

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITE MOULOU D MAMMERI DE TIZI-OUZOU



FACULTE DU GENIE ELECTRIQUE ET D'INFORMATIQUE  
DEPARTEMENT D'INFORMATIQUE

## Mémoire de Fin d'Etudes de MASTER ACADEMIQUE

Domaine : Mathématiques et Informatique

Filière : Informatique

Spécialité : **Conduit de projet informatique**

*Présenté par*

**Siham BOULAHIA**

**Chahrazed DJEMA**

Thème

# **Local features vs global features pour la reconnaissance des plaque d'immatriculation**

*Mémoire soutenu publiquement le 29/10/ 2016 devant le jury composé de :*

**Président : M Rabah RAHMANI**

**Encadreur : M Ahmed DIB**

**Examineur : M**

**Examineur : M**

# *Remerciements*

*L*ouanges à dieu, tout puissant, qui nous a donné la force et le courage d'avoir accompli ce travail.

*Nous tenons à exprimer nos vifs remerciements à notre promoteur Mr DIB, qui nous a fait l'honneur de diriger ce travail et pour son aide toujours pertinente, ses conseils et sa grande disponibilité.*

*Que les membres de jury trouvent ici nos remerciements les plus vifs pour avoir accepté d'honorer par leur jugement notre travail.*

*Enfin, nous remercions toutes les personnes ayant contribué d'une façon ou d'une autre à réaliser ce projet de fin d'études.*

SIHAM, CHAHRAZED

# Dédicaces

**N**ous dédions ce modeste travail :

- A tous ceux qui nous ont quittés, à tous les membres de nos familles (Grands-parents, frères, sœurs, cousins, cousines, oncles et tantes, nièces, neveux, à nos parents surtout qui nous ont soutenus) et à tous nos ami(e)s qui se reconnaîtront ;
- A tous ceux que nous avons connus de près ou de loin.

SIHAM, CHAHRAZED

# Le sommaire

## Introduction générale

### Chapitre I : les systèmes de vidéosurveillance

#### [I Introduction](#)

#### [II Concept de base](#)

##### [II.1 La vidéo](#)

##### [II.1.1 Définition](#)

##### [III.1.2 Les types de la vidéo](#)

##### [II.1.3 Indexation et recherche de vidéo](#).....

#### [III La vidéo surveillance](#)

##### [III.1 Définition :](#)

##### [III.2 Technologie :](#)

##### [III.2.1 Architecture des systèmes de vidéosurveillance](#)

##### [III.3 Evolution des systèmes de vidéosurveillance](#)

##### [III.4 Architecture d'un réseau de vidéosurveillance intelligent](#)

##### [III.4.1 Architecture centralisée](#)

##### [III.4.1 Architecture distribuée](#)

##### [III.5 Etat de l'avancement technologique en analytique vidéo](#)

##### [III.6 Tendances technologiques pour la vidéosurveillance](#)

##### [III.7 Domaines d'application](#)

##### [III.8 Indexation et recherche de vidéo surveillance](#)

#### [IV Conclusion](#)

### Chapitre II : Traitement d'image

#### [I Introduction](#)

#### [II Quelques notions de terminologie](#)

##### [II.1 Définition de l'image](#)

##### [II.2 Les attributs de l'image](#)

##### [II.3 Les types d'images](#)

##### [II.4 Les formats d'image](#)

#### [III Le traitement d'image](#)

##### [III.1 Introduction à la vision par ordinateur](#)

##### [III.2 Acquisition d'une image à partir d'une séquence vidéo](#)

##### [III.3 Avantages et inconvénients du traitement d'image](#)

# Le sommaire

[III.4 Exemples d'emploi du traitement d'image](#)

[IV Indexation et recherche d'image](#)

[IV.1 La recherche locale vs globale](#)

[IV.1.1 Ontologie](#)

[IV.1.2 Retour de pertinence](#)

[V Les caractéristique](#)

[V.1 Définition d'une caractéristique](#)

[V.2 Les types des caractéristiques](#)

[V.2.1 caractéristiques au niveau des pixels \(Pixel-level features\)](#)

[V.2.2 Caractéristiques locales \(Local features\)](#)

[V.2.3 globales \(Global features\)](#)

[V.3 Caractéristique local contre caractéristique global](#)

[V.4 Détection de caractéristique](#)

[V.4.1 Définition](#)

[V.4.2 Les détecteurs de caractéristique local](#)

[V.4.3 Détecteurs de caractéristique global](#)

[V.5 Description de caractéristique](#)

[V.5.1 Descripteur de caractéristique local :](#)

[V.5.2 Descripteur de caractéristique global](#)

[VI Conclusion](#)

## **Chapitre III :les systèmes de reconnaissance de plaque d'immatriculation**

[I Introduction](#)

[II Présentation de RAPI](#)

[II.1 Définition d'une plaque d'immatriculation](#)

[II.2 système de reconnaissance automatique des plaques d'immatriculation\(RAPI\)](#)

[III Domaines d'utilisation des RAPI](#)

[IV Caractéristiques des Plaques d'immatriculation de Véhicule algérien](#)

[V Quelques méthodes existantes pour la reconnaissance des plaques d'immatriculation](#)

[V.1 Utilisation des informations globales](#)

[V.1.1 méthode basée sur la CCA\(Connected Component Analysis \)](#)

[V.1.2 méthode basée sur la PCA\(L'Analyse en Composantes Principales \)](#)

[V.1.3 Utilisation des méthode d'extraction de caractéristiques GLCM \(Gray Level Co-occurrence Matrix\) et EDMS\(edge direction matrix\)](#)

[V.1.4 Méthodes basées sur les couleurs et les formes des plaques d'immatriculation](#)

[V.2 Utilisation des informations locales](#)

# Le sommaire

[V.2.1 La reconnaissance des caractéristique a laide la méthode surf](#)

[V.2.2 Maximally Stable Extremal Region \(MSER\) :](#)

[V.2.3 basé sur SOFT \(La méthode de Viola et Jones.\)](#)

[V.3 Les méthodes globales vs les méthodes locales pour la RPIV](#)

[VI Conclusion](#)

[I Introduction](#)

[II Implémentation des systèmes de reconnaissance des plaques d'immatriculation de véhicules](#)

[II.1 La reconnaissance basée sur les caractéristiques locales\(avec MSER\)](#)

[II.1.1 Le principe de MSER](#)

[II.1.2 Multi - résolution MSER](#)

[II.1.3 Implémentation](#)

[II.1.4 Résultats et caractérisation](#)

[II.2 Reconnaissance des plaques d'immatriculation avec GLCM et EDMS \(approche global\)](#)

[II.2.1 La méthode GLCM :](#)

[II.2.2 Implémentation](#)

[II.2.3 Résultats et caractérisation](#)

[III comparaison entre les resultats des deux methodes](#)

[IV Conclusion](#)

# Table des figures et liste des tableaux

## Table des figure

### Figure de chapitre I

Figure I.1 : Représentation de vidéos non scénarisées.

Figure I.2 : des caméras de surveillance

Figure I.3 : Les composantes d'un système de vidéo surveillance

Figure I.4 : Camera IP en méga pixel

Figure I.5: les étapes d'indexation

Figure I.6 : les approches basées sur la fusion précoce

Figure I.7 :les approches basées sur la fusion tardive.

Figure I.8 : Processus de génération PA,SCATet MCAT .

Figure I.9 :Appariement des objets.

Figure I.10: module vision (Intermediate Vision) pour la détection et suivi des objets.

### Figure de chapitre II

Figure II.1 : Représentation d'une lettre sous la forme du groupe de pixels.

Figure II.2 :Représentation de dimension d'une image.

Figure II.3 :image représente les différents sorte du bruit.

Figure II.4 :image en niveau de gris

Figure II.5 : Image en couleur

Figure II.6 :image en mode monochrome

Figure II.7 :image matricielle

Figure II.8 :image vectorielle

Figure II.9 : Représente l'acquisition d'image à partir d'une séquence vidéo

Figure II.10:aperçu d'un système d'indexation et de recherche d'image.

Figure II.11 :Indexation et recherche d'images avec une ontologie.

Figure II.12 :correspondance entre des descripteurs a bas niveau et des termes à intermédiaire niveau pour la luminance.

Figure II.13 : Illustration de retour de pertinence basé (a) sur la modification de requête (b) et sur le rôle de chaque descripteur dans le cas où deux descripteurs  $f_1$ ,  $f_2$  sont utilisés

Figure II.14: l'importance des coins et des jonctions dans la reconnaissance visuelle

Figure II.15 : Différents types de points d'intérêts ( coins, jonction en T et point de fortes variations de texture)

Figure II.16 :détecteur de point de harris

Figure II.17 : Détection de points d'intérêt invariants à l'échelle

Figure II.18 : Illustration de la segmentation d'une image en superpixels

Figure II.19 : Principe de l'image intégrale

Figure II.20 : les étapes à suivre pour créer bof

Figure II.21 :Détermination de l'angle de recalage du SURF

Figure II.22 :Masque d'analyse de la méthode GLOH

Figure II.23 :Exemple de gardiens d'une image

Figure II.24: L'extraction de l'enveloppe spatiale d'image avec le descripteur GIST

### Figure de chapitre III

## **Table des figures et liste des tableaux**

**Figure III.1 : *Etapes d'un RAPI.***

**Figure III.2 : Gestion des véhicules prioritaires**

**Figure III.3 : *Péage de la SR91, Californie***

**Figure III.4 : la plaque d'immatriculation algérienne**

**Figure III.5 : la segmentation des caractères en utilisant des opérations morphologiques**

### **Figure de chapitre IV**

**Figure IV.1 : Pyramide à l'échelle avec une octave entre les échelles.**

**Figure IV.2 : Les résultats de la détection MSER**

**Figure IV.3 : résultats de détection de la plaque d'immatriculation**

**Figure IV.4 : Quelques exemples de plaques d'immatriculation utilisées dans le système proposé**

### **Liste des tableaux**

**Tableau III.1 : tableau de comparaison des résultats des méthodes présentées**

**Tableau IV.1 : temps d'exécution par image**

**Tableau IV.2 : la performance par étape**

**Tableau IV.3 : taux d'exécution**

**Tableau IV.4 : comparaison des résultats des méthodes proposées**

# Introduction générale

L'essor du numérique dans le monde audio-visuel s'est propagé depuis le son vers l'image et vu son influence remarquable dans la vie moderne, on a accordé une primauté de l'image dans le domaine de la communication. En effet, l'image représente le moyen de communication le plus efficace, le plus attirant et le plus approprié pour transmettre l'information entre différentes personnes. De nombreuses opérations relevant de l'analyse d'image et auparavant effectuées "manuellement" peuvent aujourd'hui être résolues automatiquement par des systèmes de vision artificielle, et ceci dans des domaines très diversifiés. Nous pouvons citer entre autre le contrôle de qualité lors de la fabrication de matériaux, les systèmes d'assistance à la conduite de véhicules les applications en imagerie médicale,... ou encore les systèmes automatiques de reconnaissance de plaques d'immatriculation avec la vidéo surveillance qui correspond au cadre dans lequel notre travail s'est déroulé.

