesurer la similarité entre deux images. Il n'est donc pas surprenant de remarquer que c'est le descripteur de bas niveau le plus utilisé dans la récupération d'images. De même la répartition de couleur sur une image est stable et a peine affecté par la translation, la rotation et la déformation. Sa définition et son efficacité sont liés à l'espace colorimétrique utilisé par l'image.

La distribution de couleur est une caractéristique très utilisée pour la représentation d'images. Il s'agit vraisemblablement du descripteur le plus répandu en indexation. Il est en théorie invariant aux translations et rotations, et change seulement légèrement en cas de changements de la prise de vue ou de l'échelle. La distribution de couleur est habituellement estimée à l'aide d'un histogramme, mais il existe d'autres descripteurs [20]

2.1 Espace colorimétrique

L'espace colorimétrique, également appelé espace de couleur ou espace chromatique est une méthode de présentation des couleurs selon des coordonnées généralement à 3 dimensions. Les couleurs possibles dans cet espace sont réparties sur un espace volumique (donc à 3 dimension). Ces 3 coordonnées (lié chacune à une dimension), permettent de retrouver une couleur de manière précise sur cet espace. Chaque couleur contenue dans un espace colorimétrique peut ainsi être associée à des coordonnées déterminant un point précis permettant de le retrouver.

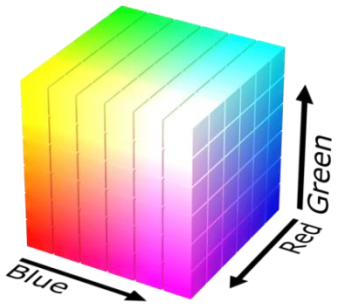
L'espace colorimétrique de la CIE (Commission International de l'Eclairage) arrive à représenter l'ensemble des couleurs perceptibles par l'œil humain (8 millions de couleur). Elles sont donc un étalon international de la colorimétrie et servent de référence en raison de leur déclinaison graphique (Tel que CIE XYZ, CIE L*a*b).

Les espace les plus communément utilisés dans les CBIR sont : l'espace RGB, XYZ, Cie L*a*b, L*u*v, HSV. Le choix de l'espace colorimétrique peut être déterminant pour l'efficacité du descripteur de couleur. Par exemple la projection de l'image dans l'espace HSV permet de séparer les informations relatives à la teinte, la saturation et l'intensité [21]. Il a été démontré que la teinte est mieux invariante aux conditions d'éclairage et de prise de vue [22]. D'autres espaces également fréquents dans le domaine revendiquent d'être perpétuellement uniformes et indépendant de l'intensité telle que CIE, XYZ, et CIE-LUV [23]. Là encore ce sont des modèles de représentations et il n'existe pas un espace de couleur

idéal. On trouve une comparaison entre les espaces de couleurs ainsi que leurs caractéristiques et une analyse avantages/inconvénients dans [24]. Un histogramme HSV [25] est largement utilisé dans le CBIR, offrant des images de meilleure couleur que les images en niveaux de gris. Comme la vision humaine est plus sensible à la luminosité qu'à la couleur intensité, le système visuel humain utilise souvent l'espace colorimétrique HSV, qui est plus en ligne avec des caractéristiques visuelles humaines que l'espace colorimétrique RVB, pour faciliter le traitement des couleurs et la reconnaissance **[26].**

2.1.2 Système RGB

L'espace colorimétrique RGB ou standard RGB (Red-Green-Blue, Rouge-Vert-Bleu) est l'espace de couleur utilisé sur les appareils numérique. En 1999 la CIE définie le RGB comme « un espace chromatique commun pour le stockage ». L'espace colorimétrique RGB est très utilisé pour la représentation de couleur. Ce modèle a pour avantage d'être très basique puis qu'aucun traitement n'est nécessaire. Les autres espaces de couleurs sont souvent calculé par référence à celui-là.



[Images by Michael Horvath, available under Creative Commons Attribution-Share Alike 3.0 Unported license]

Figure 13: Espace colorimétrique RGB

2.1.3 Système XYZ

L'espace colorimétrique XYZ a été définie en 1931 par la CIE (Commission Internationale de l'éclairage). Cet espace dérive de l'espace RGB, l'une possédant les coordonnée r, g, b et l'autre x, y, z. Les relations entre les grandeurs physiques et la perception étant linéaire, les trois grandeurs associées à une couleur constituent un espace vectoriel.

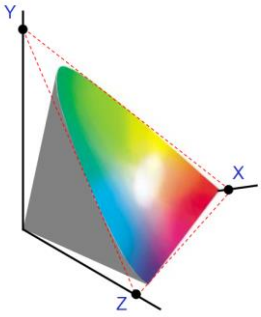


Figure 14: Espace colorimétrique XYZ [Site 2]

2.1.4 Système L*u*v

L'espace colorimétrique L*u*v définie en 1976 par la CIE est l'espace utilisé pour la caractérisation des écrans défini par la Commission. Elle sera définie au même moment que l'espace CIE L*a*b pour la caractérisation des surfaces.

2.1.5 Système YCrCb

Le modèle YCbCr ou plus précisément le modèle Y'CbCr est une méthode de représentation de l'espace de chrominance. Y est a ne pas confondre avec Y', le premier étant le signal de luminance, le second le signal de luminance avec une correction gamma appelé luma.
Une image numérique (en noir, blanc ou en couleur) peut être créée grâce à la somme des couleurs qui la compose. Ainsi dans tous image représenté sous l'espace colorimétrique Y'CbCr, le signal Y' sera l'addition des couleurs rouge, bleu et vert comme observé dans sa formule ci-dessous.
Ainsi, un récepteur recevant : Y', le signal de luminance (noir et blanc), ainsi que les informations de chrominance Cb (Y' moins le bleu) et Cr (Y' moins le rouge). Peut obtenir le vert en utilisant l'équation mathématique Y' = 0.3R' + 0.6V' + 0.1B' avec Y', R', B' Connue.

$$Y' = 0{,}299\ R' + 0.587\ V' + 0{,}114B'$$
$$Cb = -0{,}1687\ R' - 0{,}3313\ V' + 0{,}5\ B' + 128$$
$$Cr = 0{,}5\ R' - 0{,}4187\ V' - 0{,}0813\ B' + 128$$

2.1.6 Système HSV

L'espace colorimétrique HSV (Hue, Saturation, Value) en français TSV (Teinte, Saturation Value) est l'un des espace les plus utilisé dans le traitement d'image afin de permettre une efficacité du descripteur tout en étant relativement simple. Hue ou Teinte est le point représentant la variation de la teinte (de la couleur), Value ou valeur représente la luminosité (la quantité de blanc ajouté a la couleur).

Hue (Teinte : Rouge, Vert, Jaune, Blue) :
Saturation (Saturation : Rose -> Rouge) :
Value (Luminosité : Sombre -> Clair) :

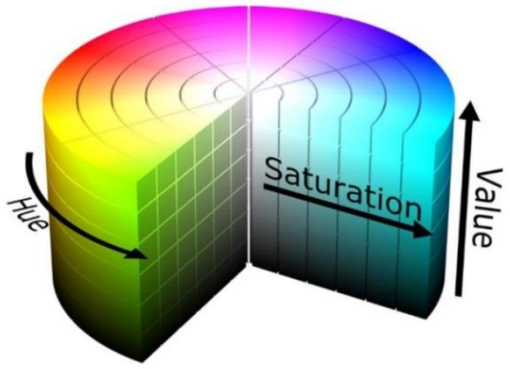


Figure 15: Espace colorimétrique HSV

2.2 Récapitulatif des espaces colorimétriques

Quel que soit l'objectif voulu, il faut toujours rechercher la représentation, i.e. l'espace couleur, qui sera le mieux adapté aux données et à l'algorithme que l'on souhaite utiliser [13].

Espace Couleur	Calcul Linéaire	Distance Uniforme	Avantages Inconvénients
RGB	Oui	Non	-Format de base -Nombreux algorithmes -Axes fortement corrélés
XYZ	Oui	Non	-Espace incontournable -Décomposition Luminance/Chrominance -Nécessite de connaître les conditions d'acquisitions
X1 X2 X3	Oui	Non	-Axes décorréler -Forte complexité algorithmique -Espace lié à l'image étudiée
i1 i2 i3	Oui	Non	-Approximation de la décorrélation -Calcul beaucoup plus rapide que X1X2X3 -Dépend des images sélectionnées pour le calcul de la matrice de passage
HSV	Non	Oui	-Séparation Luminance, Teinte et Saturation -Bonne corrélation avec la représentation humaine des couleurs -Extraction de la teinte -Transformation non-linéaire : création d'artéfacts numériques
Y I Q	Oui	Non	-Décomposition Luminance/Chrominance
L a b	Non	Oui	-Distance adaptée à la perception humaine -Décomposition Luminance/Chrominance -Temps de calcul important -Transformation non-linéaire : création d'artéfacts numériques -Nécessite de connaître les conditions d'acquisition

Tableau 1 : Récapitulatif des espaces colorimétriques [13]

2.3 Les modèles de caractérisation des couleurs

Nous appelons modèles de caractérisation, les modèles permettant d'utiliser la couleur comme descripteur optimale pour la recherche d'images par le contenue. Le modèles basé sur les histogrammes et celle basé sur des méthodes statique

La fusion de caractéristiques de couleur et de texture offre un ensemble de fonctionnalités vigoureux pour approches de récupération d'images en couleur. Les résultats obtenus des expériences (Wang et al. [26]) révèlent que la méthode proposée a récupéré images plus précises que les autres méthodes traditionnelles [20]

La figure suivante représente les trois catégories de modèles utilisés pour caractériser les attributs des données de l'image relatives à la couleur ; le modèle de l'histogramme, le

modèle statistique, et la méthode, MPEG-7 [27]. L'histogramme associe à chaque valeur d'intensité lumineuse le nombre d'échantillons (pixels) ayant cette valeur :

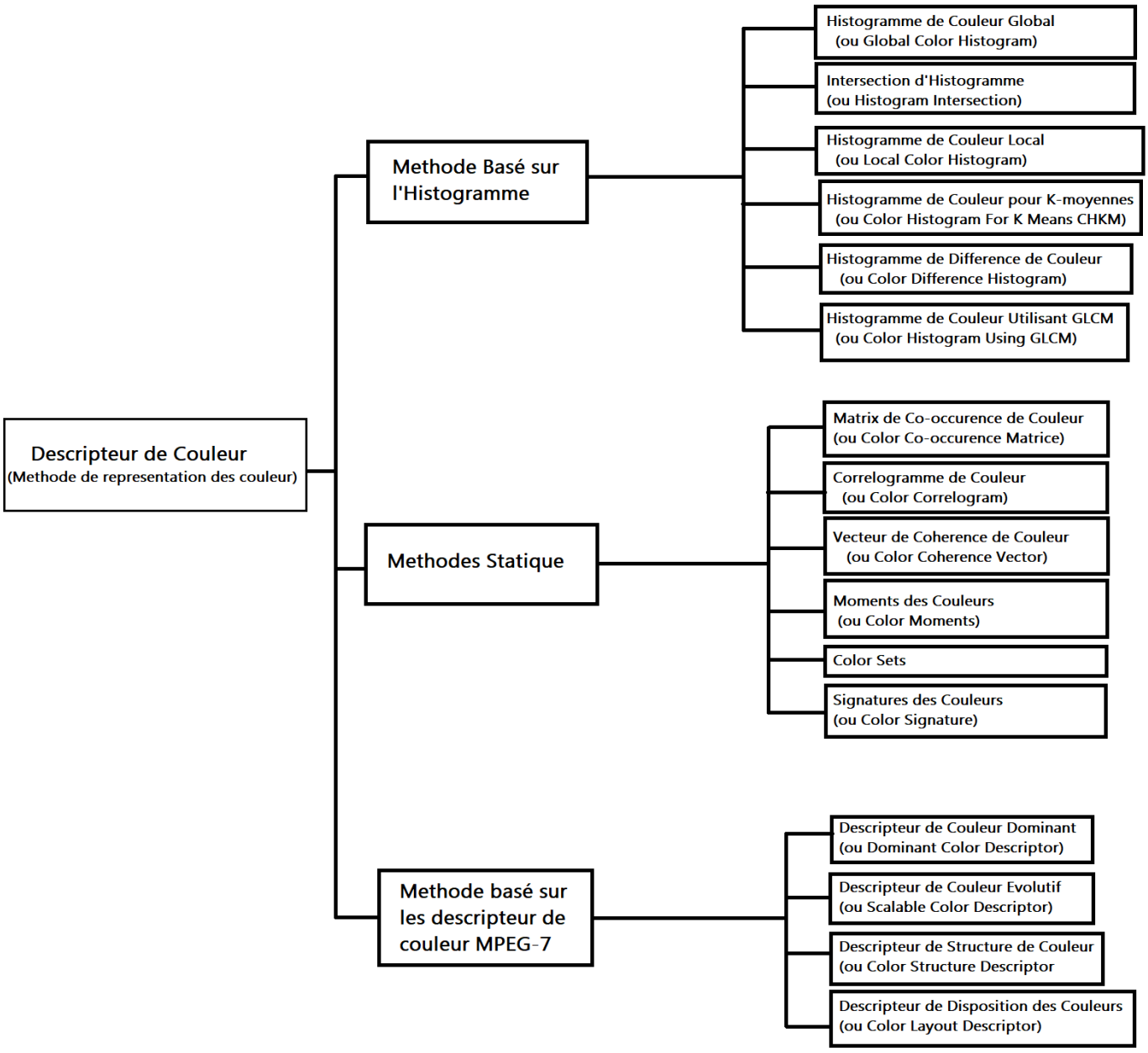


Figure 16: Les modèles de caractérisation des couleurs

Pour utiliser la caractéristique de couleur comme descripteur de couleur, de nombreuse technique sont utilisé. Une technique très utilisée pour la couleur est l'intersection d'histogrammes. Les histogrammes sont faciles et rapides à calculer, et robustes à la rotation et à la translation. On peut également utiliser les Descripteur de couleur dominante (DCD), Descripteur de couleur évolutif (SCD), Descripteur de structure de couleur (CSD), Descripteur de disposition des couleurs (CLD)

2.3.1 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 [39], 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 [17].

2.3.2 Moment Statique

La méthode d'histogramme utilise la distribution complète de la couleur. On doit stocker de nombreuses données. Au lieu de calculer la distribution complète dans les systèmes de recherche d'images, on calcule seulement des caractéristiques dominantes de couleur tel que l'espérance, la variance et d'autres moments [28]. Cette approche consiste à calculer l'espérance, l'équivalent d'une moyenne pondérée de tous les couleurs, puis la variance et les moments d'ordre 3 pour chaque composante couleur par les formules suivantes :

L'espérance : $\boldsymbol{E_i = \frac{1}{n} \sum_{j=1}^{n} P_{ij}}$

La variance : $\boldsymbol{\delta_i = (\frac{1}{N} \sum_{j=1}^{N} (P_{ij} - E_i)^2)^{\frac{1}{2}}}$

Moment d'ordre 3 : $\boldsymbol{S_i = (\frac{1}{N} \sum_{j=1}^{N} (P_{i,j} - E_i)^3)^{\frac{1}{3}}}$

P_{ij} représente la position du pixel, N représente le nombre total des pixels dans cette image. Dans [29], les auteurs ont prouvé que les méthodes des moments statistiques utilisées archent plus vite et donnent des résultats meilleurs que les méthodes d'histogrammes.

Ainsi, la distribution des couleurs dans l'image est caractérisée par quelques moments dominants comme l'espérance et la variance.