Les avancées dans la technologie de reconnaissance des plaques d'immatriculation (RAPI) ont permis aux agences policières à travers le monde d'appréhender plus de suspects recherchés et de récupérer plus de véhicules volés, tout en améliorant la sécurité des agents en service. Choisir le bon système RAPI est cependant crucial pour assurer une solution fiable afin de détecter les véhicules suspects et les menaces possibles. En général, tout système de reconnaissance de plaque d'immatriculation se compose de trois parties : la détection, la localisation et la reconnaissance de la plaque d'immatriculation. Ces tâches sont réalisées par des opérations de traitement d'image. L'algorithme de la localisation de la plaque du véhicule localise l'emplacement de la plaque dans l'image. La phase de reconnaissance identifie les numéros de la plaque d'immatriculation du véhicule. La reconnaissance représente un grand défi des images de véhicule, à cause des différents formats des plaques, les différents styles, la prise de vue et la variation des conditions d'éclairage lors de l'acquisition des images de véhicule. Plusieurs méthodes ont été proposées pour la détection et la reconnaissance de plaques, que ce soit en utilisant les caractéristiques ou informations globales d'une image contenant une plaque d'immatriculation qui décrivent l'image dans son ensemble ou en utilisant les caractéristiques locales calculées sur les résultats de la subdivision de la bande image sur la segmentation d'images ou de détection des bords. La comparaison entre ces deux approches a été réalisée dans plusieurs domaines tels la vidéo synthèse, la reconnaissance d'objets, les systèmes CBIR (*Content-Based Image Retrieval*) ou système de recherche et d'indexation d'image.

Ce travail se focalise essentiellement sur le problème de choix et de comparaison des caractéristiques (globales ou locales) pour la reconnaissance des plaques d'immatriculation en terme de performance qui a été et sera l'objectif majeur des méthodes de reconnaissance en particulier, autrement dit « quelles méthodes sont les plus performantes pour la reconnaissance des plaques d'immatriculation, celles basées sur les caractéristiques locales ou les méthodes basées sur les caractéristiques globales ? ». nous allons en premier lieu citer quelques travaux réalisés dans le cadre de la reconnaissance de la plaque, puis grâce aux résultats expérimentaux de ces dernières nous allons pouvoir classer ces méthodes et élire les plus performantes.

L'objectif principale de ce travail est d'aider à la décision du choix de caractéristique qui sera prise lors de la conception d'un système RAPI.

Pour la réalisation de cette comparaison, nous avons élaboré un plan s'articulant sur quatre (4) chapitres :

Le premier chapitre intitulé « les systèmes de vidéo surveillance » dans le quelle nous présentons laconiquement les systèmes de vidéo surveillance, leur évolution, leurs domaines d'utilisation leur architecture et leurs tendances

Dans le deuxième chapitre , nous introduisons les notion de « caractéristique locale et globale » ainsi que la détection et l'extraction de ces dernières après une présentation des différentes définitions à la base de ce concept sur l'image et le traitement d'image ,quelques descripteurs (locaux et globaux) seront aussi présentés avec concision.

Le troisième chapitre sera consacré au systèmes RAPI (Reconnaissance Automatique des Plaques d'Immatriculation), leurs domaines d'utilisation nous allons aussi citer quelques méthodes déjà existantes que ce soit celles basées sur les caractéristiques locales ou celles basées sur les caractéristiques globales , et à partir des résultats expérimentaux des ces méthodes nous allons réaliser une comparaison de celles-ci en terme de performance.

Dans le quatrième chapitre nous allons proposé deux méthodes que nous jugerons performantes pour tester et ratifier le résultat obtenus dans le troisième chapitre. Les deux méthodes seront présentées et testées par application pour en extraire des preuves.

# chapitre I

*Les systèmes de  
vidéo surveillance*

# Chapitre I : les systèmes de vidéosurveillance

## I. Introduction

Les systèmes de vidéosurveillance existent depuis environ 25 ans. Intégralement analogiques à leurs débuts, ils ont évolué progressivement vers la technologie numérique. Les systèmes actuels ne ressemblent plus guère aux anciennes caméras analogiques branchées sur des magnétoscopes traditionnels. Aujourd'hui, ils utilisent les caméras réseau et les serveurs informatiques pour l'enregistrement vidéo dans un système entièrement numérique. Ces systèmes qu'on trouve dans de nombreux secteurs d'activité (banque, transports, industrie, grande distribution, etc.) ou lieux de vie (villes, immeubles de bureau, équipements collectifs, etc.), sont utilisés pour accroître la sécurité en protégeant les biens et les personnes par de la vidéosurveillance.

Dans ce chapitre, on présentera laconiquement les systèmes de vidéosurveillance, ainsi que ses technologies et ses domaines d'application.

## II. Concept de base

### II.1 La vidéo

#### II .1.1 Définition

Une vidéo peut être définie comme une succession d'images à une certaine cadence. Le principe de l'analogique est de reproduire le signal à enregistrer sous forme similaire sur un support magnétique (en général). Le signal est en forme d'onde, il est très sensible aux perturbations extérieures. Ces perturbations numérique désigne l'ensemble des techniques qui permettent d'enregistrer, de traiter et de lire, la vidéo ne pouvait être enregistrée qu'en mode analogique. L'évolution des technologies micro-informatique permet aujourd'hui d'enregistrer des images vidéo à haute définition au format numérique.

#### II .1.2 Les types de la vidéo

- **Les vidéos scénarisées**

Une vidéo est scénarisée si elle est produite selon un script. Comme Les journaux télévisés et les films.

- **Les vidéos non scénarisées :**

Une vidéo est non scénarisée si elle n'est pas produite selon un script. Comme Les enregistrements de réunion, et la vidéosurveillance. Dans les vidéos non scénarisées, les événements sont spontanés.

#### II .1.3 Indexation et recherche de vidéo

##### II .1.3.1 Indexation et recherche de vidéo scénarisée

L'indexation consiste à analyser la structure et le contenu des vidéos scénarisées. L'analyse de la structure des vidéos scénarisées consiste à représenter une vidéo par ses composantes (plans, images clés, groupes, scènes) tandis que l'analyse du contenu vise à extraire l'information de ces composantes. L'analyse de la structure des vidéos est composée de la segmentation temporelle de vidéos, de la détection des images clés et du regroupement des plans en groupes et en scènes. La vidéo est tout d'abord décomposée en plans (par la segmentation temporelle de vidéos). Les images clés sont ensuite détectées pour chacun des plans (par la détection des images clés). Les plans peuvent être regroupés en groupe et en scène (par le regroupement des plans en groupes et en scènes). La segmentation temporelle de vidéos en plans et la détection des images clés sont le plus souvent présentées dans les approches d'indexation et de recherche de vidéos scénarisées, néanmoins le regroupement les plans en groupes et en scènes est rarement effectué. La segmentation temporelle de vidéos en plans détecte la transition entre les plans. C'est pourquoi, elle est également connue sous le nom de détection de transitions. Les transitions peuvent être soit brusques, soit progressives. Dans le premier cas, on passe directement d'un plan à l'autre, alors que dans le second, un effet est ajouté entre les deux plans. La détection des images clés consiste à

# Chapitre I : les systèmes de vidéosurveillance

déterminer dans un plan un ou plusieurs frames qui représentent significativement le contenu du plan. En général, le premier frame et le dernier sont pris comme images clés. Un ou plusieurs frames qui sont largement différents d'autres frames selon une mesure de similarité sont considérés comme des images clés. Les approches de détection des images clés se différencient par leurs descripteurs, leurs mesures de similarité et leurs stratégies de choix. Le regroupement des plans en groupe et en scène consiste à retrouver le script sur lequel la production de la vidéo s'appuie. En effet, les scènes sont définies comme un groupe de plans cohérents qui présente un sens pour l'utilisateur. Le problème principal repose sur la définition de la cohérence des plans dans une scène. Est-ce une même unité de lieu, de personne, de thématique ? Contrairement aux plans, la définition des scènes repose sur une corrélation sémantique subjective.

## II .1.3.2 L'indexation et recherche de vidéo non scénarisées

L'analyse d'une vidéo non scénarisée est effectuée d'une manière ascendante. Elle consiste à détecter les plays and breaks, identifier les marqueurs auditifs et visuels, déterminer les highlight candidates et les regrouper en highlight groups. L'indexation et la recherche de vidéos de sports, de réunion et de vidéosurveillance font partie de l'indexation et de la recherche de vidéos non scénarisées. Nous remarquons qu'il existe deux types de vidéos de sport : l'un contient des vidéos de sport prises par les caméras et l'autre sont des vidéos de sport télédiffusées. Les vidéos de sport télédiffusées sont créées par la tâche d'édition à partir des plans de vidéos acquis par les caméras. L'édition correspond au choix des plans et à leur montage avant diffusion. Les vidéos du premier type sont des vidéos non scénarisées tandis que celles du deuxième type sont des vidéos scénarisées. Pour les vidéos de deuxième type, après une phase de prétraitement permettant d'identifier les plans, l'analyse de ces plans est celle de vidéo non scénarisée.