2.3.3 Histogramme de couleur

Statistiquement, un histogramme de couleur est un moyen d'approximer la probabilité conjointe des valeurs des trois canaux de couleur. La forme la plus courante de l'histogramme est obtenue en divisant la plage des données en tranches de taille égale. Il y a deux paramètre importante à déterminer dans lors pour un histogramme : La largeur des cases (qu'on appellera bacs) et les emplacements des cases. Ce n'est pas très difficile de voir que le choix de la largeur du bac a un énorme effet sur l'apparence de l'histogramme résultant. Choisir un tout petit bac largeur donne un histogramme dentelé, avec un bloc séparé pour chaque observation. Une très grande largeur de bac donne un histogramme avec un seul bloc. Les largeurs de bac intermédiaires conduisent à une variété de formes d'histogramme entre ces deux extrêmes. Les positions des bacs sont également importantes pour la forme des histogrammes. De petits déplacements des bacs peuvent entraîner un changement majeur de la forme de l'histogramme.
Puis pour chaque bac, le nombre de points du jeu de données (ici les couleurs des pixels dans une image) qui tombent dans chaque bac sont comptés et normalisés au nombre total de points, ce qui nous donne la probabilité qu'un pixel tombe dans cette case. Par souci de simplicité, étant donné une image couleur I(x, y) de taille X × Y, qui se compose de trois canaux I = (I_R, I_G, I_B), l'histogramme de couleur utilisé ici est :

$$h_c(m) = \frac{1}{XY} \sum_{x=0}^{X-1} \sum_{y=0}^{Y-1} \begin{cases} 1 \; si \; I(x,y) dans \; le \; bac \; m, \\ 0 \; si \; non \end{cases}$$

Où un bac de couleur est défini comme une région de couleurs.
Les régions de l'espace colorimétrique peuvent être définies de manière non paramétrée par algorithmes de clustering non paramétriques, ou simplement donnés par des frontières fixes dans certains espaces colorimétriques. Par exemple dans l'espace colorimétrique RVB, si nous divisons chaque canal R, G, et B en 8 intervalles égaux de longueur, nous aurons un histogramme de couleur 8 x 8 x 8 de 8 × 8 × 8 = 512 cases de couleur. Un exemple de l'apparence d'un histogramme de couleur est illustré [30 Adapter en français par 'Google Traduction']

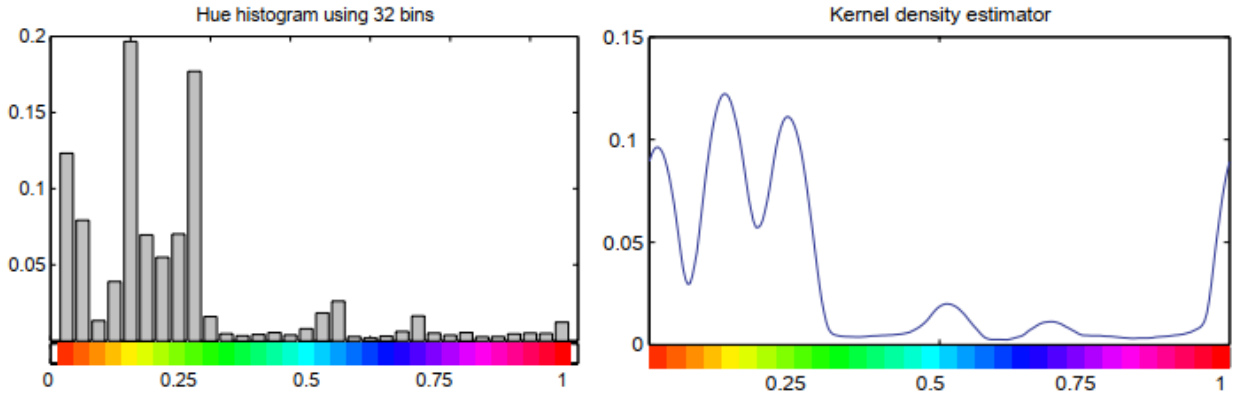


Figure 17: La distribution de la densité de teinte d'une image [30]

Dans une image espace colorimétrique R, V, B ; un histogramme de couleur est associé à chacune des composantes de couleur R, V, B. Les histogrammes d'images sont faciles à calculer mais présentent trois inconvénients majeurs. Premièrement l'histogramme ne nous

fournit pas l'emplacement des objets (l'information spatiale) du contenu dans une image. Deuxièmement, L'histogramme est sensible aux changements de luminosité, contraste et artefacts de compression. Puis finalement le calcul de la distribution complète des couleurs engendre la construction de grands vecteurs ce qui complique l'indexation [31].
Dans notre implémentation, nous utiliseront un descripteur de couleur en se basant sur le modèles des histogrammes, nous tenteront de corriger chacune de inconvénient cité précédemment. L'absence d'information spatiale sera corrigée en calculant plusieurs histogrammes, un pour chaque emplacement spatial. La sensibilité au changement de luminosité sera corrigée en utilisant l'espace colorimétrique HSV qui est peu affecté par les changements de luminosité.

3 Descripteur de Texture

De nombreuses définitions ont été proposées, mais aucune ne convient parfaitement aux différents types de textures rencontrées. Dans une définition couramment citée [32], la texture est présentée comme une structure disposant de certaines propriétés spatiales homogènes et invariantes par translation [17]. La texture est une description de l'arrangement spatial de la couleur ou des intensités dans une image ou une région sélectionnée d'une image [33]. Ci-dessous quelque exemple de textures :

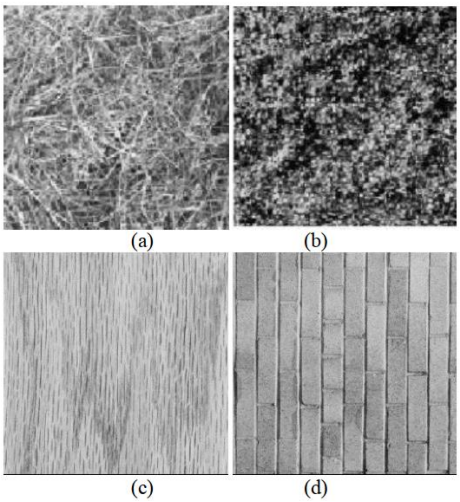


Figure 18: exemple de texture. (a) Herbe, (b) feuilles, (c) Bois, (d) Mur de brick [33]

On peut trouver la textures dans presque chaque surface, qu'elle soit naturel ou artificiel. La texture est lié à la sensation de touché que procure la surface d'un élément (sable, eau, murs, tissu). C'est un motif visuel homogène qui représente l'agencement structurel de la couleur ou l'intensité dans une image. Les caractéristiques de texture peuvent représenter les informations de structure spatiale interne des images. Le plus utilisé est le LBP [34], le LBP standard encode la relation entre une référence pixel et ses voisins environnants en comparant les valeurs de niveau de gris. [26]
Il existe un grand nombre de textures. On peut les séparer en deux classes: les textures structurées (macro textures) et les textures aléatoires (micro textures). Une texture qualifiée de structurée est constituée par la répétition d'une primitive à intervalle régulier (exemple : damier, mure de bique, damier, grain de café). Les textures qualifiées d'aléatoires se distinguent en général par un aspect plus fin (sable, herbe, etc.). Il y a certains aspects concernant les textures telles que la taille de granularité, directionnalité, caractère aléatoire ou régularité et éléments de textures. L'analyse de ces aspects peut être utile pour classer

les textures dans une ou plusieurs classes. L'analyse de texture est l'une des bases étapes dans les cas de vision par ordinateur tels que le suivi d'objets, reconnaissance visuelle des formes, reconnaissance faciale, détection de la peau, récupération d'images et etc. Depuis, de nombreuses approches ont été proposées pour résoudre ce problème.
Quelque modèle de caractérisation des textures sont présent dans la Figure ci-dessous :

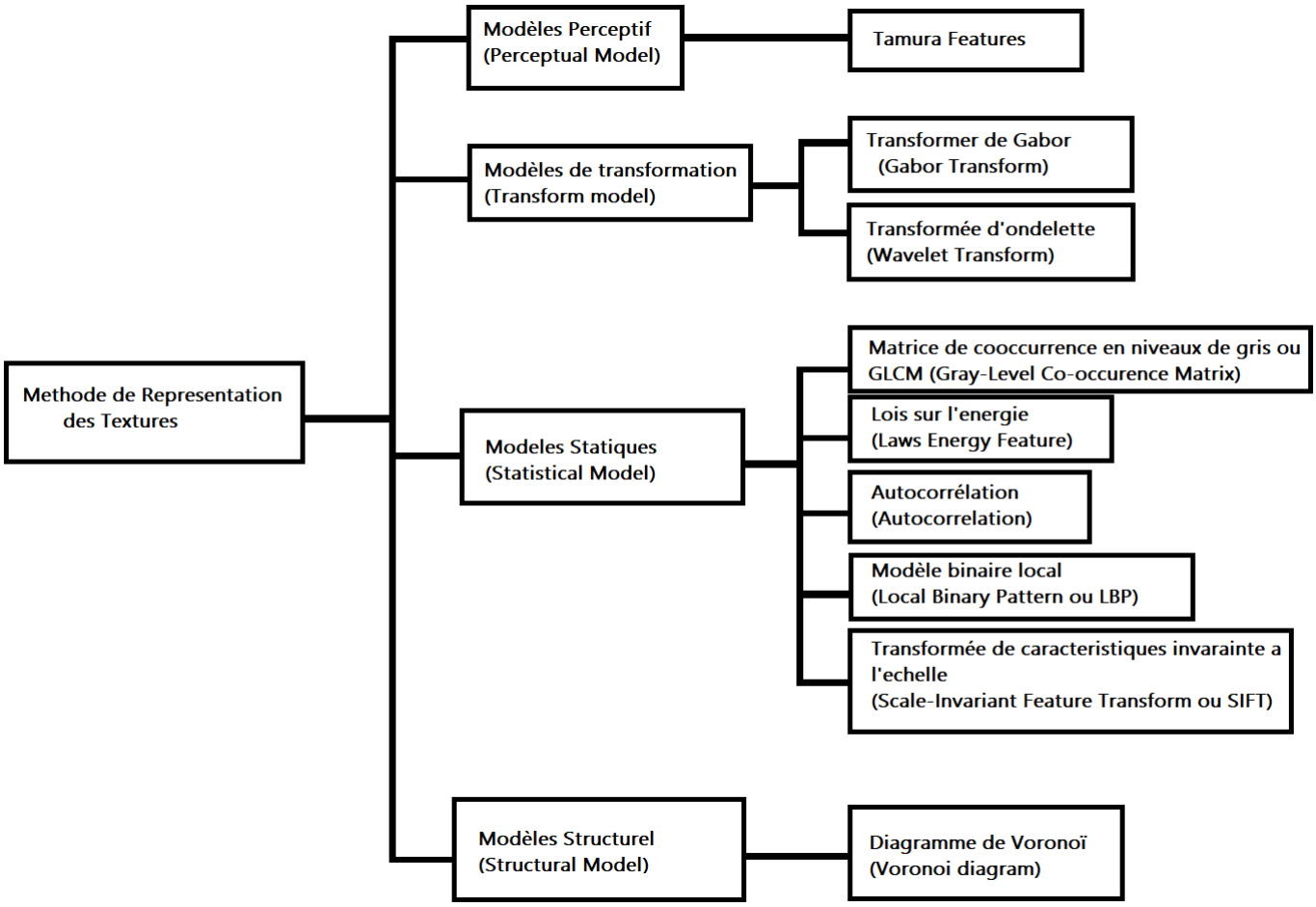


Figure 19: Les modèles de caractérisation des Textures

L'approche statique basée sur les filtres est les approches les plus populaires. Certaines approches proposent des algorithmes robustes qui sont des combinaisons de caractéristiques structurelles et statistiques.

3.1 Méthode basée sur l'approche statique

Une texture peut être décrite statistiquement, en rendant compte de la manière dont les niveaux de gris de l'image s'organisent les uns par rapport aux autres. Cette approche traite la texture comme un processus stochastique déterministe. Les notions qualitatives visuelles tel que la granularité, le contraste, homogénéité, la répétitivité, la fragmentation ainsi que l'orientation doivent être modélisé dans ce model. La granularité est un motif dominant dans une texture et est parfois considéré comme la texture elle-même. Les grains sont constitués de pixels voisins qui ont un niveau de gris similaire. Ainsi la taille et le niveau de gris de la texture détermine les niveaux de finesse (macro ou micro) de la texture. Le contraste quant à lui, est basé sur le nombre de niveaux de gris et leur taux de variation. Un

changement de contraste entraine donc une modification dans la qualité de l'image mais pas dans sa structure. Enfin, l'orientation est, pour chaque région, la propriété globale qui traduit la direction générale prise par les motifs ou grain d'une texture. On peut citer la matrice de cooccurrence comme descripteur de texture basé sur l'approche statique :

- **Matrice de cooccurrence de niveaux de gris :** est une méthode qui représente la référence en analyse statistique de texture d'une image. Les matrices de cooccurrence sont très riches en information de texture et servent souvent de méthode comparative pour les nouvelles approches. Les matrices de cooccurrence font l'objet de plusieurs recherches. Une technique simple à mettre en œuvre et offrant de bonnes performances, est de mesurer la probabilité d'apparition des paires de valeurs de pixels situés à une certaine distance. Elle est basée sur le calcul de la probabilité $P_{i,j}(d, \vartheta)$ qui représente le nombre de fois où un pixel de niveau de gris i apparait à une distance d relative d'un pixel de niveau de gris j suivant une orientation ϑ *donnée.* Les matrices de cooccurrence sont exploitable par extraction des attributs numériques calculés appelés paramètres de texture. Généralement, les quatre attributs utilisés dans la recherche d'image par le contenu sont : l'énergie, le contraste, l'entropie et le moment inverse de différence, obtenus après normalisation des matrices $P_{i,j}(d, \vartheta)$ par $N \times M$ [31].

3.2 Méthode basée sur l'approche structurale

La texture est considérée comme un phénomène linéaire et organisé avec des techniques structurales. Les techniques d'analyse de texture structurelle définissent la texture comme la structure de composants de texture bien définis tels que des lignes parallèles régulièrement espacées.

Les modèles structurels purs de texture sont basés sur l'idée que les textures sont constituées de primitives qui apparaissent dans des arrangements spatiaux répétitifs quasi réguliers. Pour décrire la texture, il faut décrire les primitives et les règles de placement. Une primitive est un ensemble connexe de cellules de résolution caractérisées par une liste d'attributs. La primitive la plus simple est le pixel avec son attribut de ton gris. Un exemple d'une tel primitive est un ensemble de pixels connectés au maximum et ayant même ton de gris ou ayant la même direction de bord [35].

3.3 Méthode basée sur l'approche spectrale

L'expression des périodicités et autres régularités dans une image ou dans un signal se fait naturellement dans le cadre de l'analyse spectrale. On peut donc fabriquer comme descripteur, un transformé de fourrier discrète. Des méthodes plus populaires consiste à utiliser un ensemble de filtres de Gabor disposé à plusieurs échelles et orientations. Cette dernière méthode à l'avantage de permettre l'identification des « traits perceptif majeurs ». Certain travaux ont montré qu'un descripteur basé sur une telle description spectrale pouvais rendre compte de la structure spatiale dominante d'une scène naturelle.

Une autre alternative à la méthode basée sur la transformée de fourrier est de calculer la transformée en cosinus discrète (DCT). En pratique, les descripteurs basés sur les coefficients

DCT ont permis de discriminer des images d'intérieur et d'extérieur, des images de paysages urbains contre des paysages naturels et, combinés à d'autres descripteurs, plusieurs catégories de scènes naturelles simultanément.
Intéressons-nous à quelque méthode de caractérisation de textures :

- **Les filtres de Gabor :** sont très utilisés en indexation, pour la description de la texture. Ils sont notamment utilisés par la norme MPEG-7. Ces filtres sont généralement exploités dans l'espace de Fourier dans le but de caractériser des textures locales. L'utilisation des filtres de Gabor consiste à analyser indépendamment différentes parties de l'espace de Fourier, à l'aide de plusieurs filtres [13]. Les principales étapes d'extraction des caractéristiques des textures de l'image en utilisant le filtre de Gabor sont :

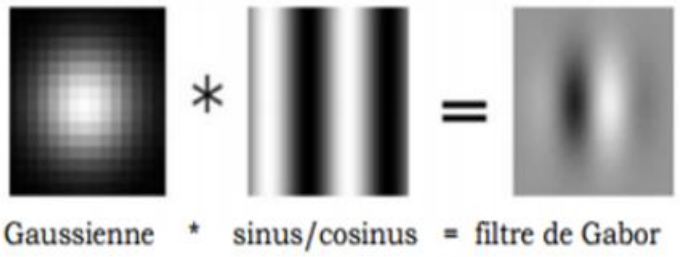


Figure 20: Filtre gabor [36]

L'expression du filtre de Gabor en deux dimensions est donnée par la formule suivante:

$$f_{Gb} = g(x', y')\exp\left(2\pi j[\mu_0(x - x_0)^2 + v_0(y - y_0)^2]\right)$$

$$g(x', y') = \exp\frac{x'^2}{2\sigma x^2} - \frac{y'^2}{2\sigma y^2}$$

x' et y' sont des constantes d'espace de l'enveloppe gaussienne qui déterminent l'étendue de l'onde suivant les axes x et y respectivement.(x_o,y_o) : représente le point d'origine où s'applique la fonction fGb (fGb est maximale en ce point) [36]. Un filtre de Gabor est un filtre de convolution obtenu en appliquant cette fonction à un masque de convolution.