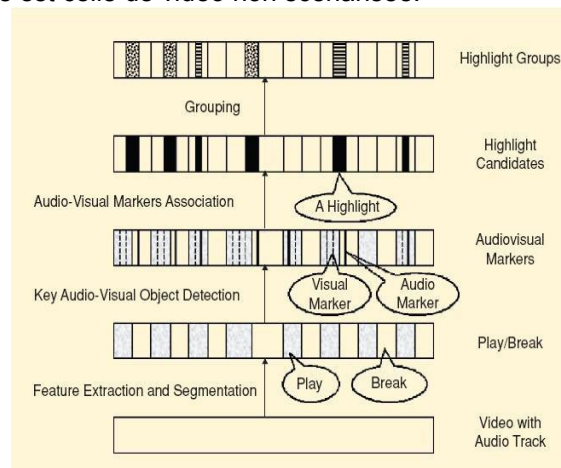


Figure I.1 : Représentation de vidéos non scénarisées.

## III. La vidéo surveillance

### III.1 Définition :

La vidéosurveillance, parfois désignée par le sigle anglais CCTV1, est un système de caméras et de transmission d'images, disposé dans un espace public ou privé pour le surveiller. Les images obtenues avec ce système, peuvent être traité automatiquement et/ou visionnées puis archivées ou détruites. La surveillance a pour but de contrôler les conditions de respect de la sécurité, de la sûreté ou de l'exécution d'une procédure particulière. Selon ses partisans, un tel système permet de prévenir la criminalité (vols, agressions, etc.) et d'opérer un contrôle social (mouvements de foule, etc.). Ses détracteurs lui reprochent son inefficacité, son coût et son atteinte à la vie privée

# Chapitre I : les systèmes de vidéosurveillance



Figure I.2 : des caméras de surveillance

## III.2 Technologie :

### III.2.1 Architecture des systèmes de vidéosurveillance

#### III .2.1.1 Le schéma de principe :

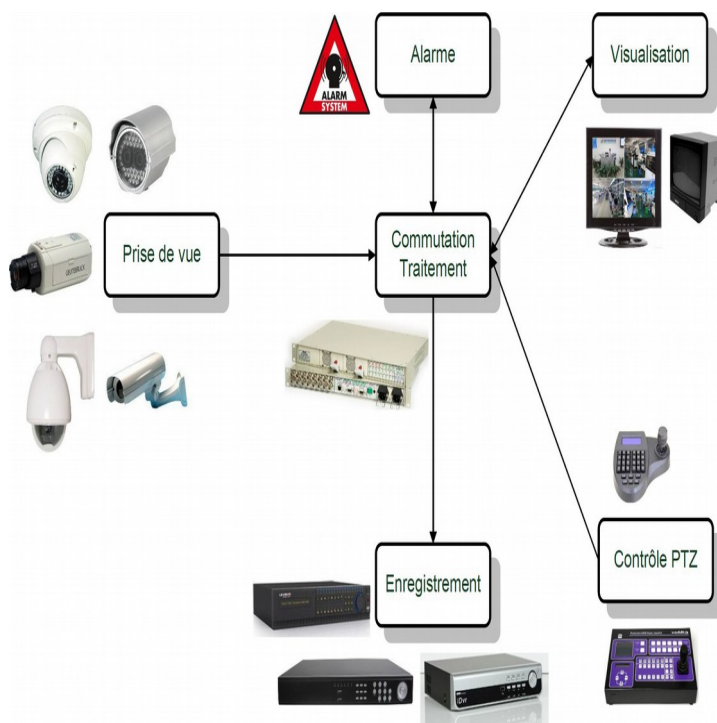


Figure I.3 : Les composantes d'un système de vidéo surveillance

Un système de vidéo surveillance typique se compose de :

- **La prise de vue**

Permet l'acquisition des images à l'aide de caméras plus ou moins perfectionnées. Certaines situations conduisent à utiliser des caméras avec des spécifications bien particulières (étanches, grand-angle, IR, etc....)

Les différentes caractéristiques à prendre en compte lors du choix d'une caméra sont :

## Chapitre I : les systèmes de vidéosurveillance

- Taille du capteur exprimée généralement en fraction de pouce (CCD 1/3' équivaut à un capteur de 8,4mm de diagonale)
- Nombre de lignes constituant l'image (330, 480, 504 ou plus)
- Résolution (définition) de l'image dépendante du nombre de lignes
- Type de monture d'objectif généralement de type CS



Figure I.4 : Camera IP en méga pixel

- **Acquisition**

Il existe une panoplie de modèles de caméras répondant à différents besoins de surveillance. Elles sont analogiques ou numériques et peuvent être motorisées ou non. Plus spécifiquement, on retrouve les types de caméras suivants :

- Fixe : Pointée dans une direction unique, elle couvre une zone définie (une entrée, une portion de stationnement, etc.). C'est la caméra de surveillance traditionnelle. Elle constitue un excellent choix lorsqu'on désire que la présence de la caméra, ainsi que sa direction de surveillance, soient visibles.
- (Pan-Tilt-Zoom): Motorisée, elle peut être actionnée, manuellement ou automatiquement, dans des mouvements panoramique/inclinaison/zoom. Elle sert à suivre des objets ou des individus se déplaçant dans la scène ou à zoomer sur des régions d'intérêt (par exemple, sur une plaque d'immatriculation).
- Dome : Recouverte d'un caisson hémisphérique, ce qui la rend discrète et, dans certains modèles, résistante au vandalisme et aux intempéries. Elle peut être fixe ou mobile. Les versions motorisées couvrent une zone très large, grâce à leur balayage horizontal de 360° et de 180° à la verticale. Bien qu'en « tour de garde », elle puisse remplacer dix caméras fixes en balayant l'aire à surveiller, elle n'observe qu'une seule direction à la fois.
- Mégapixel : Offre une résolution plus élevée que les caméras standards, allant de 1 à 16 mégapixels. Elle permet soit de capter une image plus détaillée, soit de couvrir un plus large champ visuel, réduisant le nombre de caméras nécessaires pour couvrir une aire à surveiller. Lorsqu'utilisée avec un grand angle, elle possède un espace de visualisation allant généralement de 140° à 360°. Offrant la possibilité de zoomer de façon logicielle dans l'image, elle peut ainsi devenir une alternative à la caméra PTZ mécanique qui entraîne l'usure des pièces. Sa résolution élevée contribue à l'amélioration de la performance des algorithmes de

## Chapitre I : les systèmes de vidéosurveillance

détection et de reconnaissance exigeant un haut niveau de détails, telles que la lecture de plaques d'immatriculation et la reconnaissance de visage.

- Infrarouge et thermique : Sensible au rayonnement infrarouge (IR), elle est capable de produire une image de bonne qualité dans le noir pour une surveillance nocturne. De nuit, elle filme en noir et blanc, mais elle peut produire une image couleur le jour. Certaines caméras infrarouges sont équipées de leur propre source de lumière IR, allumée lorsque le niveau d'éclairage chute sous un certain seuil. Des projecteurs IR séparés (lampe ou LED18) peuvent aussi être utilisés. Les caméras thermiques enregistrent le rayonnement de chaleur des objets. Elles ne requièrent aucune source d'illumination.
- Panoramique : Grâce à une optique spéciale, elle offre 360° de visibilité avec une seule caméra. Elle permet un PTZ virtuel dans l'image. Les principales technologies panoramiques pour la surveillance sont le fisheye, la lentille à miroirs et la lentille panomorphe. Toutefois, la résolution de ces caméras est souvent insuffisante pour des analyses nécessitant un niveau de détail élevé.

- **Transmission**

La vidéo captée par les caméras de surveillance doit être transmise aux systèmes d'enregistrement, de traitement et de visionnement. Cette transmission peut se faire par câble (câbles coaxiaux ou à fibre optique, fils de cuivre torsadés) ou à travers l'air (signaux infrarouges, transmission radioélectrique).

La vidéo filaire prédomine largement dans les systèmes de vidéosurveillance. Elle offre une plus grande bande passante et une meilleure fiabilité que les connections sans fil, à un coût inférieur. Cependant, la transmission vidéo sans fil s'impose parfois comme solution, par exemple dans le cas de surveillance de grands périmètres où l'installation de câblage s'avérerait trop coûteuse, ou lorsque les zones à surveiller sont impossibles à rejoindre par câble

Qu'il transite par fil ou sans fil, le signal vidéo peut être analogique ou numérique. Encore aujourd'hui, la majeure partie des transmissions vidéo pour la surveillance sont **analogiques**. Néanmoins, les réseaux informatiques (LAN, WAN ou Internet) sont de plus en plus utilisés pour transporter la vidéo grâce au protocole IP. Les caméras IP peuvent se connecter directement sur ces réseaux, tandis que les flux vidéo émergeant de caméras analogiques doivent, au préalable, être numérisés par un **encodeur**, aussi appelé **serveur vidéo**, pour passer par les réseaux IP .

- **Compression :**

La vidéo numérisée représente une grande quantité de données à transmettre et à archiver. L'envoi d'une séquence vidéo peut nécessiter jusqu'à 165 mégabits de bande passante et la vidéo d'une seule caméra pour une journée peut occuper sept Giga octets d'espace disque. C'est pourquoi la vidéo de surveillance doit être compressée grâce à des **codecs**, algorithmes permettant de réduire la quantité de données en supprimant les redondances, par image ou entre les trames d'une séquence, ainsi que les détails imperceptibles à l'œil humain . Selon le type de compression, l'usage du processeur requis pour l'exécution du codec est plus ou moins intensif. Un compromis s'impose donc entre le taux de compression et les ressources du processeur qui sont accaparées .

Il existe deux grands groupes de standards internationaux de compression : JPEG, créés par le Joint Photographic Experts Group, et MPEG, élaborés par le Moving Photographic Experts Group. Dans le premier groupe, on retrouve les formats JPEG pour les images fixes, et MJPEG pour les séquences vidéo. Le second groupe comprend les formats MPEG-1, MPEG-2, MPEG-4 et H.26421 .