- **La transformée en ondelettes :** est à la base de nombreuses analyses de texture, telles que les filtres de Haar. La caractérisation des texture avec la transformée d'ondelette est utilisé principalement dans le but de recherche d'images.

  L'approche continue (signal continue) des ondelettes pour un signal est trop complexe pour être applicable rapidement sur une image. Une approche par décomposition du signal en signal discret permet de résoudre ce problème. Ainsi la méthode DWT (Discret Wavelet Transform) ou Transformée d'ondelette discret permis d'obtenir une transformée bien plus rapide.
  L'on ne choisit alors plus une ondelette mère mais des filtres discrétisations. Pour calculer un transformé d'ondelette, on n'a alors besoin que de 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. Les filtres de Haar sont fréquemment employés en apprentissage pour obtenir la description d'un objet (comme un visage ou une personne).

4 Descripteur de Forme

L'information de forme est complémentaire à celle de la couleur au même titre que l'information de texture. La forme peut être un descripteur très riche en information pour un objet et donc une image. De nombreuses solutions ont été proposées pour représenter la forme des objets sur une image. Parmi eux on distingue : les descripteurs basés sur les régions et les descripteurs Basés sur les frontières.

Les moments invariants (qui font référence au descripteur basé sur les régions) sont utilisés pour caractériser l'intégralité de la forme d'une région. Leurs attributions sont peu variant aux transformations géométriques comme la translation, la rotation et le changement d'échelles. Les descripteurs basé sur les frontières sont classiquement les descripteurs de Fourier et permettent de retrouver la forme des objets sur une image.

Ci-dessous, un graphe contenant la liste non- exhaustive des méthodes et descripteur permettant la caractérisation de la forme :

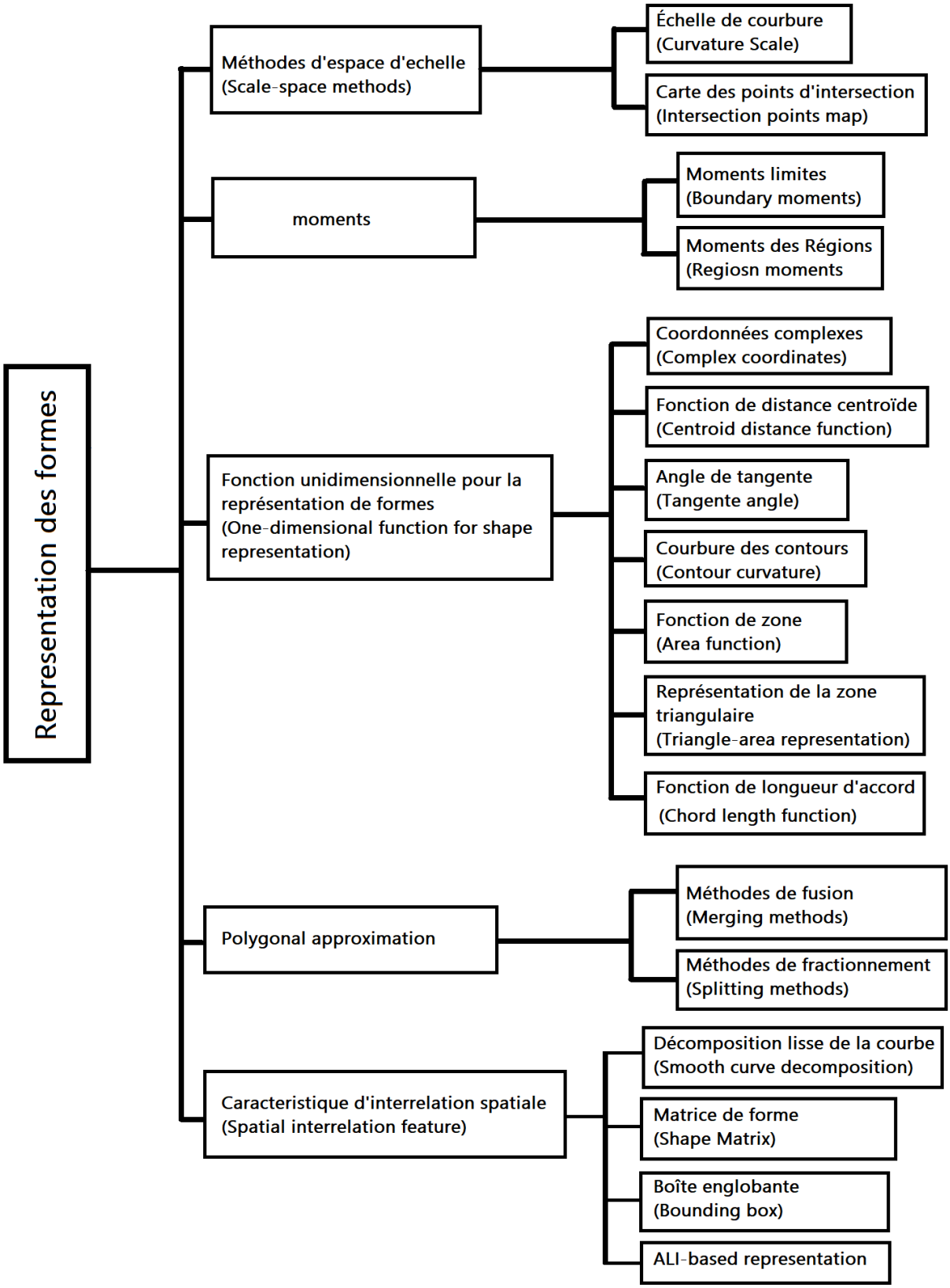


Figure 21: Les modèles de caractérisation des Formes

4.1 Transformée de Fourier

Les descripteurs de Fourier DFs font partie des descripteurs les plus populaires pour les applications de reconnaissance de formes et de recherche d'images. Ils ont souvent été utilisés par leur simplicité et leurs bonnes performances en termes de reconnaissance [29]. De plus, ils permettent de décrire la forme de l'objet à différents niveaux de détails et ils sont peu sensibles au bruit. Les descripteurs de Fourier sont calculés à partir du contour des objets. Leur principe est de représenter le contour de l'objet par un signal 1D, puis de le décomposer en séries de Fourier [28].

4.2 Transformée de Hough

La transformée de Hough est une technique d'extraction de caractéristiques utilisé dans l'analyse d'images, la vision par ordinateur et l'image numérique en traitement. Cela se fait à l'aide d'une procédure de vote qui est effectuée dans un espace paramétrique. Dans sa version classique, le transformée de Hough permet surtout l'indentification des lignes dans une image. L'idée clé de la transformée de Hough est de savoir qu'une ligne peut être déterminée de manière unique par le paramètre de pente m et le paramètre d'interception b. Partant de ce constat, une ligne droite = m x + b dans l'espace image peut être représentée par un point (b, m) dans l'espace des paramètres. Cependant, nous pouvons voir des valeurs illimitées des paramètres m et b si une droite est perpendiculaire à l'axe x. Éviter cela, on peut utiliser un couple de paramètres différent, noté r et α, pour les lignes de la transformée de Hough. Le paramètre r représente la distance entre la ligne et l'origine, tandis que α est l'angle du vecteur de l'origine au point le plus proche sur la ligne. L'équation de la droite peut donc s'écrire suite : r = x cos α + y sin α.

4.3 Moments géométriques

Les moments géométriques [37] permettent de décrire une forme à l'aide de propriétés statistiques. Ils représentent les propriétés spatiales de la distribution des pixels dans l'image. Ils sont facilement calculés et implémentés. Par contre, cette approche est très sensible au bruit et aux déformations et le temps de calcul de ces moments est très long.
La formule générale des moments géométriques est donnée par la relation suivante:

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

p + q est l'ordre du moment. Le moment d'ordre 0 $m_{0,0}$ représente l'aire de la forme de l'objet [28].
Il est possible de calculer à partir de ces moments l'ellipse équivalente à l'objet. Afin de calculer les axes de l'ellipse, il faut ramener les moments d'ordre 2 au centre de gravité :

$$m^g_{2,0} = m_{2,0} - m_{0,0} x_c^2$$
$$m^g_{1,1} = m_{1,1} - m_{0,0} x_c y_c$$
$$m^g_{0,2} = m_{0,2} - m_{0,0} y_c^2$$

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

$$\alpha = \frac{1}{2}\arctan\frac{2m^g_{1,1}}{m^g_{2,0} - m^g_{0,2}}$$

L'un des moments géométriques populaires, sont les moments invariants de Hu :

- **Les moments invariants de Hu :** Les moments centralisés non orthogonaux sont invariants en translation et peuvent être normalisés par rapport aux changements d'échelle. Cependant, pour permettre l'invariance à la rotation, ils nécessitent une reformulation. Hu ont décrit deux méthodes différentes pour produire des moments invariants en rotation. Le premier utilisait une méthode appelée axes principaux, mais il a été noté que cette méthode peut échouer lorsque les images n'ont pas d'axes principaux uniques. De telles images sont décrites comme étant symétriques en rotation. La deuxième méthode Hu décrite est la méthode des invariants de moment absolus et est discutée ici. Hu a dérivé ces expressions d'invariants algébriques appliqués à la fonction génératrice de moment sous une transformation de rotation. Ils consistent en des groupes d'expressions de moment centralisées non linéaires. Le résultat est un ensemble d'invariants de moments orthogonaux absolus (c'est-à-dire de rotation), qui peuvent être utilisés pour l'identification de modèles invariants d'échelle, de position et de rotation. Ceux-ci ont été utilisés dans une simple expérience de reconnaissance de formes pour identifier avec succès divers caractères typés. h_n et n_{th} Hu invariant moment :

  $$\boldsymbol{h_0} = \eta_{20} + \eta_{02}$$
  $$\boldsymbol{h_1} = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$$
  $$\boldsymbol{h_2} = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2$$
  $$\boldsymbol{h_3} = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$$
  $$\boldsymbol{h_4} = (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (3\eta_{21} - \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$$
  $$\boldsymbol{h_5} = (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2 + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})]$$

  Enfin, un invariant d'inclinaison, pour aider à distinguer les images miroir, est :
  $$\boldsymbol{h_6} = (3\eta_{21} - \eta_{03})(\eta_{30} - \eta_{12})[(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] + (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03})[3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]$$

  Ces moments sont d'ordre fini, donc, contrairement aux moments centralisés, ils ne comprennent pas un ensemble complet de descripteurs d'images. Cependant, des invariants d'ordre supérieur peuvent être dérivés. A noter que cette méthode tombe également en panne, comme avec la méthode basée sur l'axe principal pour des images à symétrie de révolution car les sept moments invariants seront nuls.

4.4 Moments orthogonaux

Les moments cartésiens sont formés à l'aide d'un ensemble de base de monôme $X^p Y^q$. Ces moments sont basique et non orthogonal. Les monômes augmente rapidement a mesure que l'ordre de puissance augmente, produisant ainsi des description hautement corrélées. Cela peut donc entrainer les présences d'informations descriptives importantes dans de petites différences entre les moments. Ce qui nécessite une grande précision de calcul.
Cependant, les moments orthogonaux ont l'avantage de nécessiter une précision moindre pour représenter les différences avec la même précision que les monômes. La condition

d'orthogonalité simplifie la reconstruction de la fonction d'origine à partir des moments générés. Orthogonalité signifie mutuellement perpendiculaires, exprimées mathématiquement - deux fonctions y_m et y_n sont orthogonales sur un intervalle $a \leq x \leq b$ si et seulement si :

$$\int_a^b ym\,(x)yn(x)dx = 0\,; \quad m \neq n$$

On note qu'une suite de polynômes qui sont orthogonaux par rapport à l'intégration, sont aussi orthogonaux par rapport à la sommation. Les moments de zernike sont les des moments orthogonaux populaires qui a prouvé son efficacité :

- **Les moments de Zernike :** Ce type des moments a été initialement introduit par Teague [38] et qui sont construits à partir de polynômes complexes et forment un ensemble orthogonal complet définie sur le disque unité. Ils sont invariants par rotation et changements d'échelles et présentent des propriétés intéressantes en termes de résistance aux bruits, efficacité informative et possibilité de reconstruction. Les moments orthogonaux de Zernike d'ordre *p* sont définis de la manière suivante :

$$A_{m,n} = \frac{m+1}{\pi} \iint I(x,y)[V_{m,n}(x,y)]dxdy$$

Où *m* et *n* définissent l'ordre du moment et *I*(*x, y*) le niveau de gris d'un pixel de l'image *I* sur laquelle on calcule le moment. Les polynômes de Zernike $V_{m,n}(x,y)$ sont exprimés en coordonnées polaires :

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

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

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

avec n = 0,1,2 … ∞ ; 0 ≤ |m| ≤ n et n - |m| un entier pair.

Les polynômes de Zernike sont orthogonaux, et donc les moments correspondants le sont également. Cette propriété d'orthogonalité annule l'effet de redondance de l'information portée par chaque moment [28].

5 Mesure de similarité

La mesure de la similarité est une partie importante dans la recherche d'une image. Il s'agit là des méthodes par laquelle on définit 'si deux image sont similaire ou non'. Une fois la formule ou méthode établie, elle sera utilisée par le système pour calculer la similarité (ressemblance) entre une image requête et les images se trouvant dans notre base de donnée d'image. Cette formule permet de donner un score à chaque image selon 'à quel point ses images sont similaire'. Un score minimal signifie que deux images sont très similaires. Un score maximal correspond à une distance maximale entre deux images.

La mesures de la similarité prend souvent le plus de temps du point de vu utilisateurs. Car il s'agit de comparer toute les images de la base de donnée à l'image requête fournis. Il est

important de trouver un bon compromis entre le temps de calcule des distances de similarités d'un système CBIR et la précision des calculs. L'efficacité d'un système de similarité se mesure avec 4 critères principaux :

- **La perception** : Une distance minimale dans l'espace caractéristique indique deux images semblables. Une distance maximale indique que les images sont non semblables. Cet espace de valeur est-il assez large ?
- **Le calcul** : La mesure de distance se calcule rapidement pour une faible latence. Le calcul à effectuer est-il simplifié et rapide ?
- **La scalabilité** : Les valeurs attribué lors du calcul des distance doivent être invariant au changement de base de donné. Les valeurs des distances attribué reste elle cohérente en cas de changement de base de données ?
- **La robustesse** : la mesure devra être robuste aux changements des conditions d'acquisition d'image.

Le choix de la mesure de similarité la plus appropriée dépend du niveau d'abstraction de la représentation de l'image. 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. La comparaison au niveau des pixels est très spécifique et, par conséquent, n'est utilisée que lorsque des appariements relativement précis sont nécessaires [17].

5.1 Distance de Minkowski

La méthode la plus simple pour mesurer la similarité entre deux images correspond aux distances de Minkowski. Cette distance est une famille de distances vectorielles. Soit I_1 et I_2 deux vecteurs de caractéristiques, la distance L_r est définie par :

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

Où $r \geq 1$ est le facteur de Minkowski, n est la dimension de vecteur. Les métriques de Minkowski représentent un bon compromis entre efficacité et performance. Pour cette famille de distances, plus le paramètre r augmente, plus la distance Lr aura tendance à favoriser les grandes différences entre coordonnées. Ces distances sont rapides à calculer et simples à implémenter, par contre leur calcul est réalisé en considérant que chaque composante du vecteur apporte la même contribution à la distance [28].

5.2 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. 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 [12]. Sa formule générale est donnée par :

$$\boldsymbol{D_Q = \sqrt{(f_1 - f_2)^T A (f_1 - f_2)}}$$

Où f1 et f2 sont deux vecteur dont on veut mesurer la distance
A est la matrice de similarité avec A = [a_{ij}]
a_{ij} est la distance entre deux éléments des vecteurs f_1 et f_2 ; sa formule est :

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

5.3 Distance Chi carré

La distance du chi carré est l'une des mesures de distance qui peut être utilisée comme mesure de la dissemblance entre deux histogrammes et a été largement utilisée dans diverses applications telles que la récupération d'images, la classification des textures et des objets et la classification des formes.
Le calcul de la distance du chi carré est une méthode statistique qui mesure généralement la similarité entre 2 matrices de caractéristiques. Une telle distance est généralement utilisée dans de nombreuses applications telles que la récupération d'images similaires, la texture d'image, les extractions de caractéristiques, etc.
La distance Chi-carré de 2 tableaux 'x' et 'y' avec la dimension 'n' est calculée mathématiquement en utilisant la formule ci-dessous :

$$X^2 = \frac{1}{2}\sum_{i=1}^{n}\frac{(xi - yi)^2}{(xi + y1)}$$

6 Conclusion

Afin de mesurer la similarité entre les images, on doit extraire des caractéristiques représentatives de cette image. Ces caractéristiques sont appelés descripteurs d'image et sont divisés en plusieurs catégories (couleur, texture, forme) chacune pouvant avoir plusieurs approche ou modèle de caractérisation (Couleur: moment de couleur, histogramme de couleur; Texture: Transformer d'ondelette, filtre de Gabor; Forme: Moment de Zernike, moment de hue...).
Le choix des descripteurs d'image est capital pour un système CBIR. Bien que la couleur, la texture, et la forme soit tous des descripteurs puissant. Le choix des méthodes d'extraction ou modèle de caractérisation a un rôle déterminant dans l'expression du potentiel discriminatoire ou non des descripteurs.

Chapitre IV : Méthodes de Classification

1. Introduction

Les descripteurs combinés aux méthodes de mesure de similarité entre vecteur, permettent de faire un système CBIR de pauvre performance. La nécessité de méthodes de classification d'image par rapport aux caractéristiques extraite des descripteurs peut donc se voir. Les méthodes de classifications d'image fait référence à un processus de vision par ordinateur qui permet de classer une image en fonction de son contenue visuel. Ces méthodes peuvent grandement contribuer à améliorer l'efficacité (précision et vitesse) d'un système CBIR. Dans ce chapitre, nous parleront des différents méthodes de classification d'image. Nous aborderont des méthodes un peu anciennes, mais efficace et populaire tel que le SVM. Mais également des méthodes récentes de classification basé sur l'intelligence artificiel comme les CNN (Convolutional neurone network ou Réseaux de neurone convolutif).

Avec les avancé technologique, les méthodes d'apprentissage en profondeur et les réseaux de neurones ont pris un élan immense. Les formations pour un apprentissage en profondeur d'un logiciels ou outils tels que des classificateurs, qui alimentent une énorme quantité de données, les analysent et en extraient des fonctionnalités utiles. L'intention des processus de classification est de catégoriser tous les pixels d'une image numérique dans l'une classe parmi plusieurs afin de facilité le processus de recherche. En général, ce sont les données multi spectrales qui sont utilisées pour effectuer la classification. Car le modèle spectral présente dans les données pour chaque pixel est utilisé comme base numérique pour la catégorisation. L'objectif de la classification d'images est d'identifier et de représenter, en tant que niveau de gris (ou couleur) unique, les caractéristiques apparaissant dans une image en termes d'objet que ces caractéristiques représente réellement sur le terrain. Dans l'analyse d'image numérique, la classification des images est peut-être la partie la plus importante. La classification est une tâche complexe, par conséquent elle peut être parfois difficile à réaliser. La classification des images fait référence à l'étiquetage des images dans l'une des nombreuses classes prédéfinies. Sur de grande base de données, la classification de chaque image de manière manuelle pouvant demander un effort considérable, la classification automatique grâce à la vision par ordinateur se montre fort utile. Dans notre implémentation nous utiliseront donc un CNN comme méthode de classification automatique.

2. Structure pour effectuer la classification des images

Pour effectuer une classification d'image, on passera par les mêmes étapes, quel que soit la méthode utilisé :

- **Prétraitement d'image** : le but de ce processus est d'améliorer les données d'image (caractéristique) en supprimant les distorsions indésirables et en améliorant certaines caractéristiques importantes de l'image afin que les modèles de vision par ordinateur puissent bénéficier de ces données amélioré pour travailler. Les étapes de prétraitement de l'image comprennent la lecture de l'image, le redimensionnement de l'image et l'augmentation des données (mise à l'échelle des gris de l'image, réflexion, flou gaussien, histogramme, égalisation, rotation et traduction).

- **Détection d'objet :** On parle ici de localiser les objets présents sur l'image, c'est-à-dire la segmentation de l'image et l'identification de la position des objets d'intérêt.

- **Extraction et formation de caractéristiques :** c'est l'étape la plus importante dans laquelle des méthodes statique ou d'apprentissage en profondeurs doivent être utilisé pour identifier les modèles les plus intéressants de l'image et en tiré des vecteurs ou matrice de valeur représentative. Il est important de rechercher les caractéristiques qui pourraient être uniques à une classe particulière et qui, plus tard aideront le modèle à différencier les classes. Ce processus est appeler apprentissage du modèles car c'est ici que le modèle apprend les caractéristiques du jeu de données (base de donnée).

- **Classification de l'objet :** Dans cette étape, on catégorise les objets détectés dans des classes prédéfinies en utilisant une technique de classification appropriée qui compare les motifs d'image avec les motifs cibles

3. Type de classification

Dans la classification, plusieurs type et processus d'apprentissage sont proposé chacune adapté à des situations et moyen différente.

3.1 Classification avec apprentissage supervisé

La classification supervisée est basée sur l'idée qu'un utilisateur peut sélectionner des échantillons de pixels dans une images sont représentatifs de classe spécifiques, puis demander au logiciel de traitement d'image d'utiliser ces sites de formation (base d'image de test) comme références pour la classification de tous les autres pixels de l'image. On doit sélectionner les ensembles de donnée (dans notre cas images) de test en fonction nos connaissances. L'on doit définir également les limites (ou seuil) de la similarité des autres pixels pour les regrouper. En fonction des caractéristiques spectrales de la zone d'entrainement, le seuil peut être variable. Il faut ensuite désigner le nombre de classes dans lesquels les images sont classées. Une fois qu'une caractérisation statistique a été réalisée pour chaque classe d'information, l'image est ensuite classée en examinant la réflectance de chaque pixel et en prenant une décision sur la signature à laquelle elle ressemble le plus. La classification supervisée utilise des algorithmes de classification et des techniques de régression pour développer des modèles prédictifs.

La performance de la classification dépend notamment de l'efficacité de la description d7i et de la fiabilité du système d'apprentissage pour classer efficacement tout nouvel exemple (pouvoir prédictif) [40].

Les algorithmes de classification superviser comprennent : la régressions linéaire, la régressions logistique, les réseaux de neurones (NN : neurone network), l'arbre de décision, la machine a vecteur de support (SVM : support vector machine), la forêt aléatoire, les k plus proche voisin (k-NN : k nearest neighbor)

3.2 Classification avec apprentissage semi supervisé

Ce méthode est semi-supervisé dans le sens où il apprend à la fois des données étiquetées et non étiquetées. Fondamentalement, la méthode combine des données étiquetées disponibles le long des itérations avec des informations contextuelles fournies par le grand nombre de données non étiquetées disponibles. L'utilisation de ces données non étiquetées permet de minimiser les efforts de l'utilisateur, car potentiellement moins d'étiquettes doivent être attribuées aux images au fil des itérations.

3.3 Classification avec apprentissage non supervisé

La classification non supervisée est celle ou les résultats (groupement de pixels ayant des caractéristiques communes) sont basés sur l'analyse logicielle d'une image sans que l'utilisateur ne fournisse d'échantillons. La détermination des pixels lié et es regroupement en classes sont fait par l'ordinateur. L'utilisateur peut juste spécifier le nombre de classe de sortie, ainsi que l'algorithme a utilisé par le logiciel. Cependant l'utilisateur peut parfois avoir besoin de connaissance pour les zones à classer (classe importante).

Après avoir choisi les paramètres sur lesquels va porter la classification, de nombreuses techniques peuvent alors être envisagées et on distingue deux catégories de classificateurs : hiérarchiques et non-hiérarchiques [40].

Les algorithmes les plus couramment utilisée dans l'apprentissage non supervisé comprennent l'analyse de clusters, la détection d'anomalies et les réseaux de neurone. Cette dernière méthode est d'ailleurs celle que nous avons utilisée comme méthode de classification.

4. Méthodes de classification d'image et détection d'objet

Dans le domaine de la vision par ordinateur, l'un des doutes les plus courants que la plupart d'entre nous ont est de savoir quelle est la différence entre la classification d'images, la détection d'objets

La classification d'image nous aide à classer ce qui est contenu dans une image. La localisation d'image spécifiera l'emplacement d'un seul objet dans une image tandis que la détection d'objet spécifie l'emplacement de plusieurs objets dans l'image. Enfin, la segmentation d'image créera un masque pixel par pixel de chaque objet dans les images.

4.1 Algorithme K plus proche voisin ou K-NN (K-nearest Neighbor)

L'algorithme K-Nearest Neighbors est une méthode non paramétrique utilise pour la classification et la régression. Dans les deux cas, l'entré est constituée des k exemples de formation les plus proches dans l'espaces des fonctionnalités. Cet algorithme est très simple par rapport au suivante car son apprentissage est non paramétrique et donc paresseux. Cet qui signifie que les fonctions ne sont rapprochée que localement et les calculs sont différés jusqu'à l'évaluation de la fonction. K-NN repose simplement sur la distance entre les vecteurs de caractéristiques et classe les points de données inconnus en trouvant la classe la plus courante parmi les k-exemples les plus proches. Afin de trouver les k plus proche voisin, nous devons définir, comme pour toutes méthodes de classification, une métrique de

distance ou fonction de similarité. Les choix les plus communs pour l'algorithme de K-NN est la distance euclidienne et la distance Manhattan. La sortie est une appartenance à une classe. Un objet est classé par vote de pluralité de ses voisins, l'objet étant affecté à la classe la plus commune parmi ses k plus proches voisins (k est un entier positif, typiquement petit). Par exemple si k=1, alors la classe de l'objet est simplement assigné à la classe de ce seul voisin le plus proche. Le voisin le plus proche condensé (CNN, l'algorithme Hart) est un algorithme conçu pour réduire l'ensemble de données pour la classification K-Nearest Neighbor.

4.2 Algorithme K-moyens (ou K- means)

Le k-means clustering est une méthode de quantification vectorielle, issue du traitement du signal, qui vise à partitionner n observations en k clusters dans lesquels chaque observation appartient au cluster de moyenne la plus proche (centres de cluster ou centroïde de cluster), servant de prototype de la grappe. Il en résulte un partitionnement de l'espace de données en cellules de Voronoi. Le clustering k-means minimise les variances intra-cluster (distances euclidiennes au carré), mais pas les distances euclidiennes régulières, ce qui serait le problème de Weber le plus difficile : la moyenne optimise les erreurs au carré, alors que seule la médiane géométrique minimise les distances euclidiennes. Par exemple, de meilleures solutions euclidiennes peuvent être trouvées en utilisant les k-médianes et les k-médoïdes.

La classification est utilisée pour rassembler les zones extraites de l'image de document en K groupements, fonction d'un critère de "ressemblance". Parmi les algorithmes de regroupement, la méthode des k-moyennes est la plus utilisée. n s'agit d'une approche facile à implémenter et qui consiste à répartir les objets des images en k groupes autour de k centres appelés noyaux ou centroïde : un objet image est dans un groupe (ou cluster) s'il est plus proche, en fonction d'une distance choisie, du centroïde de ce groupe que de n'importe quel autre centroïde des autres groupes. Le centroïde de chaque groupe, recalculé à la fin de l'exécution de la méthode, correspond au barycentre de chaque groupe. L'algorithme k-moyennes nécessite de choisir au départ un nombre k de points dans l'espace vectoriel des objets comme représentants des k classes à trouver. Généralement, les k images sont choisis au hasard [40].

4.3 Algorithme Support Vecteur Machine ou SVM (Support Vector Machine)

Les machines à vecteurs de support (SVM) sont des algorithmes d'apprentissage automatique supervisés puissants mais flexibles qui sont utilisés à la fois pour la classification et la régression. Les machines à vecteurs de support ont leur mode de mise en œuvre unique par rapport aux autres algorithmes d'apprentissage automatique. Ils sont extrêmement populaires en raison de leur capacité à gérer plusieurs variables continues et catégorielles. Le modèle Support Vecteur Machine est essentiellement une représentation de différentes classes dans un hyperplan dans un espace multidimensionnel. L'hyperplan sera généré de manière itérative par la machine à vecteurs support afin que l'erreur puisse être minimisée. L'objectif est de diviser les jeux de données en classes pour trouver un hyperplan marginal maximal. Il construit un hyper-plan ou un ensemble d'hyper-plans dans un espace de grande dimension et une bonne séparation entre les deux classes est obtenue

par l'hyperplan qui a la plus grande distance au point de données d'apprentissage le plus proche de n'importe quelle classe. La puissance réelle de cet algorithme dépend de la fonction noyau utilisée. Les noyaux les plus couramment utilisés sont le noyau linéaire, le noyau gaussien et le noyau polynomial.

4.4 Réseau de neurone convolutif ou CNN (Convolutional neuron network)

Les tendances en matière de récupération d'image se concentrent sur les réseaux de neurones convolutif capable de générer de meilleurs résultat à un cout de calcule élevé. Les réseaux de neurone convolutif (CNN, ou ConvNet) sont un type particulier de réseau de neurones multicouches, conçus pour reconnaître des modèles visuel directement à partir d'image de pixels avec un prétraitement minimal. Les réseaux de neurone convolutif sont en fait une architecture spéciale des réseaux de neurone artificiels. La particularité du réseau de neurone convolutif est d'utiliser certain caractéristiques retrouver dans le cortex visuel humain. Grace à cela, elle a pu obtenir des résultats de pointe dans les taches de vision par ordinateur. Les CNN (réseau de neurone convolutif) sont composés de deux éléments, la couche convolutif et les couche de regroupement. Bien que ces élément soit simple, il existe des combinaisons infinie pour l'organisation de ses couches pour un problème de vision par ordinateur donné. La partie la plus difficile dans l'utilisation des CNN est de savoir comment combiner ses couches dans la pratique afin de concevoir une architecture modèle qui utilisent au mieux ces couches. La raison pour laquelle les réseaux de neurone convolutif sont si populaires est que leur architecture n'ont pas besoin d'extraction de caractéristiques. Le système convolutif apprend à faire l'extraction de caractéristique et le concept de base est qu'il utilise la convolution de l'image et des filtres pour générer des caractéristiques invariantes qui sont transmises à la couche suivante. Les caractéristiques de la couche suivante sont alambiquées avec différents filtres pour générer des caractéristiques plus invariante et abstraite et le processus se poursuit jusqu'à ce qu'il obtienne une caractéristique/sortie finale qui est invariante à l'occlusion. Parmi les architectures de CNN les plus utilisé, on trouve LeNet, AlexNet, ZFNet, GoogLeNet, VGGNet, ResNet et Plus récemment MobileNet.

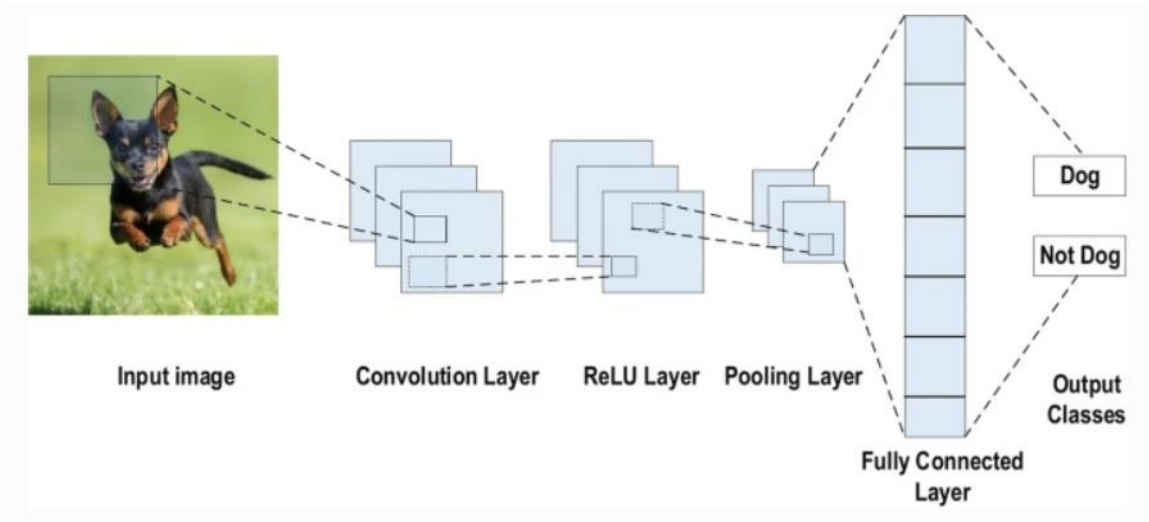


Figure 22: Fonctionnement général des CNN [Site 3]

4.4.1 Architecture

Tout CNN, se compose de couche d'entré, de couche cache (Couche convolutif, de regroupement etc…) et d'une couche de sortie. Les couches intermédiaires sont dites cachées car leur entrée et sortie peuvent être cachées par la fonction d'activation (parfois considéré comme une couche d'activation) et la convolution finale. Dans le CNN, les couches cachées comprennent des couches qui effectuent des convolutions. Cela n'inclut généralement une couche qui effectue un produit scalaire du noyau de convolution avec la matrice d'entrée de la couche. La fonction d'activation du produit obtenue a une fonction d'activation généralement ReLu (Unité Linéaire Rectifiée). Pendant que le noyau de la convolution transverse le long de la matrice d'entré (ici une matrice d'image l'image) génère une carte de caractéristiques. A son tours, cette carte caractéristique contribue en tant qu'entré pour la couche suivante. Viennent ensuite d'autres couches telles que les couches de regroupement, les couches entièrement connectées et les couches de normalisation.