À l'heure actuelle, MJPEG et MPEG-4 sont les standards les plus répandus en vidéosurveillance. Toutefois, avec les améliorations en qualité et efficacité (taux de compression, latence, résistance à l'erreur) qu'il apporte, H.264 devrait bientôt remplacer MPEG-4. En effet, sans affecter la qualité de l'image, l'encodeur H.264 permet de réduire la taille de celle-ci de plus de 80 % par rapport à la compression MJPEG, et de 50 % par rapport à la compression MPEG-4.

- **Traitement**

# Chapitre I : les systèmes de vidéosurveillance

Les systèmes de gestion vidéo opèrent les traitements des images de vidéosurveillance, tels que la gestion des différents flux vidéo, le visionnement, l'enregistrement, l'analyse et la recherche dans les séquences enregistrées. Il existe quatre grandes catégories de systèmes de gestion vidéo.

- **Enregistreur vidéo numérique (DVR22):** Appareil qui dispose d'un disque dur interne pour l'enregistrement numérique de la vidéo et d'un logiciel intégré de traitement de la vidéo. Il n'accepte que les flux provenant de caméras analogiques, qu'il numérise. Les modèles récents permettent de visionner la vidéo à distance sur ordinateur. Encore très répandus, ils laissent toutefois peu à peu leur place aux systèmes supportant la vidéo IP de bout en bout.
- **Enregistreur vidéo hybride (HDVR23):** Similaire à l'enregistreur numérique, mais accepte à la fois le branchement de caméras analogiques et IP. Il est possible de rendre hybrides plusieurs modèles d'enregistreurs vidéo numériques par l'ajout d'un logiciel.
- **Enregistreur numérique réseau (NVR24):** Conçu pour les architectures réseaux IP de vidéosurveillance, il ne peut traiter que les signaux vidéo provenant de caméras IP ou d'encodeurs.
- **Logiciel de vidéosurveillance IP:** Solution purement logicielle de gestion de la vidéo sur un réseau IP. Dans le cas de systèmes de surveillance comportant peu de caméras, un navigateur Web peut suffire à gérer la vidéo. Pour de plus gros réseaux de vidéosurveillance, un logiciel dédié de gestion vidéo doit être utilisé. Celui-ci s'installe sur un ordinateur personnel ou un serveur. Plus complexe à installer, en raison des configurations nécessaires du serveur, il offre une plus grande flexibilité pour le choix et l'ajout de composantes au réseau de vidéosurveillance. Les logiciels de vidéosurveillance IP représentent une tendance forte en gestion vidéo, surtout dans les infrastructures comportant un grand nombre de caméras. Les plateformes ouvertes permettent d'intégrer facilement des caméras et composantes matérielles de différents fabricants.
- **Archivage**

La période d'archivage des séquences vidéo varie selon les besoins de surveillance, allant de quelques jours à quelques années. En moyenne, les organisations conservent les preuves vidéo entre 30 et 90 jours. Le déploiement de vastes réseaux de caméras et l'utilisation de vidéosurveillance à haute résolution fait exploser les demandes pour les systèmes de stockage. Bien que le coût des supports d'enregistrement ait considérablement baissé dans les dernières années, l'archivage représente souvent une part importante des dépenses d'infrastructure en vidéosurveillance, en raison de la quantité toujours croissante de données vidéo à stocker. Les solutions de stockage sont de deux types:

Interne : Les disques durs intégrés aux enregistreurs vidéo numériques ou aux serveurs représentent la forme d'archivage la plus répandue. Elle peut offrir jusqu'à quatre téraoctets d'espace. Certaines caméras IP disposent même d'une carte mémoire ou d'un disque USB permettant d'enregistrer des heures, voir des jours de vidéo. Les solutions internes d'archivage conviennent bien pour les systèmes de vidéosurveillance de taille modeste, comprenant jusqu'à 50 caméras.

L'archivage se fait sur des appareils externes aux enregistreurs ou serveurs vidéo. De type NAS (*Network Attached Storage*) ou SAN (*Storage Area Network*), ces systèmes offrent un espace de stockage partagé entre les différents clients du réseau. Sur un système de stockage en réseau NAS, un fichier est archivé sur un même disque dur, alors qu'avec le réseau de stockage SAN, un fichier peut être sauvegardé en fragments répartis sur plusieurs supports de stockage. Ces solutions d'archivage rattachées sont particulièrement avantageuses pour les grands réseaux de vidéosurveillance comportant un grand nombre de caméras<sup>25</sup>. Bien que plus onéreuses que les systèmes internes d'archivage, ces solutions rattachées sont supérieures en termes d'extensibilité, de flexibilité et de redondance

## III.3 Evolution des systèmes de vidéosurveillance

La transition numérique en vidéosurveillance s'est opérée en plusieurs étapes. Amorcée avec l'apparition de l'enregistreur numérique, elle se poursuit vers une conversion totale à l'infrastructure IP,

# Chapitre I : les systèmes de vidéosurveillance

où la vidéo est transmise sur réseau intranet ou Internet de la caméra à l'écran de visionnement. Dans ce passage, l'on retrouve plusieurs systèmes hybrides, intégrant composantes analogiques et numériques. La figure ci-dessous présente les jalons importants qui ont marqué l'évolution vers la vidéosurveillance intelligente sur réseau IP

## III.3.1 Première génération : le tout analogique

Un système de vidéosurveillance analogique utilisant un magnétoscope traditionnel (VCR) est un système entièrement analogique dans lequel les caméras analogiques avec sorties coaxiales sont reliées au magnétoscope pour l'enregistrement. Les bandes utilisées sont identiques à celles utilisées par les particuliers. Les séquences vidéo ne sont pas compressées. Dans le cas d'un enregistrement à vitesse maximale, une cassette a une durée maximale de 8 heures. Dans les systèmes de plus grande envergure, un quad ou multiplexeur peut être connecté entre la caméra et le magnétoscope. Le quad/multiplexeur permet alors d'enregistrer le contenu de plusieurs caméras sur un même magnétoscope, mais selon une fréquence d'image cependant inférieure. La surveillance vidéo nécessite un moniteur analogique

- **Avantage:**

Les systèmes analogiques sont très fiables et Simple à utiliser, ils ne requièrent pas de compétences informatiques

- **Inconvénients**

La qualité de la vidéo est inférieure à celle des systèmes numériques.

Il faut changer les cassettes fréquemment (aux trois jours ou plus).

Nécessite un nettoyage et un entretien régulier des magnétoscopes.

La qualité de la vidéo enregistrée se détériore avec le temps.

Ne permet pas le visionnement à distance, comme sur les réseaux numériques.

Ce sont des systèmes propriétaires.

## III.3.2 Deuxième génération (le système hybride)

Un système de vidéosurveillance analogique utilisant un enregistreur numérique (DVR) est un système analogique permettant l'enregistrement numérique des images. Avec un enregistreur numérique, l'enregistrement vidéo ne se fait plus sur bandes magnétiques mais sur des disques durs où les séquences sont numérisées et compressées de manière à emmagasiner chaque jour un maximum d'images. Les premiers enregistreurs numériques disposaient d'un espace disque limité. La durée des enregistrements était donc assez restreinte, à moins de réduire la fréquence d'images. Grâce aux progrès récents dans ce domaine, l'espace disque ne pose plus réellement problème.

- **Avantage**

Utilisation d'ordinateurs et de matériel standards de réseau pour l'enregistrement et la gestion de la vidéo

Possibilité d'enregistrer à l'extérieur du site de surveillance (par exemple, centralisation des enregistrements).

Architecture distribuée qui offre flexibilité, extensibilité (peut ajouter une caméra à la fois) et redondance (en cas de bris ou pannes)

- **Inconvénients**

Gourmand en bande passante, si l'enregistrement est fait hors du site de surveillance (par ex., de façon centralisée).

Si le réseau tombe en panne, l'enregistrement peut être interrompu.

Si l'enregistrement centralisé n'est pas requis, l'utilisation d'enregistreurs vidéo réseau est souvent plus coûteuse que celle d'enregistreurs numériques.

Nécessitent des calculs complexes pour déterminer le nombre de flux vidéo pouvant être supportés par le serveur, la quantité d'espace disque nécessaire à l'enregistrement, le taux de trames, le niveau de compression et d'autres facteurs liés aux capacités du réseau.

## III.3.3 Troisième génération : le tout numérique IP

## Chapitre I : les systèmes de vidéosurveillance

Au sens strict, un système de vidéosurveillance est complètement IP, lorsque toutes ses composantes sont numériques et toutes les transmissions sont effectuées sous le protocole IP. Ces réseaux comprennent donc des caméras réseau, aussi appelées IP. Il s'agit de caméras intégrant leur propre encodeur. Celles-ci sont reliées, via des commutateurs réseau, à des serveurs (ordinateurs personnels), munis d'un logiciel de gestion vidéo. L'enregistrement est fait sur serveur ou sur des enregistreurs vidéo réseaux propriétaires. Les traitements sont effectués sur le serveur ou dans les périphériques. Toutefois, beaucoup de gens considèrent qu'un système de vidéosurveillance dont la vidéo est transmise sous protocole IP à partir des encodeurs, constitue un système réseau IP. Dans ces systèmes, les caméras peuvent être analogiques, en autant qu'elles soient reliées à des encodeurs.

Plus de détails sur la vidéosurveillance IP et, notamment ses avantages et inconvénients, sont fournis dans ce qui suit :

Depuis deux ans, le marché de la vidéosurveillance migre de la vidéo analogique à la vidéo IP. Cet essor de la vidéosurveillance réseau est favorisé par le perfectionnement des processeurs, la baisse des coûts de stockage et l'amélioration des algorithmes de compression. Bien que cette transition s'observe, elle ne fait que débiter et il est difficile de prédire à quel rythme le « tout IP » prendra le dessus sur les technologies analogiques en vidéosurveillance. Les caméras constituent le facteur de résistance. En effet, il est estimé que 95 % des 40 millions de caméras installées à travers le monde sont encore analogiques [1]. De plus, encore aujourd'hui, seulement un acheteur sur quatre choisit des caméras IP. Pourtant, les caméras IP offrent des avantages techniques sur les caméras analogiques, tels qu'une meilleure qualité d'image, une résolution plus élevée et de l'intelligence embarquée. Le coût élevé des caméras IP, environ le double des caméras analogiques, reste le principal frein à l'achat. Avec les améliorations apportées par les manufacturiers aux enregistreurs vidéo et l'existence d'enregistreurs vidéo hybrides, le règne des caméras analogiques pourrait se prolonger encore quelques années.