4.4.2 Couches d'un CNN

Les architectures des CNN sont formée d'empilement de couches, ayant chacun une fonction définie. Ces couches distinctes permettent de transformer le volume d'entré en un volume de sorite via une fonction différentiable. Ci-dessous quelque type de couche utilisé dans les CNN :

4.4.3.1 Couche de convolution (Conv)

Les couches convolutif (Conv) sont comme on peut s'en rendre compte au nom « Réseau de neurone Convolutif » la pierre angulaire des CNN. Les calculs les plus lourds sont effectués dans cette couche. On trouve dans cette couche des filtres (ou noyaux) qui sont convolutés vers la matrice d'entrée.

Exemple : Si l'on a comme entré des données de taille [32 x 32 x 3] et que la taille du filtre est de 5x5, alors chaque neurone de la couche Conv aura des poids pour une régions [5x5x3] dans le volume d'entrée, pour un totale de 5x5x3 = 75 poids.

Après chaque couche CONV dans un CNN, nous appliquons une fonction d'activation non linéaire, telle que ReLU, ELU ou l'une des autres variantes. Les couches d'activation ne sont pas techniquement des "couches" car aucun paramètre/poids ne sera apprise à l'intérieur.

4.4.3.2 Couche de regroupement (Pool)

Placé entre deux couches convolutif, la couche de regroupement permet de réduire la taille des images tout en préservant leurs caractéristiques importantes par des opérations de mutualisation.

Pour effectuer une opération de mutualisation, il faut d'abord découper l'image en cellules régulières, ensuite on garde la valeur maximale à l'intérieur de chaque cellule. Pour éviter de perdre trop d'information, l'on utilise souvent des petites cellules carrées. Les choix les plus courants sont le 2x2 cellules adjacentes qui ne se chevauchent pas ou les 3x3 cellules séparées les uns des autres par un pas de 2 pixels.

4.4.3.3 Couche entièrement connectée (fully connected: FC)

Les neurones des couches FC sont entièrement connectés à toutes les activations de la couche précédente, comme c'est le cas pour les réseaux de neurones à anticipation. C'est-à-

dire que les neurones dans une couches entièrement connectée ont des connexions vers toutes les sorties de couche précédente. Les couches FC sont toujours placées à la fin du réseau (c'est-à-dire que nous n'appliquons pas une couche CONV, puis une couche FC, suivie d'une autre couche CONV).

4.4.3.4 Couche de correction (ReLu)

La couche de correction ou d'activation est l'application d'une fonction non-linéaire aux cartes de caractéristiques en sortie de la couche de convolution. En rendant les données non-linéaires, elle facilite l'extraction des caractéristiques complexes qui ne peuvent pas être modélisées par une combinaison linéaire d'un algorithme de régression.

4.4.3.5 Couche de perte (Loss)

La couche de perte est la dernière couche du réseau. Elle calcule l'erreur entre la prévision du réseau et la valeur réelle.

4.4.3 Hyper paramètre

Les hyper-paramètres sont les paramètres des modèles deep learning, donc des réseaux de neurones. Ils sont de deux types :

- **Les hyper-paramètres du modèle :** Ces hyper-paramètres sont principalement lié à la taille du réseau de neurones, souvent prédéfinis par l'architecture de du CNN choisi. Ces paramètres sont rarement ou peu changé, car l'architecture est souvent utilisée telle qu'elle est proposée par son auteur.
- **Les hyper-paramètres d'algorithme** : Ces hyper-paramètres sont ceux qui sont le plus souvent modifier celons les besoins. Elles permettent de contrôler la vitesse d'apprentissage du modèle. Elles sont : le chargement des données en entrée du modèle (la taille du lot de données, la méthode de chargement des données), le déroulement de la phase d'entraînement (le nombre d'itérations, la fonction de perte, le taux d'apprentissage), le déroulement de la phase de validation

4.4.4 Architecture de détection d'objet

Les Architecture de détections d'objets sont des Architecture généralement combiner avec l'architecture des CNN dans le but spécifique de la détection d'objet sur une image. La détection d'objet est similaire a la classification « multi-étiquette » d'images. C'est-à-dire qu'une image peut correspondre à plusieurs classes. La détection d'objets contrairement à la classification d'image, renvoie également la position des objets détectés sur l'image. Nous utiliserons un algorithme de détection d'objet dans notre implémentation en raison de la présence de multiple objets sur une majorité des images de pascal VOC 2012.Quelque description d'algorithme de détection d'objets ci-dessous :

4.4.4.1 Architecture YOLOR

YOLO a été proposé par Joseph Redmond et al. en 2015. Il a été proposé de faire face aux

problèmes rencontrés par les modèles de reconnaissance d'objets à cette époque, Fast R-CNN est l'un des modèles à la pointe de la technologie à cette époque mais il a ses propres défis tels que ce réseau ne peut pas être utilisé en temps réel, car il faut 2-3 secondes pour prédire une image et ne peut donc pas être utilisé en temps réel. YOLOR est aussi une méthode basé sur l'approche neuronale.

4.4.4.2 Architecture SSD

SSD (Single Shoot Multi Box Detector) est un algorithme de détection d'objets basé sur le deep learning. En tant que l'un des algorithmes de détection les plus courants, il peut grandement améliorer la vitesse de détection et assurer la précision de la détection. SSD utilise VGG16 pour extraire les cartes de fonctionnalités. Ensuite, il détecte les objets à l'aide de la couche Conv4_3. L'on utilisera l'algorithme SSD en raison de sa rapidité surprenante avec un CNN tel que Mobilenet. L'algorithme SSD combiné à Mobilenet crée un modèle rapide et adaptable à tout ordinateur ou Smartphone.

4.4.4.2 Exemple de CNN populaire

Avec l'avancer des nouvelles technologies on peut s'attendre à voir le nombre et la puissance des CNN croitre. Voici quelque exemple de CNN actuellement disponible :

4.4.4.2.1 GoogLeNet

GoogLeNet est également connue sous le nom de GoogLeNet version 1. Le réseau GoogLeNet a été une étape importante dans le développement des classificateurs CNN. Avant sa création, les CNN les plus populaires empilaient simplement des couches de convolution de plus en plus profondes, dans l'espoir d'obtenir de meilleures performances. A mesure des versions l'architecture de GoogLeNet évoluera grandement. L'architecture GoogLeNet est :

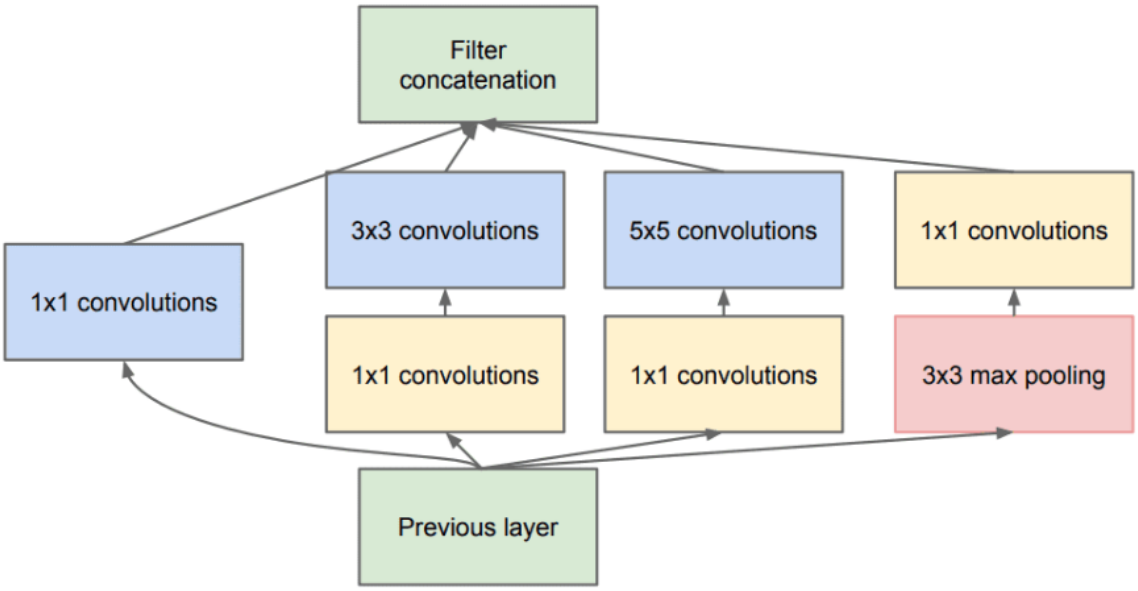


Figure 23: Architecture général de GoogLeNet [Site 4]

4.4.4.2.2 ResNet

Ceci est construit sur le concept de "connexions sautées" et utilise beaucoup de normalisation par lots pour lui permettre de former des centaines de couches avec succès

sans sacrifier la vitesse au fil du temps. Ces connexions de saut sont également connues sous le nom d'unités fermées ou d'unités récurrentes fermées et présentent une forte similitude avec les éléments réussis récents appliqués dans les RNN. Les ResNet sont constitués de ce qu'on appelle un bloc résiduel.

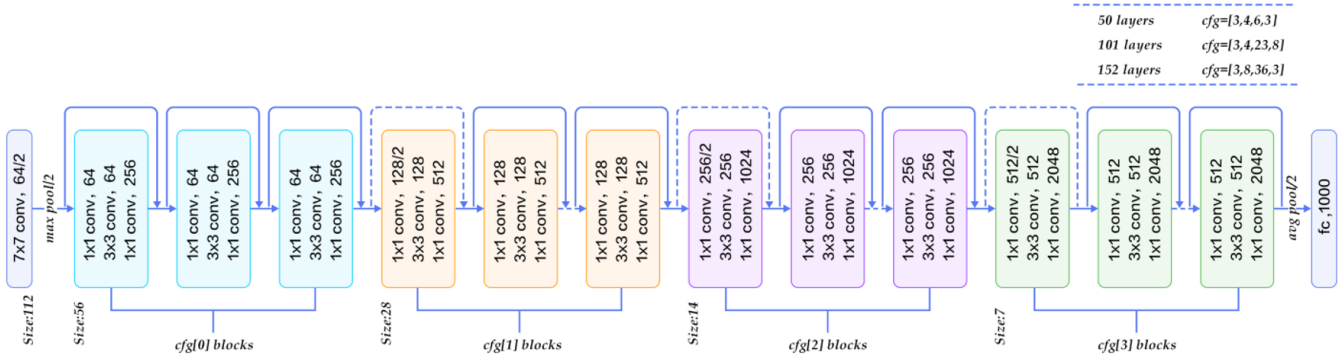


Figure 24: Architecture général de ResNet [Site 5]

4.4.4.2.3 MobileNet

Le modèle MobileNet est construit sur des convolutions séparables en profondeur, qui sont un type de convolution factorisée qui divise une convolution régulière en une convolution en profondeur et une convolution ponctuelle. La convolution en profondeur utilisée par MobileNet applique un seul filtre à chaque canal d'entrée. Les sorties de la convolution en profondeur sont ensuite combinées à l'aide d'une convolution 11 par la convolution ponctuelle. En une étape, une convolution conventionnelle filtre et mélange les entrées pour créer un nouvel ensemble de sorties. La convolution séparable en profondeur divise cela en deux couches : une pour le filtrage et l'autre pour la combinaison. L'architecture MobileNet est calculée en fonction de la profondeur circonvolutions séparables (DS). Le concept de décomposition de convolution appelée factorisation est considérée comme factorisant une convolution standard en une convolution en profondeur.

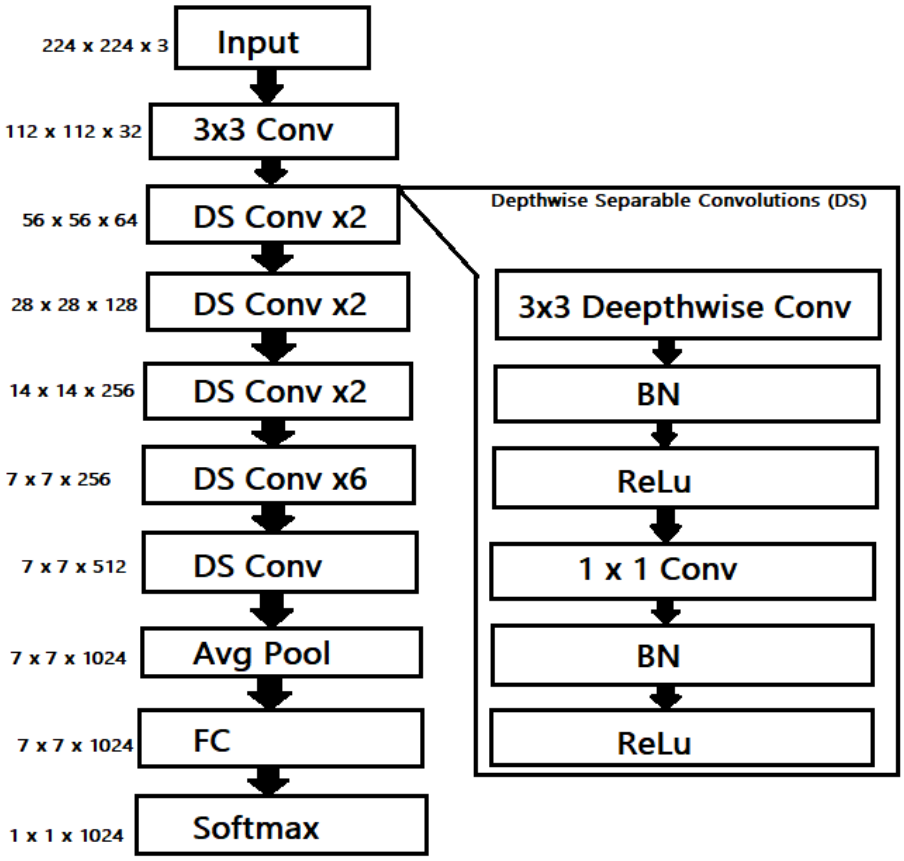


Figure 25: Architecture de MobileNet

5 Conclusion

Dans ce chapitre nous avons présenté des méthodes de classifications. Quel que soit les descripteurs d'image choisie, les méthodes de classification sont de puissant outil pour améliorer de manière significative les performances d'un système CBIR.

Nous avons présenté des méthodes d'apprentissage supervisé, semi-supervisé, et non superviser tel que: Support Vecteur Machine ou SVM, K-plus proche voisin ou K-NN, K-Moyens.

Nous avons également présenté des réseaux de neurones convolutifs, cité quelque exemple d'architecture et expliquer le fonctionnement des principale couche de ces réseaux. Avec l'évolution des technologies, les méthodes d'apprentissage basé sur les réseaux de neurone convoluté sont de plus en plus performantes et utilisé. Ces méthodes utilise des caractéristiques abstrait et ne nécessite pas toujours l'extraction de descripteur d'image précis. Le système convolutif apprend à faire l'extraction des caractéristiques qui sont transmises à la couche suivante.