Les réseaux IP de vidéosurveillance de dernière génération présentent les caractéristiques suivantes :

caméras, réseaux, enregistrement, accès.

Ils sont numériques de A à Z :

Ils utilisent du matériel non propriétaire

Ils fonctionnent avec des serveurs distribués et multiplateformes.

On peut y accéder à distance, n'importe où, n'importe quand, soit d'un centre de contrôle, via Internet ou un réseau LAN ou WAN, sur un cellulaire ou un assistant numérique personnel, etc.

Ils peuvent inclure de l'analytique vidéo

La vidéosurveillance sur réseau IP offre de nombreux avantages. Elle repose sur une infrastructure plus souple que la vidéo analogique, combinant transmission câblée et sans fil. De plus, l'infrastructure réseau est rapide et facilement extensible. Il n'y a pas de limite au nombre de caméras qui peuvent s'y ajouter. Considérant que les systèmes actuels peuvent facilement comporter entre 400 et 500 caméras, comparativement à 25 à 30 il y a quatre ans, cette facilité d'étendre le réseau de vidéosurveillance est très attrayante lorsqu'on pense déployer un nouveau système. Pour les installations étendues comportant de grandes distances entre les caméras et les systèmes d'enregistrement, les caméras IP deviennent un choix avantageux, car le coût du câble coaxial est très élevé. C'est pourquoi les caméras IP sont surtout déployées pour la surveillance d'institutions, telles que les établissements scolaires, les campus de grandes entreprises, les municipalités et les bases militaires. Les caméras IP apportent de nombreux bénéfices techniques. Par exemple, les caméras mégapixel IP offrent des avantages indéniables au niveau de la vidéosurveillance. Ces caméras, présentant une résolution nettement supérieure à celle des caméras analogiques, ont le potentiel de résoudre plus de crimes. Bien que son coût soit élevé, une seule caméra mégapixel peut souvent remplacer plusieurs caméras. Les caméras intelligentes, intégrant des processeurs, apportent des gains techniques aux systèmes de vidéosurveillance de nouvelle génération. Elles permettent de distribuer les calculs pour les traitements d'analytique vidéo et, ainsi, de ménager la bande passante en ne transmettant que les données pertinentes pour la sécurité. L'architecture ouverte des réseaux de vidéosurveillance IP permet de combiner le matériel de différents manufacturiers. Il est ainsi possible de sélectionner les composantes les plus adaptées à son application de surveillance. De plus, l'architecture IP facilite l'intégration de différents systèmes de sécurité (vidéo, alarmes de feux, contrôle d'accès, etc.). La configuration de ces réseaux avec redondance, tant au niveau de la transmission que de l'archivage, assure une meilleure fiabilité. Dans un réseau entièrement analogique, le mauvais fonctionnement d'une caméra ou d'un enregistreur peut signifier une perte définitive de vidéo. De plus, l'archivage sur réseau, dont les coûts ne cessent de baisser, permet de conserver des

## Chapitre I : les systèmes de vidéosurveillance

années de vidéo plutôt que quelques jours ou semaines avec les magnétoscopes ou enregistreurs numériques. Il suffit d'ajouter des disques durs pour augmenter la capacité de stockage. Dans ces réseaux, la transmission numérique de la vidéo, de bout en bout du système, réduit les pertes de qualité qu'occasionnent les conversions d'analogique à numérique. De plus, par des traitements logiciels intelligents, il est possible de filtrer les événements pertinents dans la vidéo et de ne transmettre que les métadonnées les décrivant. Ceci permet d'économiser de la bande passante. Ces métadonnées servent ensuite à indexer le contenu vidéo pour permettre des recherches plus performantes dans les séquences captées. Il est aussi possible, dans ces réseaux, de transmettre des commandes aux caméras, en fonction de la vidéo traitée. Par exemple, en réponse à un événement suspect, le logiciel de gestion vidéo pourra activer automatiquement le zoom optique ou numérique sur une caméra du réseau pour un suivi plus précis de la situation.

Le passage à la vidéosurveillance IP comporte tout de même certains inconvénients. Par exemple, dans le cas d'infrastructures analogiques existantes, refaire tout le filage réseau peut représenter un coût élevé. Dans ce cas, l'utilisation d'encodeurs pour numériser les signaux analogiques permet de conserver les caméras en place et d'implanter un réseau hybride. De plus, la technologie des enregistreurs numériques a beaucoup progressé, afin de rester compétitive. Les enregistreurs numériques offrent désormais plusieurs avantages comparables aux technologies sur réseaux IP : accès à distance, capacité de gérer un grand nombre de caméras, intégration avec les autres systèmes (contrôle d'accès, détection d'intrusion, systèmes de points de vente, etc.), capacité analytique. Pour de petites installations de vidéosurveillance, l'utilisation de caméras analogiques avec enregistreurs numériques offre donc un excellent rendement, à des coûts moindres que la vidéo IP.

Bien que sur le plan technologique, les réseaux IP de vidéosurveillance offrent de nombreux avantages, le déploiement de ceux-ci pose certains problèmes organisationnels. Le premier réside dans l'installation. La plupart des intégrateurs et installateurs en vidéosurveillance sont issus du secteur de la sécurité physique. Or, l'installation de systèmes de vidéosurveillance sur réseau IP requiert des compétences en technologie de l'information, particulièrement en réseautique. Encore peu d'intégrateurs et installateurs de vidéosurveillance sont spécialisés dans ce domaine.

Le second problème touche à la gestion de ces systèmes. Une fois installés, qui, dans l'organisation, est responsable du système de vidéosurveillance : le personnel de sécurité physique ou les services informatiques ? Le flou qui existe encore sur cette question dans les organisations retarde bien souvent le passage à la vidéosurveillance IP. Finalement, passer d'un réseau de télévision en circuit fermé à un système de vidéosurveillance, dont les données transitent par les réseaux Ethernet ou Internet, soulève les questions de sécurité informatique. Évidemment, ceci nécessite de mettre en place les mécanismes de sécurité adéquats pour protéger la vidéo de surveillance. Aussi, le bon fonctionnement de la vidéosurveillance devient tributaire de la fiabilité du réseau. À cet égard, l'intégration de la vidéosurveillance sur réseaux IP suivra peut-être celle de la téléphonie IP, qui est maintenant bien répandue.

### III.4 Architecture d'un réseau de vidéosurveillance intelligent

Les systèmes de vidéosurveillance intelligents peuvent être déployés selon deux grands types d'architecture, soit centralisée ou distribuée. L'une comme l'autre opèrent des traitements intelligents pour extraire des données sur les images vidéo et peuvent, au besoin, commander le déplacement d'une caméra PTZ pour opérer une surveillance active.

#### III.4.1 Architecture centralisée

Dans une architecture centralisée, tous les traitements intelligents sont concentrés en un même endroit. Dans les systèmes traditionnels, c'est l'enregistreur numérique qui récoltera tous les flux vidéo pour en faire l'analyse. Pour les infrastructures en réseau, c'est un serveur informatique (un PC) qui opérera les traitements analytiques. L'analyse vidéo nécessite une grande puissance de calcul qui monopolise une part importante des ressources du processeur. Comme ils doivent aussi gérer l'encodage, l'enregistrement et le visionnement des flux vidéo, les enregistreurs numériques et serveurs ne peuvent accomplir des tâches d'analytique vidéo que sur un nombre restreint de caméras. De plus, dans le cas des systèmes de vidéosurveillance sur réseau, la transmission de tous les flux vidéo en un point centralisé consomme beaucoup de bande passante. Le réseau informatique doit pouvoir soutenir ce trafic.

#### III.4.2 Architecture distribuée

# Chapitre I : les systèmes de vidéosurveillance

Comme son nom l'indique, l'architecture distribuée répartit les traitements intelligents dans les différents nœuds du système de vidéosurveillance. Ainsi, les calculs nécessaires à l'analytique pourront être faits en périphérie, sur des caméras intelligentes dotées de processeurs, les encodeurs ou dans les commutateurs réseau. Dans cette architecture, plutôt que d'envoyer des flux vidéo à l'enregistreur numérique ou à un ordinateur central, seules des métadonnées extraites sur la vidéo sont transmises. L'intelligence peut aussi être partagée entre les périphériques et l'unité centrale de traitement. Cette infrastructure aide à diminuer les coûts de câblage et la bande passante requise dans un réseau de vidéosurveillance. De plus, elle offre une meilleure extensibilité, puisque l'ajout de caméras n'est pas nécessairement limité par la puissance de calcul de l'enregistreur numérique ou du serveur.

## III.5 Etat de l'avancement technologique en analytique vidéo

L'analytique vidéo et la vidéosurveillance intelligente sont des domaines de recherche très actifs. En particulier, depuis les dix dernières années, plusieurs conférences, éditions spéciales de revues ou articles y ont été consacrés. Quelques produits commerciaux ont vu le jour suite aux recherches menées dans ce secteur. De grands projets financés par les gouvernements, tels que VSAM, aux États-Unis et PRISMATICA, en Europe, ont favorisé le développement de techniques de détection, suivi et reconnaissance d'objet à partir de séquences vidéo. Toutefois, il s'agit de technologies émergentes. Leur utilisation dans des applications à grande échelle reste encore récente et, somme toute, marginale. Le coût de ces technologies, ainsi que leurs limites techniques actuelles pour la surveillance en environnements complexes, freine une adoption plus généralisée. Cette section dresse un portrait de l'avancement technologique de la vidéosurveillance intelligente. Elle présente les principales analyses que les logiciels d'analytique vidéo peuvent accomplir. On y décrit sommairement les fonctionnalités actuellement implantées dans les systèmes commerciaux, ou en développement dans les laboratoires de recherche. Une énumération des principaux défis que confrontent ces technologies, ainsi que des besoins actuels et futurs qui sont envisagés par les utilisateurs, y est aussi présentée. Une revue plus poussée des différents procédés et algorithmes d'analyse vidéo intelligente, ainsi qu'une bibliographie plus étendue du sujet, peuvent être trouvées dans les publications scientifiques [8, 9, 10, 11, 12, 14].