Nous avons en fin présenté quelque algorithme de détection d'objets souvent utilisé en combinaison avec les réseaux de neurone convolutif.

Chapitre V : Implémentation et évaluation expérimental

1 Introduction

Dans le but de trouver des moyens de plus en plus rapide et fiable de faire de la recherche d'image par le contenue. Nous avons implémenté un logiciel permettant de testé les capacités d'un réseau de neurone convolutif. Dans cet études, nous utiliseront un model pré-entrainer de Mobilenet (permettant la détection d'objet), ainsi que les caractéristique de couleurs des images. Dans le but de mettre en évidence l'efficacité du logiciel sur des images contenant plusieurs objets, nous choisissons la base d'image VOC2012. Nous avons également mené des tests avec d'autre CNN comme ResNet et GoogLeNet avant de finaliser notre choix sur Mobilenet.

2 Outil d'implémentations

Pour ce logiciel, nous utiliseront le langage Python sur un IDE (environnement de développement intégré ou integrated development environment) célèbre appelé vscode.
Notre logiciel utilise plusieurs librairies pour son fonctionnement. Ses librairies sont faciles à télécharger à partir de commande sur cmd (l'invite de commande) ou power Shell utilisé hors connexion. Elles sont peu gourmandes en poids de donnée. Les Outils utilisées sont :

- **OpenCV2:** est une bibliothèque graphique libre, initialement développée par Intel, spécialisée dans le traitement d'images en temps réel. Nous utiliserons la librairie dans nos processus de traitement d'image tel que la modification de l'espace colorimétrique, le calcule d'histogramme de couleur, l'utilisation d'un modèle DNN ou Deep Neuronal Network (fonction qu'on utilisera ici pour : l'utilisation de model CNN) etc…
- **Tkinker:** est la bibliothèque graphique libre d'origine pour le langage Python, permettant la création d'interfaces graphiques. Nous utiliserons cette librairie pour la création d'interface via sa fonction grid.
- **NumPy:** est une bibliothèque pour le langage de programmation Python, ajoutant la prise en charge de grands tableaux et matrices multidimensionnelles, ainsi qu'une grande collection de fonctions mathématiques de haut niveau pour opérer sur ces tableaux. Nous utiliserons Numpy dans nos créations et calcule sur les vecteurs ou matrice.
- **glob:** est utilisé pour renvoyer tous les chemins de fichiers qui correspondent à un modèle spécifique. Nous pouvons utiliser glob pour lire l'ensemble des fichiers se trouvant dans un dossier. Pour cet implémentations nous utiliseront glob pour indexer la base de donné en renvoyant les images une par une à l'extracteur de caractéristique.
- **Imutils :** Une série de fonctions pratiques pour faciliter les fonctions de traitement d'image de base telles que la traduction, la rotation, le redimensionnement, la squelettisation et l'affichage des images Matplotlib avec OpenCV et Python 2.7 et

Python 3. Nous utilisons Imutils pour calculer l'histogramme de couleur et pour d'autre tâche différente.

Utiliser lors des tests de fonctionnalité (ResNet, HOG, Zernike) :

- **Scipy:** est une bibliothèque visant à unifier et fédérer un ensemble de bibliothèques Python à usage scientifique. Scipy utilise les tableaux et matrices du module NumPy. Nous utiliserons cette librairie dans plusieurs tests, bien que la version finale du logiciel ne l'utilise plus
- **Tensorflow :** est un outil open source d'apprentissage automatique développé par Google. Souvent utiliser pour l'intelligence artificielle, la vision d'ordinateur (classification, segmentation d'image et vidéo…). Nous utiliserons cet outils, plus précisément la fonction readNetFromTensorflow lors des tests sur d'autre CNN.

Nous utiliserons également un modèle pré-entrainé de Mobilenet Version 3 créé le 14/01/2022. Ce Modèle a été entrainé avec un algorithme de détection d'objet sur la base de données « large coco (Microsoft Common Objects in Context) dataset» créé par Microsoft et contenant 328.000 images.
Ce modèle ainsi que les fichiers de configuration de Mobilenet doivent être téléchargé puis utiliser. A deux ces fichiers pèsent 13Mo.
On doit également télécharger un fichier label permettant de convertir les ID d'objet en nom compréhensible (Exemple : id= 1 correspond au label « personne »)

3 Etape d'implémentation

Pour utiliser notre logiciel ou créer un logiciel similaire. Il faut installer les outils cité plus haut.

3.1 Télécharger les outils requit a l'utilisation de l'implémentation

Voyons voir étape par étape comment installer chacun de ces outils afin d'être à mesure d'utiliser ou créer notre application :

- **Vscode** : l'IDE (Environnement de développements intégré) dans lequel on écrira notre code peut être télécharger officiellement sur le site :
  https://code.visualstudio.com/download
- **Python** : le langage de programmation utilisé peut être télécharger depuis le site officiel de python : https://www.python.org/downloads/
  L'installation de python permettra également 'l'installation de « pip » qui nous permettra de facilité l'installation de librairie sur python. Il est possible en cas d'installation raté que « pip » ne soit pas installer. Dans ce cas il faut installer « pip » grâce au terminal intégré de vscode, on utilise la commande : python get-pip.py
  Pour Ouvrir le terminal intégré, il suffit de faire clic droit sur un fichier python puis « Ouvrir vers le terminal intégré » ou « Open in integrated terminal ». Le terminal s'ouvrira en bas de l'interface de vscode nous permettant de lancer des commande. Ecrire puis lancer la commande de téléchargement de python

- **OpenCV** : OpenCV 2 qui nous permettra grandement nos outil de traitement d'image peut être téléchargé directement sur le terminal intégré de VSCODE avec la

commande : pip install opencv-python

- **Tkinker** : Cette librairie nous facilitera la création d'interface. Le processus d'installation est le même que celui de opencv2, sur le terminal integré de vscode, on utilise la commande : pip install tk
- **Scipy** : Cette libraire permet des traitement sur des vecteurs, la commande utiliser pour l'installer est : pip install scipy
- **NumPy** : Cette librairie permet des traitements sur des vecteurs, la commande utilise pour l'installer est : pip install numpy
- **Glob** : Pour installer cette librairie on utilise la commande : pip install glob2
  Ou la commande : pip3 install glob2
  Celons la version de python l'un ou l'autre peut fonctionner
- **Pillowpy** : Cette libraire permet des traitement sur des images, la commande utilisé pour l'installer est : pip install Pillowpy
- **Imutils** : Cette librairie peut être utilisé grace a la commande : pip install imutils

Une fois tous ces outils installer, notre programme pourra être lancé.
Le modèles pré-entrainer de Mobilenet utiliser est ouvert au public et peut être télécharger à partir du lien suivant : https://gist.github.com/dkurt/54a8e8b51beb3bd3f770b79e56927bd7

3.2 Télécharger le programme

Afin de rendre notre programme ouvert au plus grand notre, nous l'avons uploader en ligne. Le système implémenter peut donc être télécharger avec le lien suivant : https://github.com/Trast00/Systeme-CBIR-Master

3.3 Utiliser le programme

Une fois télécharger, il suffit de décompressé le fichier avec WinRar ou WinZip (disponible gratuitement sur internet).
Pour tester le programme il faut lui fournir une base d'image et l'indexer(remarque notre base d'image était trop lourd pour être mise en ligne). Il suffit d'indexer une base d'image grâce au fichier indexer.py (disponible dans le programme). Cet fichier peut être utiliser en lui fournissant un dossier d'image à indexer. Si cet dossier a pour chemin d'accès « JPEGImages » car se trouvant dans le programme ; on utilise la commande « python indexer.py --dataset JPEGImages --index indexs/index.csv ».
Si ce dossier a pour chemin d'accès « C:\Memoire\pythonTest\program\OtherJPEG» ; on peut l'indexer par « python indexer.py --dataset C:/Memoire/pythonTest/program/OtherJPEG --index indexs/index.csv »
Des instructions supplémentaire (méthode d'utilisation, fonctionnement, utilité) sont parfois disponible à l'intérieur des fichiers python.
En fin, pour lancer le programme, il suffit de lancer la commande « python interface.py » dans le terminal intégré de vscode.

4 Justification de choix

Lors de la création de notre application, nous avons dû faire plusieurs choix. Nous avons fixé plusieurs paramètre de fonctionnement de notre application. Parmi ces choix, nous avons :

4.1 Choix des CNN et Mobilenet

Les réseaux de neurone profond sont de plus en plus utilisés avec l'amélioration des systèmes d'apprentissage automatique. Cette méthode a prouvé ces performance dans des études récente et peut rester néanmoins à faible cout de calcule et faible empreinte mémoire. C'est pourquoi nous avons choisi de les utiliser dans notre étude.

L'un des point les plus négative de l'utilisation des CNN est le coût de calcule élever et le empreinte mémoire important. Nous remédions à cet problème en choisissant d'utiliser Mobilenet qui un système à très faible cout de calcule et très faible empreinte mémoire tout en conservant une précision correcte. Mobilenet est un système embarqué donc étant optimiser pour les systèmes Mobile. C'est pourquoi notre logiciel résultant, est aussi optimiser pour les systèmes Mobile car sacrifiant légèrement la précision pour la vitesse et la légèreté. Mobilenet est donc un choix judicieux pour les systèmes de faible performance.

En effet, nous avons eu l'occasion de tester un modèle pré-entrainer ResNet lors de nos recherches : ResNet semble avoir une précision bien plus grande que Mobilenet, mais traite une image en 60-90 seconde alors que Mobilenet traite une image en moins de une seconde.

Avec un calcul grossier, nous pouvons en déduire que indexer une base d'image tel que VOC2012 avec 17 250 images aurait pris :

Avec ResNet : (17 250 image)/60 = 287.5 heurs = 11.97 jours

Avec Mobilenet, il nous faut environ 20-40 minutes pour indexer les 17 250 images. Cela justifie également notre choix. Le système utiliser a aussi l'avantage d'être utilisable sur des serveurs de site web et application Mobile.

4.2 Choix des histogrammes de couleur

La couleur est l'un des caractéristiques les plus importantes pour la recherche de similarité entre deux images. Nous choisissons alors d'utiliser l'histogramme de couleur afin de conserver le descripteur simple mais suffisante. Nous calculons également plusieurs histogramme par régions pour chaque image afin d'avoir un descripteur de couleur qui tient compte non seulement des couleurs, mais également de leur répartition sur l'image.

5 Utilisation de Mobilenet et d'un Descripteur de couleur

Notre logiciel utilisera Mobilenet afin de détecter les objets présents sur une image ainsi que leur taille. Apres quelque donnée d'entré, Mobile renvoie une liste d'index correspondant aux types d'objets présentes sur l'image.

La détection de couleur se fera grâce à l'extraction de 5 histogrammes de couleur par images. Nous mesurons la similarité des histogrammes grâce à la formule du Chi Carré.

5.1 Utilisation de Mobilenet

Nous configurons Mobilenet a fin de redimensionner les images en 320x320 pixel comme exigé par le modèle pré-entrainer. Nous configurons également Mobilenet à fin d'avoir 255 valeurs pour chacune des 3 valeurs caractéristiques de la couleur. Nous configurons

également un changement automatique de l'espace colorimétrique.
En fin, nous choisissons un seuil de confiance lors de la détection d'objet. En dessous de cet seuil, l'objet est ignorer et au-dessus l'objet sera renvoyé pour indexation. Apres plusieurs test lors de la création du programme nous avons choisi un seuil de 0.6 ; car au-dessus certain objet évidant n'était pas détecter et au-dessus des objets inexistant était détecter.
Envoyé une image aux modèle pré-entrainer de Mobilenet après configuration et ajout de label nous donnes comme résultat la listes des ID d'objets contenue dans l'image ainsi que les coordonnée permettant de mettre l'objet en boite. Ces données une fois utilisé avec de code simple d'affiche donne des résultats comme ci-dessous:

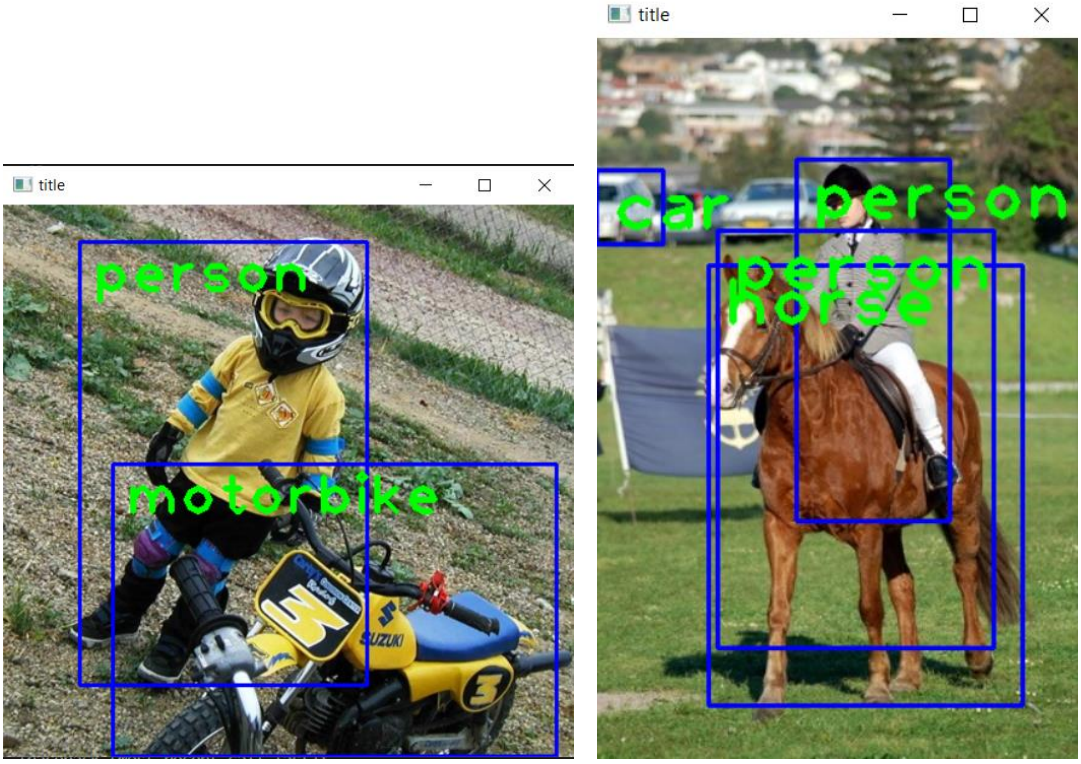


Figure 26: Utilisation de MobileNet

Par division de (taille d'objet)/ (taille d'image). Nous pouvons calculer à partir des coordonnées de boite le contenu de chaque image en pourcentages d'objet. Exemple :
2007_000733.jpg: person= 0.40587484035759896, motorbike = 0.41302681992337165
Nous sauvegardons dans liste des données (ID d'objet contenue dans l'image et coordonnée de boite) ainsi que le nom de l'image dans notre base de d'indexes sous format .csv.
L'architecture de cette version 3 de Mobile est ci-dessous

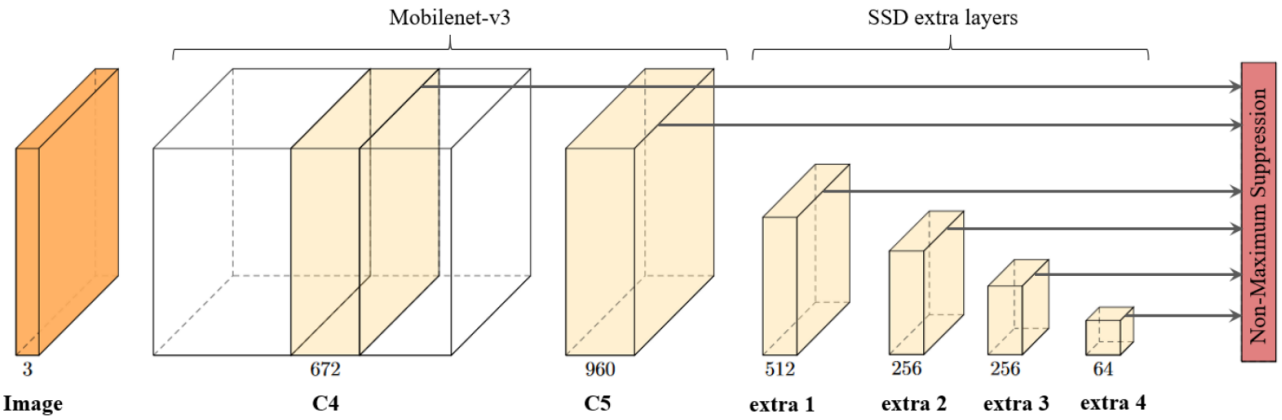


Figure 27: Architecture du détecteur d'objets : SSD avec extracteur de caractéristiques MobileNet-V3[41].