### III.5.1 Description des techniques d'analytique vidéo

L'analytique vidéo consiste à reproduire, à l'aide d'un ordinateur, l'analyse qu'effectuerait un humain en regardant les séquences vidéo provenant de caméras de surveillance. Les logiciels d'analytique traitent les images d'un flux vidéo pour y détecter automatiquement les objets (personnes, matériel, véhicule) et les événements d'intérêt pour des fins de sécurité. Une fois détectés, les objets peuvent être identifiés, suivis et localisés (Figure 4). Leurs actions et interactions sont analysées et classées, afin d'interpréter l'activité d'une scène pour la soumettre à l'attention d'un opérateur humain. Les logiciels d'analytique vidéo servent en deux modes, soit en temps réel et en temps différé. En temps réel, ils détectent, sur les flux vidéo, les situations représentant une menace à la sécurité et déclenchent une alarme. En temps différé, ils permettent de retrouver plus rapidement les séquences vidéo présentant un incident sous enquête.

- **Alertes en temps réel**

La majorité des alertes sont définies par l'utilisateur du système de vidéosurveillance intelligente. Il peut s'agir d'alertes génériques, telles que la détection d'un objet abandonné ou d'un élément de la scène se déplaçant au-dessus d'un seuil de vitesse fixé. Pour déclencher ces alarmes, seules les propriétés de mouvements des objets sont analysées par le système. Des alertes plus spécifiques peuvent être lancées suite à la classification des objets ou de leur mouvement (par exemple, la discrimination entre le passage d'un humain ou d'un animal dans une zone extérieure). Les alertes comportementales basées sur la conformité ou non à un modèle de comportement enseigné au système (par exemple, un individu tentant d'ouvrir plus d'une voiture dans un stationnement), constituent des alertes prédéfinies. Certaines alertes en temps réel sont identifiées automatiquement. Au fil du temps, le système apprend un modèle d'activité et finit pas détecter les activités dérogeant de la norme. Par exemple, un logiciel d'analytique pourra apprendre que des véhicules circulent dans la rue et les piétons, sur le trottoir. L'inverse pourra sonner l'alarme.

- **Recherche de vidéo pour enquête**

# Chapitre I : les systèmes de vidéosurveillance

Les traitements d'analytique permettent d'indexer le contenu vidéo, en fonction de caractéristiques telles que la forme des objets, leur taille, leur apparence, leur trajectoire, leur type, ainsi que leur modèle d'activité. Stockées sous forme de métadonnées, ces informations permettront d'effectuer des recherches spatiotemporelles du type « trouver toutes les séquences où l'on voit une personne vêtue de rouge passer devant un certain édifice entre deux dates données ». Les fonctions analytiques développées pour les systèmes de vidéosurveillance comportent différents niveaux d'analyse. Hiérarchiquement, elles s'exécutent du niveau des pixels, à celui des objets, pour atteindre l'échelle des comportements. Elles se regroupent selon les tâches suivantes: Détection de changements, Segmentation d'objets en mouvement, Suivi d'objets, Classification et identification d'objets et classification d'activités et de comportements.

## III.5.2 Analytique dans les systèmes commerciaux

Depuis le début des années 2000, plusieurs produits d'analytique vidéo ont été mis sur le marché, afin de permettre une vidéosurveillance intelligente. Une liste des principaux développeurs de logiciels d'analytique vidéo est fournie à l'Annexe 4. Dans l'ensemble, les produits disponibles offrent sensiblement les mêmes fonctionnalités de base. Certains logiciels se distingueront par quelques fonctionnalités plus spécialisées (par ex., reconnaissance de plaques d'immatriculation) ou par leurs performances. Globalement, les principales fonctions d'analytique vidéo disponibles sur le marché se résument à :

- Suivi par caméra PTZ activé sur alarme.
- Détection de bris et d'altérations subis par les camera.
- Détection d'objet disparu ou abandonné .
- Détection d'intrusion (personne ou véhicule) dans des aires ou périmètres contrôlés : limites virtuelles interdites, zones de sécurité, passage de plusieurs individus ou véhicules à un contrôle d'accès sous une seule identification (*tailgating*), intrusion par une voie de sortie, personne passant par dessus ou dessous une guérite.
- Détection et suivi de personnes ou véhicules : analyse de la vitesse et direction de déplacement, détection de personne /voiture / objet stationnaire près d'une infrastructure sensible.
- Comptage de personnes ou véhicules, détection d'attroupements ou surnombre. Détection de vols et de fraudes aux points de ventes.
- Reconnaissance des comportements : détection de rôdeurs, bagarre, course, chutes, déplacement à contresens, graffitis, vandale .
- Monitoring de trafic : vitesse/densité/direction du trafic, accidents, objets sur la chaussée, limites de vitesse enfreintes, congestion, piétons, etc.
- Détection et reconnaissance de visages.
- Détection de plaques d'immatriculation.

Les technologies d'analytique vidéo sont encore jeunes et appelées à se développer. Beaucoup d'applications commerciales fonctionnent bien dans des environnements contrôlés. Toutefois, la plupart d'entre elles performant moins bien lorsque déployées chez les clients, dans des cas réels d'utilisation. De nombreux ajustements et paramétrages sont souvent nécessaires pour amener ces systèmes à opérer correctement pour une application spécifique. Des fonctions de base, telles que la détection de mouvement, l'amélioration d'image et la détection d'altérations sur les caméras, sont en général assez robustes pour fonctionner dans une gamme étendue d'applications. Le comptage de personnes, la détection d'intrusion (limites virtuelles) et la reconnaissance de plaques d'immatriculation constituent des analyses suffisamment matures. Bien qu'elles nécessitent certains

## Chapitre I : les systèmes de vidéosurveillance

paramétrages, leur usage se répand. De façon générale, les techniques de détection et de suivi d'objet se sont grandement développées en recherche et certaines d'entre elles sont intégrées dans des systèmes commerciaux, pour des cas simples d'utilisation. Toutefois, beaucoup de travail reste à faire pour développer ces techniques à des niveaux de performance et de robustesse nécessaires pour l'opération en environnements complexes. La détection et le suivi d'objet dans des foules, ou à travers un réseau de multiples caméras, constituent des problèmes difficiles. Il existe plusieurs produits commerciaux spécialisés pour la détection et reconnaissance d'images (voir Annexe 4). Leur efficacité dépend largement de l'application et de la complexité des environnements dans lesquels ils doivent opérer. Généralement, ces technologies requièrent de contrôler les conditions d'acquisition (éclairage, pose, etc.) pour fournir des résultats satisfaisants. Certains comportements sont détectés par les logiciels d'analytique sur le marché. Il s'agit d'activités pouvant être déduites de propriétés de mouvements simples (vitesse, direction, etc.), telles que, personne courant, rôdant ou tombant. Toutefois, il est souvent difficile de discriminer le niveau de menace que représente ce mouvement. Par exemple, une personne coure-t-elle parce qu'elle s'enfuit ou parce qu'elle ne veut pas rater son vol à l'aéroport? La reconnaissance de comportement demeure un domaine expérimental en analytique. Les attentes face à l'analytique vidéo dépassent souvent largement l'efficacité des technologies actuelles. De plus, celles-ci s'avèrent complexes à intégrer et à utiliser. À ce sujet, Hearing et al. suggèrent une liste de caractéristiques souhaitables pour un système commercial d'analytique vidéo :

- **Minimiser les configurations nécessaires:** Un système commercial ne devrait pas nécessiter beaucoup de temps en configuration, paramétrage et entraînement. Idéalement, l'utilisateur devrait fournir un minimum d'informations supplémentaire au flux vidéo lors de l'installation.
- **offrir une performance adéquate :** Le nombre de fausses alarmes doit être maintenu aussi bas que possible pour éviter de déranger le personnel de sécurité inutilement. Par contre, il est encore plus critique que le système ne laisse pas passer une menace réelle. L'équilibre à viser entre le nombre de fausses alarmes et celui des détections manquées dépend de l'application de sécurité.
- **Ne pas nécessiter de calibrage des caméras ou automatiser celui-ci :** Le calibrage des caméras permet de calculer la taille et la vitesse réelles des objets de la scène. Il établit la correspondance entre la géométrie de la scène et celle de l'image.
- **Supporter une vaste gamme de caméras :** L'analyse vidéo devrait pouvoir se faire indifféremment sur les images provenant de caméras couleur, noir et blanc, infrarouge, thermique ou omnidirectionnelle.
- **Demeurer aussi générique que possible :** Idéalement, la détection d'événements pertinents devrait fonctionner dans des environnements variés et dans différents cas d'application. Toutefois, certains milieux recèlent des problèmes très spécifiques, tels que les zones comprenant des plans d'eau. Le mouvement de l'eau et les réflexions spéculaires perturbent les algorithmes de détection de changements dans l'image sur lesquels se base la détection et le suivi d'objets.
- **Offrir des mécanismes de protection de la vie privée:** Les systèmes d'analytique devraient permettre de masquer les détails sur les séquences vidéo (les visages, par exemple) qui ne sont pas essentiels au traitement d'une alarme de sécurité.

### III.6 Tendances technologiques pour la vidéosurveillance

Encore beaucoup des systèmes actuels de vidéosurveillance reposent sur des technologies traditionnelles : caméras analogiques, câbles coaxiaux, enregistreurs numériques. Toutefois, les progrès techniques dans ce domaine entraînent des changements dans les architectures, les

## Chapitre I : les systèmes de vidéosurveillance

composantes et les capacités des systèmes de vidéosurveillance. Voici les principales tendances technologiques envisagées par les experts du domaine pour les prochaines années :

Un passage à la vidéosurveillance sur réseaux IP s'opère. Surtout adoptée, présentement, pour les grandes infrastructures, la vidéosurveillance IP pourrait gagner la faveur de plus petits utilisateurs si les coûts des caméras IP baissent.

Les caméras Méga pixel constituent une avancée technique majeure pour la vidéosurveillance, offrant une meilleure résolution et une couverture plus large. De plus, elles améliorent le rendement des traitements analytiques. Elles devraient donc se répandre. Un expert questionné pour ce rapport prédisait que, d'ici 2012, la moitié des caméras de surveillance seront de haute résolution. Le standard de compression H.264 (MPEG-4 Part 10/AVC for Advanced Video Coding) devrait s'imposer en vidéosurveillance. Les taux de compression supérieurs qu'il offre par rapport aux standards existants permettent d'économiser bande passante et espace de stockage. Ce standard devrait favoriser l'adoption des caméras Méga pixel en vidéosurveillance.

Les systèmes de surveillance comprendront des réseaux de multiples caméras opérant en temps réel. Dans ces systèmes, la tendance est à distribuer les enregistrements et les traitements dans les différents nœuds (caméras et encodeurs), afin de diminuer la quantité de données transmises. Il devrait y avoir une plus grande intégration et une meilleure interopérabilité des différents systèmes de sécurité pour permettre leur gestion sous un même centre de commande et contrôle. Ainsi, la vidéosurveillance, le contrôle d'accès, les solutions biométriques, RFID, les systèmes de détection d'intrusion pourraient être reliés. La vidéosurveillance intelligente devrait progresser et les produits d'analytique vidéo devenir plus robustes et efficaces pour les applications de sécurité. La détection, le suivi et la reconnaissance d'objets (personnes, visages, véhicules, etc.) dans des scènes chargées, ainsi que la reconnaissance de comportements, seront les domaines de recherche les plus actifs pour la vidéosurveillance. L'utilisation des systèmes de vidéosurveillance dépassera de plus en plus le cadre de la sécurité. Des applications se développeront pour les analyses marketing, l'étude des comportements des consommateurs et la surveillance des opérations d'entreprises.

### III.7 Domaines d'application

La vidéosurveillance a d'abord été utilisée par les services publics (police, transports, administrations). Elle fut ensuite adoptée par des entreprises désirant protéger des actifs stratégiques, telles que les raffineries, les centrales nucléaires, les barrages fluviaux, les usines agroalimentaires et les complexes pharmaceutiques. Les casinos apparaissent aussi comme des précurseurs dans le déploiement de grands systèmes de vidéosurveillance. De nos jours, les caméras de surveillance sont présentes dans divers lieux publics et privés : immeubles, commerces, stationnements, gares, aéroports, routes, transports publics, banques, etc. Les systèmes de vidéosurveillance sont déployés à différentes échelles. Pour la surveillance de crimes mineurs (par ex., attaques, vandalisme, vols), la vidéosurveillance est principalement utilisée pour l'enquête après le fait. Ce niveau de surveillance requiert des technologies simples, le plus souvent analogiques, qui n'intègrent pas d'intelligence vidéo. C'est le type de système qu'on peut retrouver au dépanneur du coin, dans un petit commerce ou dans une résidence privée. La surveillance d'immeubles ou de commerces de plus grande taille nécessite souvent un réseau plus étendu de caméras. La surveillance vise principalement le contrôle des voies d'accès, des stationnements et, s'il s'agit de magasins, des rayons et points de vente. La vidéosurveillance y est surtout utilisée pour les fins d'enquête. Cependant, ces usagers souhaiteraient obtenir des systèmes de surveillance capables de générer des alertes en temps réel pour permettre une intervention immédiate. Les produits d'analytiques vidéo accomplissent donc une percée dans ces secteurs. Le retour sur l'investissement constitue un facteur déterminant pour l'achat d'équipements de vidéosurveillance intelligente pour cette clientèle. La vidéosurveillance à grande échelle se retrouve dans les villes et les quartiers, les systèmes de transport, les campus académiques, les grands événements (festivals, sommets économiques, jeux olympiques, etc.), les périmètres de sécurité étendus. Elle nécessite le déploiement de plusieurs dizaines, voire de centaines de caméras. Celles-ci doivent parfois être accessibles à des centaines d'intervenants en sécurité, provenant de différentes agences gouvernementales, corps policiers ou services d'urgence. Dans ces installations, la vidéosurveillance s'ajoute à une panoplie de systèmes de sécurité et de contrôle : contrôle d'accès, d'incendie, téléphonie, communications radio, systèmes géomatiques, etc. Étant donné le nombre de caméras vidéo impliquées et l'importance des interventions d'urgence, ces applications sont particulièrement propices à l'utilisation d'analytique vidéo pour le traitement automatisé des flux vidéo générant des alarmes lors d'événements suspects. Comme les budgets alloués aux systèmes de surveillance dans ces infrastructures sont souvent importants, l'ajout de logiciels d'analytique peut plus facilement y être envisagé. Il faut noter une utilisation assez récente de

## Chapitre I : les systèmes de vidéosurveillance

la vidéosurveillance dans des unités mobiles : voitures de patrouille, ambulances, autobus, etc. La transmission sans fil des signaux vidéo à un centre de surveillance peut poser problème. Il faut noter une utilisation assez récente de la vidéosurveillance dans des unités mobiles : voitures de patrouille, ambulances, autobus, etc. La transmission sans fil des signaux vidéo à un centre de surveillance peut poser problème.

### III.7.1 Gouvernement et sécurité publique :

Les différents paliers de gouvernement doivent assurer la sécurité de la population, ainsi que des infrastructures publiques. Au niveau national, elle sera utilisée, par exemple, pour surveiller les éléments suivants : Infrastructures sensibles, Frontières, Édifices et sites gouvernementaux, laboratoire, bases militaires et prisons. À un niveau local, la vidéosurveillance est implantée dans plusieurs villes dans le monde pour surveiller la criminalité et servir d'outil pour les interventions d'urgence. Elle contribue aussi à assurer la sécurité lors de grands rassemblements (spectacles, manifestations, événements sportifs). Londres est la ville la plus souvent citée pour le nombre de caméras déployées dans ses rues. La surveillance vidéo sert aussi à gérer le stationnement, notamment pour le contrôle des permis de stationnement, l'application des règles, la détection des vols, vandalismes ou méfaits, et le contrôle d'accès. La vidéosurveillance est abondamment utilisée par les forces de l'ordre pour mener les enquêtes, surveiller les personnes et les véhicules recherchés, détecter les activités dangereuses ou criminelles. On la retrouve aussi à bord des autos-patrouilles pour la vérification des interventions policières. On retrouve de plus en plus de caméras le long des tronçons routiers pour surveiller la circulation et détecter les incidents, les comportements dangereux ou les infractions. Montréal possède un réseau de plusieurs caméras et des stations de détection de véhicules permettant le suivi en temps réel des conditions de la circulation et la détection automatique des incidents<sup>13</sup>. Le secteur gouvernemental présente plusieurs besoins pour la vidéosurveillance intelligente: identification d'individus et de véhicules, comptage de personnes et surveillance de foules, reconnaissance de comportements suspects ou violents (bagarres, méfaits), détection d'intrusions, monitoring routier. Des analyses plus poussées, telles que la reconnaissance d'émotions par ordinateur ou du mensonge chez un individu, sont aussi envisagées.

### III.7.2 Education

La vidéosurveillance est de plus en plus présente dans les établissements scolaires. Son utilisation vise à assurer la sécurité des enseignants et étudiants, ainsi que la protection des actifs contre le vandalisme et le vol. Les tueries tragiques survenues en milieu scolaire, tels qu'à l'école Colombine, aux États-Unis, et ici au Québec, à l'école Polytechnique et au Collège Dawson, ont fait ressortir l'importance de surveiller plus adéquatement les campus scolaires. Ceux-ci, particulièrement dans le cas des universités, peuvent être étendus et posséder plusieurs bâtiments, de nombreux accès à contrôler et des stationnements à surveiller. Dans ce milieu, la vidéosurveillance sert notamment à: surveiller le matériel, les équipements et les données, détecter et suivre les actes de vandalisme, vols, méfaits, comportements inappropriés, reconnaître les plaques d'immatriculation, supporter les enquêtes criminelles et contrôler les accès. Comme les établissements d'éducation possèdent souvent une infrastructure réseau IP, il peut s'avérer avantageux pour eux d'y déployer des systèmes de vidéosurveillance numériques.

### III.7.3 Commerce de détail

Le commerce de détail représente un marché en expansion pour la vidéosurveillance. Celle-ci y est utilisée autant pour des aspects de sécurité intérieure (magasin, entrepôt) qu'extérieure (stationnement). Même les plus petits commerces se dotent de caméras pour, minimalement, conserver des preuves vidéo en cas de vols ou d'incidents. Dans les chaînes de magasins, des systèmes de vidéosurveillance beaucoup plus sophistiqués sont déployés pour permettre la surveillance centralisée des différentes succursales. Dans l'ensemble du secteur, la vidéosurveillance visera notamment:

- la surveillance des caisses et des transactions (vols, fraudes par les employés)
- La protection des biens matériels et des infrastructures.
- La surveillance de l'inventaire, des marchandises (livraisons)

## Chapitre I : les systèmes de vidéosurveillance

- La protection du personnel et des clients
- Le contrôle d'accès aux aires verrouillées
- La vérification des situations d'urgence (incendie, alarmes, etc.)
- La surveillance des stationnements, des véhicules, des entrées et sorties

Considérant les risques élevés de vols et d'attaques auxquels sont exposés les commerces de détail, ainsi que les pertes importantes qui en découlent, la vidéosurveillance s'impose comme un outil incontournable pour assurer la sécurité des employés et de la marchandise. De plus, il existe dans ce marché un potentiel réel pour l'analytique vidéo. Par exemple, on retrouve des systèmes qui combinent l'information vidéo et les données de la caisse, afin de vérifier que les articles sortis par les clients ont effectivement été facturés. Cette technologie permet, par exemple, de prévenir les fraudes aux points de vente impliquant la complicité des caissiers. De plus, la vidéosurveillance intelligente est de plus en plus utilisée à des fins qui dépassent la sécurité, comme la gestion des opérations et la mise en marché. Dans ce contexte, l'analytique vidéo servira, notamment, à compter les clients, à analyser leurs comportements et déplacements en magasin et à compiler des statistiques sur les habitudes de consommation.