MobileNetV3 est la troisième version de l'architecture alimentant les capacités d'analyse d'images de nombreuses applications mobiles populaires. Il a été développé en supprimant des couches complexes et en utilisant la fonction H-SWISH au lieu de ReLU standard pour augmenter encore l'efficacité et la précision du réseau. Les MobileNet sont l'une des architectures les plus avancées en matière de vision par ordinateur mobile.

5.2 Extractions des histogrammes de couleurs

Afin de d'extraire les caractéristiques de couleur d'une image, plusieurs paramètres sont à prendre en compte.

Espace de couleur : Nous choisissons TSV (teinte saturation valeur) afin de garder le modèle relativement simple tout en étant meilleur que l'espace RVB car imitant mieux la perception humaine des couleurs.

Caractéristique d'histogramme : Il faut déterminer un nombre de d'intervalle de couleur (Bacs). Exemple de choix de bacs différent :

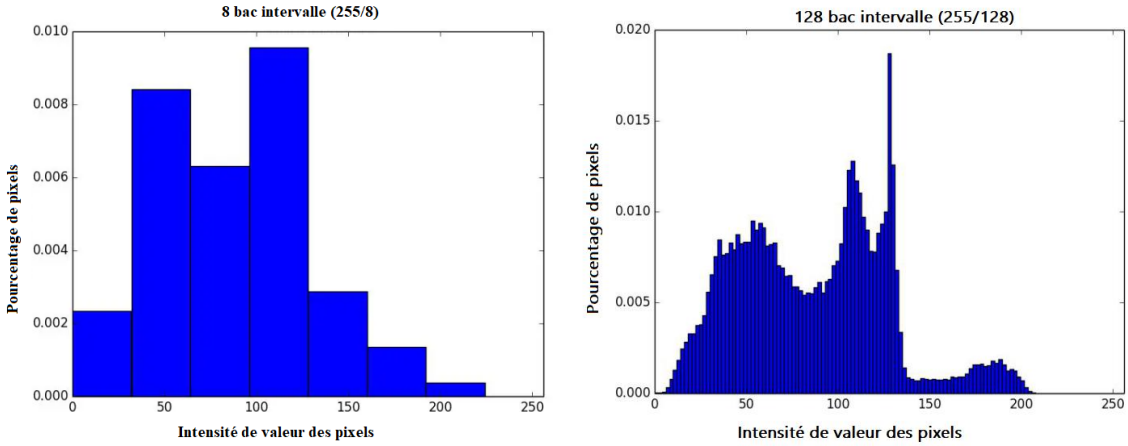


Figure 28: Illustration de différence de choix de bac d'histogrammes

Si l'on choisit un trop grand nombre d'intervalle d'intensité par histogramme, le système perd sa capacité à généraliser certaine couleurs. Choisir un trop peu d'intervalle d'intensité et le système ne distingue pas les nuance de couleur importante.

Un choix de 8 intervalle d'intensité pour la teinte, 12 intervalle d'intensité pour la saturation et 3 intervalle pour la valeur (pourcentage de blanc ajouté) semble donné des résultats que

correcte tout en limitant la dimension du vecteur caractéristique (8x12x3=288 valeurs).
Histogramme celons la localisation : On peut améliorer le vecteur caractéristique obtenue en le faisant tenir compte de la localisation de couleur. Ainsi le vecteur caractéristique final aura non seulement l'histogramme de couleur, mais aussi grossièrement les positions des couleurs sur l'image. Pour cela, on calcule 5 histogrammes de couleur sur 5 régions différente de l'image. Les régions des images sont sélectionnées comme montré ci-dessous :

Figure 29: Illustration de la méthode séparation régional de l'image pour les histogramme

Nous avons alors 5 histogramme de couleur qu'on sauvegardera dans la base de données en suivant l'ordre montré sur l'image afin de comparer par la suite chaque régions d'une image requête à chaque régions d'une autre image (exemple : régions haut-gauche avec régions haut-gauche, régions centre avec régions centre etc…).

5.3 Mesure de similarité des descripteurs

Lorsque les données sont indexer et sauvegarder dans la base de donnée. Il nous sera nécessaire de comparer les données sauvegardé lors d'une recherche. Les méthodes de comparaison utilisée sont simples mais efficace.

5.3.1 Mesure de la similarité celons les objets détecter

La base de données VOC2012 pouvant contenir plusieurs objets de type diffèrent par image. Il est difficile d'utiliser la classification d'image simple.

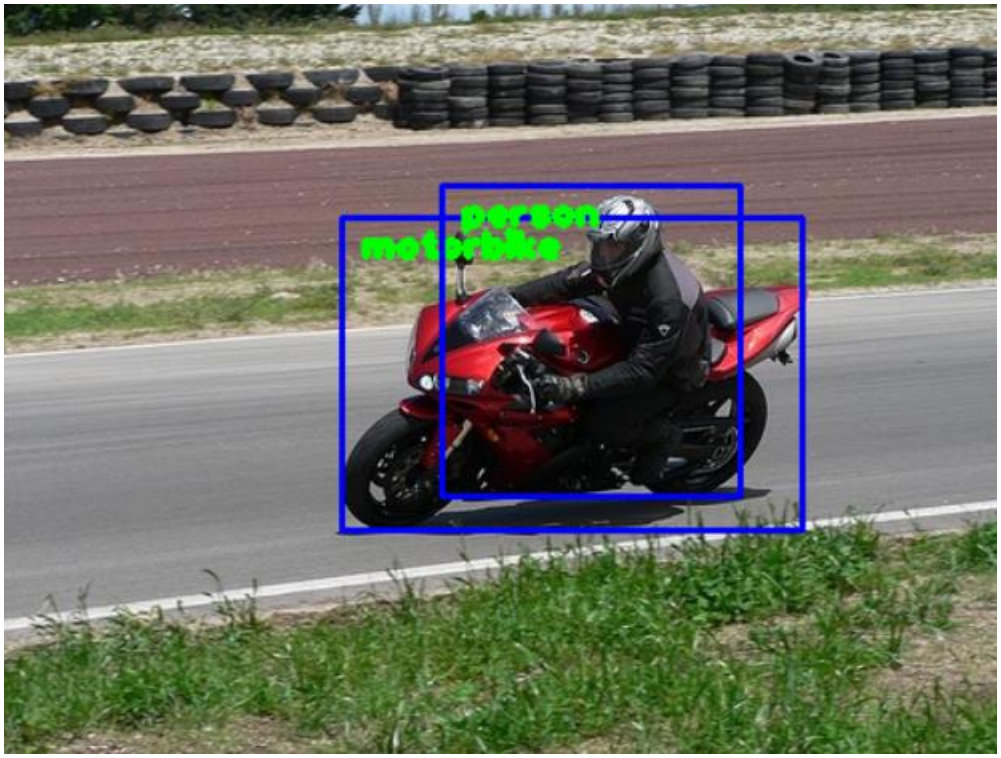


Figure 30: Exemple1 d'image contenant plusieurs objets

En effet, cet image ne peut ni vraiment être classifier comme une moto, ni comme personne. Elle appartient aux deux classes. D'où l'intérêt de la détection d'objet sur la classification d'image. On peut envisager de classifier l'image comme étant dans la classe moto et personne.

Mais il ne serait pas tout à fait correcte de classifier l'image ci-dessous comme appartenant à la classe personne :

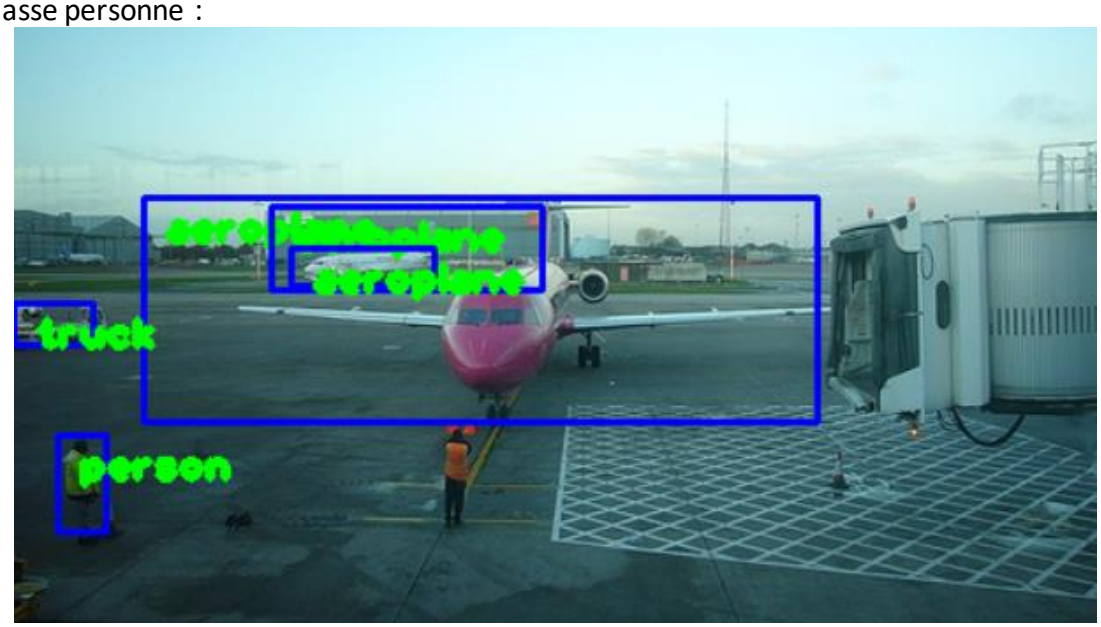


Figure 31: Exemple 1 d'image contenant plusieurs objets

Sur l'image ci-dessus, les objets 'personne' et 'camion' présente sur l'image ont été détecter, mais il ne semble pas être correcte de classifier image dans la classe 'personne' ou 'camion,

en raison de la taille des objets ‘personne’ et ‘camion’ par rapport à la taille des objets ‘avion’.
Pour comparer les images, nous allons donc utiliser non seulement la classe des objets, mais aussi leur taille afin d’avoir un calcule de score optimal pour la recherche.
Pour trouver des résultats similaires à une image requêtes donnée, nous rechercherons tous les images contenant au moins un objet de même classe et attribuons un score de différence selon la différence de taille entre les objets de la même classe.
Pour chaque objets présent sur l’image requête, si l’objets est présents sur l’image de la base, ajouté au score la différence de taille sinon, ajouter au score la taille de objets.
Cela permet de maximiser le score en cas d’absence de grand objet et de minimiser le score en cas d’absence de petit objet sur l’image de la base. Ainsi les images avec les scores les plus petites seront les images les plus proches en termes de taille d’objet. En fin, Nous trions les images selon la taille des objets les plus similaire jusqu’au moins taille d’objets les moins similaire.
Sans cette méthode de calcul :

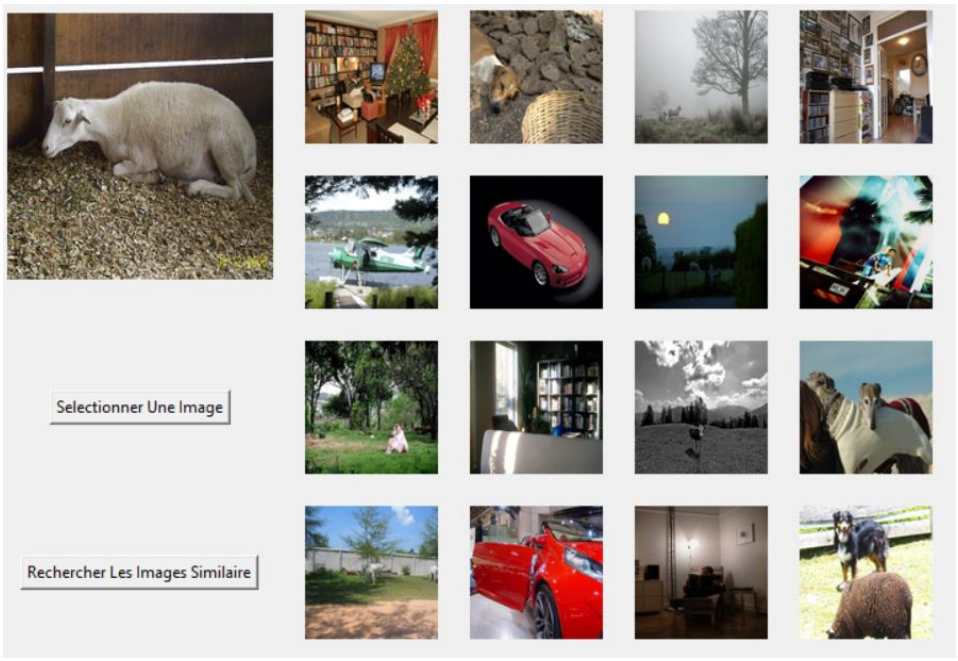


Figure 32: Exemple de résultat de recherche sans tenir compte de la taille

Une image requête comme celui en haut à gauche (Montons : ~50%) a autant de différence avec une image de voiture (Voiture : ~80%) une image de Mouton avec un chien (Mouton : ~30%, Chien : ~30%), car tous deux ont un objet de différence. En tenant compte de la taille des objets, la différence entre l’image requête est plus proche de l’image avec chien et mouton qu’avec l’image de voiture. « Ce n’est alors plus tel nombre objet de différence, mais tel pourcentage d’image de différence »

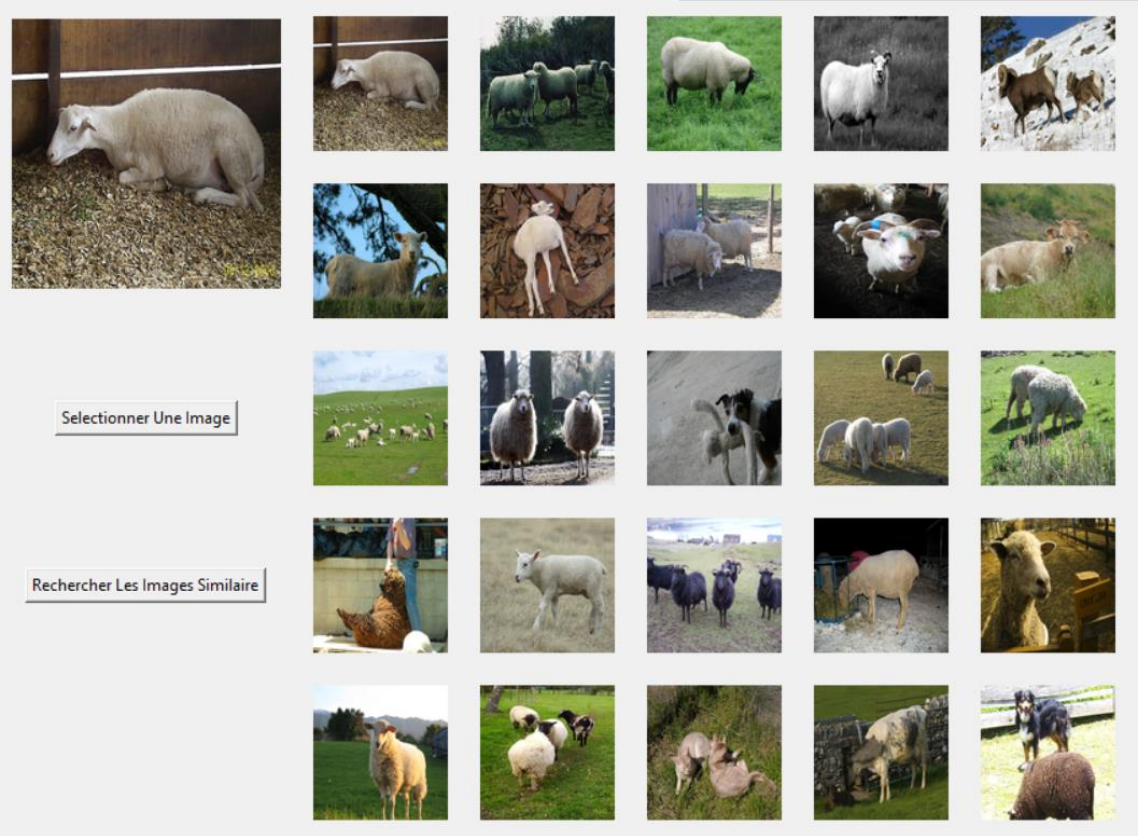


Figure 33: Exemple de résultat de recherche en tenant compte de la taille

5.3.2 Mesure de la similarité des histogrammes de couleurs.

Tout d'abord, pourquoi ajouter les caractéristiques de couleurs ? Si aucun objet n'est présent sur une image, la détection d'objets que mène Mobilenet est vaine. Lors de nos premier test, les descripteurs de couleur se sont avéré être un choix correcte pour trouver des images similaire a une image requête sans objets détecter par Mobilenet.

Pour mesurer la similarité les caractéristiques de couleur de deux images. On compare leur histogrammes régional (régions 1, 2, 3, 4, 5 entre eux) en utilisant la distance Chi carré de formule :

$$x^2 = \frac{1}{2}\sum_{i=1}^{n}\frac{(xi - yi)^2}{(xi + yi)}$$

Le calcul de la distance chi carré est une méthode statistique, mesure généralement la similitude entre 2 matrices d'entités. Une telle distance est généralement utilisée dans de nombreuses applications telles que l'extraction d'images similaires. La distance chi carré a montré de bonne performance dans la mesure de la similarité entre deux histogramme.

6 Résultat

Dans un système CBIR, la précision et le rappel sont les deux valeurs caractéristiques des performances du système. Nous allons donc essaies de déterminer la précision de notre

système avec ces deux valeur. Pertinente s
La précision se calcule :

$$\boldsymbol{Precision = \frac{nombre\ d\ image\ pertinente\ retrouvées}{nombre\ d'image\ retrouver}}$$

Le rappel se calcule :

$$\boldsymbol{Rappel = \frac{nombre\ d\ image\ pertinente\ retrouvees}{nombre\ \ d\ image\ pertinente\ totale}}$$

Nous utilisons la base de données VOC 2012, la mesure des performances de notre système a été mise en difficulté par différente facteur, notamment :

- Un trop grand nombre d'image pour déterminer les vraies négatives (donc calculer le rappel)
- Une grande partie des images appartienne a plusieurs classe (Donc pouvant être considérer pertinent comme résultat de recherche pour chacune de ses classe)
- Pour certain images, il est difficile de déterminer de manière purement objective si elles doivent être considéré comme correcte, ou non.

Pour remédier à ces problèmes, nous avons procédées comme suite :

- La base de donnée VOC 2012 contient des indexes sous format XML, dont nous pouvons estimer la fiabilité a proche de 100%. La base de donné contient également plusieurs autre type d'indexes (segmentation par classe, nombre, action etc…) qui permettre d'aider les utilisateurs dans la mesure des performances, ou pour un apprentissage supervisé, semi supervisé ou non supervisé. Nous avons considéré ses indexes comme étant correcte, et avons comparé les résultats fournie par notre systèmes au index présent dans la base de donné. Grace à cette comparaison nous avons pu relever les Vraies négatives (résultat correcte non afficher par notre système) avec une précision correcte. Cela nous a permis de faire une estimation du taux de rappel de notre système. A noter que cette estimation n'est pas forcement fiable, mais reste néanmoins la meilleure alternative pour estimer le rappel sur cette base de données
- En raison de la présence de multiple classe par image, si un résultat contient les objets recherchés le résultat est considérer comme suffisamment pertinente pour être afficher
- Pour déterminer les résultats incorrecte, si un résultat ne présente pas de manière évidente l'objet recherché, ce résultat est considéré comme incorrecte

selon ces critères d'évaluation, notre système à donner les résultats suivant :

Classes	**avions**	**bicyclette**	**oiseau**	**bateau**	**bouteille**	**bus**
Classes	**aeroplane**	**bicycle**	**bird**	**boat**	**bottle**	**bus**
Précision	97,45%	97,86	96,48%	96,62%	90,16%	99,28%
Rappel	90%	66,83%	71,85%	58,28%	28,48%	84,58%

Classes	**voiture**	**Chat**	**chaise**	**vache**	**table à**	**chien**

					manger	
Classes	**car**	**Cat**	**chair**	**cow**	**Dining table**	**dog**
Précision	94,40%	93,79%	98,35%	88,23%	97,26%	94,83%
Rappel	58,87%	83,33%	49,12%	73,23%	51,37%	72,85%

Classes	**cheval**	**Moto**	**personne**	**pot de plante**	**mouton**	**Canapé**
Classes	**horse**	**motorbike**	**person**	**potted plant**	**sheep**	**sofa**
Précision	96,30%	94,71%	98,40%	95,08%	94,39%	98,28%
Rappel	84,60%	80,52%	88,45%	40,52%	82,63%	51,07%

Classes	**train**	**télévision**
Classes	**train**	**tv monitor**
Précision	96,98%	97,51%
Rappel	86,58%	60,00%

Tableau 2: Liste des résultat de précision et rappel pour chaque classe.

La précision moyenne pour notre système, selon ces critères est de **95,81%**.

Le rappel moyenne estimer pour notre système, selon ces critères est de **68,15%**

On peut remarquer notre système possède une haute précision, bien que le taux de rappel estimer demeure faible. Remarque que l'on peut grandement jouer sur les valeurs de précisions et rappels en augmentant ou diminuant le seuil de confiance (nous avons choisi un seuil de 0,6 soit 60% comme préciser plus haut). Une augmentation du seuil de confiance, augmenterai en théorie la précisions et diminuerai le rappel. Une diminution du seuil de confiance, augmenterai en théorie le rappel et diminuerai la précision.

Le système semble engament confondre les motos et les bicyclettes, les petits chiens et les chats, les vaches et les moutons flou.

Nous pouvons également remarquer que le système a de grande difficulté à détecter (rappel faible) les petits objets tels que les bouteilles, les pots de plante et les chaises. En général, plus les objets sont gros, plus il est facile pour les systèmes de les détecter.

Rappelons également, que le système, grâce aux descripteurs de couleurs, est capable d'effectuer des recherches sur des images sans objets.

7 Conclusion

Dans ce chapitre, nous avons implémenté notre système de recherche d'image. Nous avons obtenue des précisions correctes et une vitesse de recherche acceptable. Les tests ont été effectués sur la base d'image pascale VOC 2012. Le système est suffisamment léger (13 Mo sans les indexes) pour être déployer sur Smartphone.

Conclusion générale

Les systèmes CBIR seront déterminants dans tout processus impliquant la vision par ordinateur. Ces systèmes peuvent être d'une importance exponentiel au fur et à mesure de l'évolution d'internet et de ses bases de medias (image et vidéo).

L'objectif de cet mémoire est de tenté la mise en place d'un système d'indexation et de recherche d'image par le contenue. Dans ce but, nous nous somme intéressé au traitement d'image et au fonctionnement des systèmes CBIR. Nous avons testé quelques descripteurs d'images et avons choisie d'utilisé les réseaux de neurones convolutif comme méthodes de classification d'images. Le domaine de la recherche d'image étant vaste nous avons dû faire des choix aux niveaux des descripteurs et des méthodes de classification.

Nous avons décidé d'utiliser la couleur comme descripteurs de bas niveau et de le combiner a une méthode de classification.

Nous avons utilisé une approche plus récente de la classification d'image avec les réseaux de neurone convolutif. Nous avons implémenté et testé un modèle pré-entrainé d'Inception version 2, un modèle pré-entrainé de ResNet et un modèle pré-entrainé de MobileNet.

En raison des faibles performances de nos machines nous avons eu du mal à utiliser les réseaux de neurone tel que ResNet qui était fort groumant en ressource, d'autant plus que la base de données VOC 2012 choisie contenait 17 250 images. Nous avons donc choisie Mobilenet afin de tenté de montrer comment une personne disposant de machine de faible performance pourrait implémenter des systèmes de recherche de haute précision et peu groumant en ressource.

Nous avons adapté le modèle de Mobilenet à la recherche d'image. Nous avons développé une interface afin de pouvoir présenté nos résultat de manière convenable.

Nous espérons avoir donné une vue générale sur l'utilisation des réseaux de neurone convolutif dans les systèmes CBIR et atteint les objectifs présentés par le sujet.

Perspectives :

Les approches permettant la mise en place des systèmes CBIR sont nombreuses. Néanmoins, l'approche via les réseaux de neurone convolutif offre déjà plusieurs perspectives d'évolution :

- Les réseaux de neurone sont très souvent utilisés avec leur architecture proposée par le concepteur, sans modification (comme nous l'avons fait). Néanmoins il est possible de modifier l'architecture des réseaux afin d'améliorer leur performance ou d'étudier efficacité des différente couches des réseaux.
- Nous avons utilisé un modèle pré-entrainé, il est néanmoins possible d'entrainer ses propres modèles, sur notre base de données spécifique. On peut s'attendre à une amélioration des résultats sur cette base de donné.
- Nous avons choisi Mobilenet afin de privilégier l'accessibilité de l'implémentation. Une étude des performances de ResNet, VGGnet etc… pourrait être effectué celons les performances des systèmes utilisé.
- Cette implémentation aurait pu être testée sur un appareil mobile, serveur web.

Bibliographies

[1]: G. K. Zipf, “Human Behavior and the Principale of Least- Effort”, Addison-Wesley, Cambridge, MA, 1949 avril 2013.
[2]: Henri Maitre, "Le traitement des images", La voisier, 2003.
[3]: A. Zerougui & N. Sari, "Traitement d’images monochromes", mémoire de master, université de Oum el bouaghi, 2017.
[4]: D. Boukhlouf, "Chapitre 03 : généralités sur traitement d’image", Université de Biskra, 2005.
[5]: D. Kaidi, "Classification non supervisée de pixels d’images couleur par analyse d’histogrammes tridimensionnels", mémoire de master, université de Tizi- Ouzou, 2017.
[6]: G. Bouthaina, "Représentation d'une image numérique". Polycopié de cours 1. Université Ferhat Abbas - Sétif 1. 2020
[7]: F. A. Hadjila & R. Bouabdallah, Reconnaissannce des visages en utilisant les réseaux de neurones. Mémoire d’ingénieur. Université de Tlemcen. 2003.
[8]: L. Jérôme Analyse multirésolution pour la recherche et l’indexation d’images par le contenu dans les bases de données images - Application à la base d’images paléontologique trans’tyfipal. Thèse de Doctorat. Université de bourgogne. 2005
[9]: F. Belgharbi & Latti F. Moteur de recherche d’image à base de contenu. Mémoire de master. Université de Tlemcen. 2016.
[10]: B. Alain "Indexation et recherche d’images par le contenu". Mémoire de master. Institut polytechnique de Hanoi. 2005.
[11]: A. Sameer & Al. "Columbia Object Image Library (COIL-100) ". Technical Report No. CUCS-006-96. Columbia University. February 1996.
[12]: B. Hichem "Une approche sémantique basée sur l’apprentissage pour la recherche d’image par contenu". Université de Monastir. 2009
[13]: H. Kamel "Recherche d’images par le contenu". Thèse de doctorat. Université de Constantine. 2010
[14]: Z. Houcemddin & al . "Développement d'un système de recherche d'image par le contenue". Memoir de master. Université abou belkaïda tlemcen 2018
[15] : L. Amira "Recherche d’images sémantique basée sur la sélection automatique des concepts" , Mémoire Master Académique, 2014.
[16] : A. Boucher, Recherche D'image Basee Sur Le Contenu Semantique, 2005
[17] M. Nabila & al. « Recherche d’image par le contenu ». Mémoire de master en informatique. Université de Tlemcen 2011
[18] H. Abed. Système d'indexation et de recherche d'image par le contenu. Université des Science de la technologie d'Oran. B.P 1505 EL M'NAOUR-ORAN (ALGERIE). 2009
[19] K. HANANE & al. « Mise en place d'un systeme de recherche d'image par le contenu ». Memoir de master. Université Tlemcen 2021
[20]: A. Latif & al. « Content-Based Image Retrieval and Feature Extraction: A Comprehensive Review » . Review Article Vol Article ID 9658350. 2019.
[21]: J.R. Smith & S.F Chang. Visualssek : “a fully automated content based image query system”. in ACM Multimedia Conference , pages 87-98, 1996.
[22]: P. Gros & al. « Utilisation de la couleur pour l’indexation et l’appariement d’images ”. Technical Report RR- 3269, INRIA, 1997.
[23]: I.Gagliari and R. Shettini. “A method for the automatic indexing od color images for

effective image retrieval". The New Review of Hypermedia an multimedia, 3 :201 224 , 1997
[24]: R.Kasturi and al, "An evaluation of color histogram based methods in video indexing". Technical report, CSE-96-053, Penn State University, Departement of Computer Science and Engineering, 1996
[25]: S. Sural; G. Qian; S. Pramanik "Segmentation and histogram generation using the HSV color space for image retrieval". IEEE ICIP 2002
[26]: X.-Y. Wang, B.-B. Zhang, and H.-Y. Yang, "Content-based image retrieval by integrating color and texture features," Multimedia Tools and Applications, vol. 68, no. 3, pp. 545– 569, 2014
[27] : C. A. Hussain, D. Rao and S. Mastani, "Low level feature extraction methods for content based image retrieval," in *Electrical, Electronics, Signals, Communications and Optimization (EESCO) 2015*, Visakhapatnam, 2015.
[28]: B. Saïda. "Recherche d'image par le contenu". Mémoire de magister. Université de Tizi-Ouzou 2011
[29] M.A. Stricker et M. Orengo. "Similarity of color images". In SPIE, Storage and Retrieval for image Video Databases, pages 381-392. pàà 1995.
[30]: L. V. Tran. "Efficient Image Retrieval with Statistical Color Descriptors". Dissertation No. 810. Department of Science and Technology Linköping University, Norrköping, Sweden May 2003.
[31]: Y. Akalal. "Moments de Zernike pour la recherché d'images par le contenu". Mémoire de magister. Université du Québec en Outaouais 2019
[32]: F. Policarpo, The Computer Image, ACM Press. Pages 298-308. 1998
[33]: S. Fekri-Ershad. "Innovative Texture Database Collecting Approach and Feature Extraction Method based on Combination of Gray Tone Difference Matrixes, Local Binary Patterns, and K-means Clustering". Paper(CCITIC 2014). Department of Computer Science and Engineering 2014.
[34]: T. Ahonen; A. Hadid; M. Pietikäinen. "Face recognition with local binary patterns". In Proceedings of the European Conference on Computer Vision, Prague, Czech Republic, 11–14 May 2004; pp. 469–481 In Proceedings of the International Conference on Image Processing, Rochester, Volume 2, p. II ; NY, USA, 22–25 September 2002.
[35]: R. M. Haralick. "Statistical and Structural Approaches to Texture". Proceeding of the IEEE, VOL 67 No. 5, may 1979.
[36]: L. Hamza & al. "Indexation Et Recherche D'image Fixe Basé Sur Le Contenu". Master en informatique. Université Mohammed Seddik Benyahia jijel 2020
[37]: M. Sonka, V. Hlavac, R. Boyle. "Image Processing, Analysis and Machine Vision". PWS Publishing, seconde edition, 1999.
[38]: M.R. Teague: "Image analysis via the General Theory of moments, Applied optics", vol. 19, n° 8 (1980), pp. 1353-1356, 1980
[39]: T. S. Lai, "CHROMA, a photographic image retrieval system", PhD thesis, School of computing, engineering and technology, University of Sunderland, UK, 2000
[40]: Y. Fataicha. "Recherche d'information dans les images de documents". Thèse de Ph.D. Université du Québec 2005.
[41]: J. Cohen & al. "MobileNet SSD : étude d'un détecteur d'objets embarquable entraîné sans images réelles". ORASIS 2021, Centre National de la Recherche Scientifique [CNRS], Saint Ferréol, France, 2021.

Sitographie :

[Site 1]: http://www.crdp.ac-grenoble.fr/image/general/general.htm . Visité le 06/06/2022
[Site 2]: https://studylib.net/doc/10086294/image-science-and-technology Visité le 02/06/2022
[Site 3] : https://journalofbigdata.springeropen.com/articles/10.1186/s40537-021-00444-8 .Visité le 02/06/2022
[Site 4] : https://iq.opengenus.org/different-types-of-cnn-models/ Visité le 02/06/2022
[Site 5] : https://blog.devgenius.io/resnet50-6b42934db431 .Visité le 02/06/2022