### III.7.4 Transport

La sécurité et le bon fonctionnement des aéroports, des gares, des ports et des transports publics est critique pour l'économie d'un pays. Un incident de sécurité peut gravement perturber leurs opérations et se traduire en pertes significatives. Or, étant donné les flots importants de passagers qui y transitent et l'étendue de leurs infrastructures, les systèmes de transports font face à des défis de sécurité hors du commun. Les actes terroristes commis dans différents systèmes de transports dans le monde ont exacerbé ces défis. Les logiciels d'analytique ciblent le secteur des transports en offrant différentes fonctions adaptées : détection d'intrusion dans un périmètre ou une zone contrôlée, détection de personnes pénétrant dans une voie de sortie, détection de bagage abandonné, reconnaissance de visages, comptage de personnes, reconnaissance de plaques d'immatriculation pour le contrôle d'accès aux stationnements, détection de comportements suspects (rôdeurs, vandalisme, graffiti), détection de personnes sur les voies. Toutefois, le milieu des transports représente de grands défis techniques pour les systèmes de vidéosurveillance intelligente, considérant le nombre des personnes qui passent dans le champ de la caméra, la diversité des comportements des passagers et les conditions défavorables pour les caméras (vibrations, poussière, etc.). De plus, pour la surveillance extérieure, les caméras et algorithmes analytiques doivent pouvoir fonctionner en dépit des variations météorologiques (brume, neige, pluie, etc.)

### III.7.5 Aéroports

Après les événements du 11 septembre 2001, les mesures de sécurité ont connu un resserrement, particulièrement dans les grands aéroports, et de nouvelles technologies ont été déployées. Pour les aéroports, la priorité consiste à contrôler l'accès aux zones sécurisées, principalement l'accès aux avions, mais aussi à assurer la sécurité des passagers, du personnel et des biens sur l'ensemble du périmètre (pistes, stationnements, voies d'accès, etc.). Selon les recherches faites pour ce rapport, la vidéosurveillance intelligente semble encore peu adoptée dans les aéroports canadiens. Par exemple, à l'aéroport Pierre-Elliott Trudeau de Montréal, la vidéosurveillance est surtout utilisée pour assister les interventions d'urgence. Lorsqu'un incident est signalé, une caméra mobile pivote pour suivre les opérations en direct, à partir d'un centre de surveillance. Le contrôle d'accès repose principalement sur des technologies biométriques d'identification (empreintes, iris), l'utilisation de cartes à puces et de NIP15. La vidéo de surveillance est archivée et sert également comme élément de preuve pour les enquêtes des corps policiers. Elle se retrouve aussi à bord des véhicules de patrouille. Pour le moment, les caméras sont majoritairement analogiques et peu d'analytique vidéo est utilisée. Parmi les fonctions intelligentes déjà en place, on retrouve le comptage de personnes pour évaluer, par exemple, les temps de traitement à la douane, ainsi que la reconnaissance de plaques d'immatriculation qui permet de recenser les voitures dans les stationnements et de détecter des vols. Aéroports de Montréal devrait prochainement ajouter une nouvelle partie à son système de vidéosurveillance qui sera entièrement IP et devrait inclure caméras intelligentes et logiciels de

## Chapitre I : les systèmes de vidéosurveillance

traitement vidéo efficace, la STM envisage certaines autres applications, telles que la détection de colis suspects, la surveillance des portes et la reconnaissance de comportements suspects.

### III.7.6 Milieu bancaire

La vidéosurveillance est très utilisée pour assurer la sécurité des banques. Dans un premier temps, la présence de caméras opère un certain pouvoir de dissuasion sur la perpétration de vols à mains armées et d'agressions. Si un tel délit survient, les séquences vidéo archivées servent à l'enquête et l'identification des malfaiteurs. Les guichets automatiques sont particulièrement à risque pour les actes criminels. Les caméras de surveillance aident à détecter les fraudes, comme par exemple, l'installation d'un dispositif pour lire l'information magnétique des cartes de guichet. Pour le milieu bancaire, la vidéosurveillance intelligente peut augmenter l'efficacité de la surveillance. Elle permet d'assurer un suivi sur l'ensemble du réseau des succursales, afin d'y détecter les individus ou comportements suspects. Elle permettra, entre autres, de retrouver toutes les séquences vidéo provenant de toutes les succursales où apparaît un certain individu, grâce à des techniques de reconnaissance faciales.

### III.7.7 Autres

La vidéosurveillance est déployée dans beaucoup d'autres milieux. On la retrouve dans les commerces et pour la surveillance d'édifices. Elle sert, le plus souvent, à filmer les voies d'accès et les stationnements, à surveiller les valeurs matérielles et à assurer la protection des employés et des clients. La surveillance vidéo trouve aussi sa place en santé pour assister les interventions. On peut même la retrouver à bord des ambulances, en combinaison avec d'autres instruments de mesure, pour permettre le suivi du patient à distance. L'entreprise américaine Indigo Vision a développé des solutions d'analytiques embarquées dans des caméras intelligentes spécifiquement pour détecter le mouvement des hélicoptères qui atterrissent sur le site d'un hôpital américain. Le marché grand public se développe aussi. De plus en plus de systèmes de vidéosurveillance pour les résidences sont en vente, même dans les magasins non spécialisés. Les caméras complètent maintenant les différents dispositifs de surveillance et de sécurité pour le secteur résidentiel.

## III.8 Indexation et recherche de vidéo surveillance

En se basant sur l'hypothèse que les vidéos sont prétraitées par un module d'analyse vidéo, l'approche générale pour l'indexation et la recherche de vidéos pour la vidéosurveillance comprend deux phases : la phase d'indexation et celle de recherche. Afin d'utiliser les résultats de différents modules d'analyse vidéo, un modèle de données comprenant deux concepts, objets et événements, est proposé. La phase d'indexation visant à préparer des données déterminées dans ce modèle de données effectue trois tâches. Premièrement, deux nouvelles méthodes de détection des blobs représentatifs de la tâche représentation d'objets déterminent un ensemble de blobs associés à leurs poids pour chaque objet. Deuxièmement, la tâche extraction de descripteurs consiste à analyser des descripteurs d'apparence et aussi temporels sur les objets indexés. Finalement, la tâche indexation calcule les attributs des deux concepts et les stocke dans une base de données. La phase de recherche commence avec une requête de l'utilisateur et comprend quatre tâches. Dans la tâche formulation de requêtes, afin de permettre à l'utilisateur d'exprimer ses requêtes, un nouveau langage est proposé. La requête est traitée par la tâche analyse syntaxique. Une nouvelle méthode dans la tâche mise en correspondance permet de retrouver efficacement les résultats pertinents. Deux méthodes dans la tâche retour de pertinence permettent d'interagir avec l'utilisateur afin d'améliorer les résultats de recherché. Les approches d'analyse vidéo pour la vidéosurveillance se différencient par les algorithmes proposés pour chacune des tâches. Nous classons les approches en deux catégories selon la présence du retour de pertinence.

# Chapitre I : les systèmes de vidéosurveillance

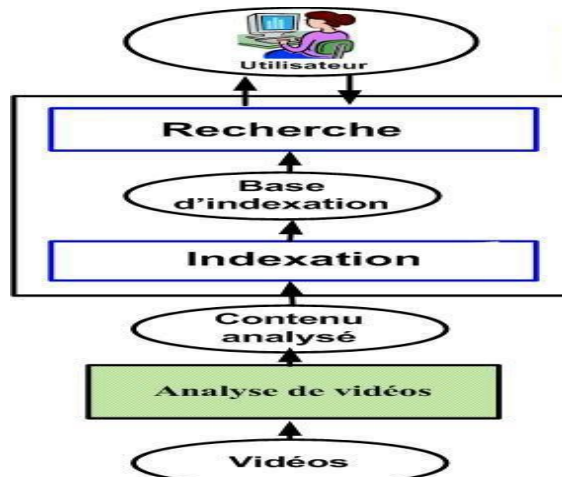


Figure I.5: les étapes d'indexation

## III.8.1 Approche sans retour de pertinence

Les résultats des modules d'analyse peuvent être les objets et les événements. Les approches d'indexation et de recherche de vidéos pour la vidéosurveillance basées sur la reconnaissance peuvent être effectuées au niveau objets, au niveau événements et au niveau composé.

- **Approche au niveau objet** L'objectif est de retrouver les objets observés par une ou un ensemble de caméras qui sont semblables à une image ou un objet recherché. Nous divisons les approches proposées en trois catégories : celles basées sur la fusion précoce, celles sur la fusion tardive et les approches hybrides. Nous présentons le cas général où une scène est observée par plusieurs caméras. Dans le cas où il y a une seule caméra, la recherche des objets est celle de la première catégorie.
- **Les approches basées sur la fusion précoce** : Les approches de la première catégorie fusionnent des vidéos provenant des caméras dans la tâche de détection et de suivi des objets. La figure I.6 montre le processus commun des approches dans cette catégorie. Le travail de Conaire et al. fait partie de cette catégorie. Les auteurs ont combiné l'information d'une caméra infrarouge et d'une caméra CCTV (Closed Circuit Television). Ils ont rapporté que la combinaison de deux caméras permet non seulement d'observer des objets dans des conditions différentes d'illumination mais aussi d'obtenir de bons résultats de détection et de suivi d'objets. Afin de mettre en correspondance entre les objets, les descripteurs tels que la taille de l'objet sont calculés.

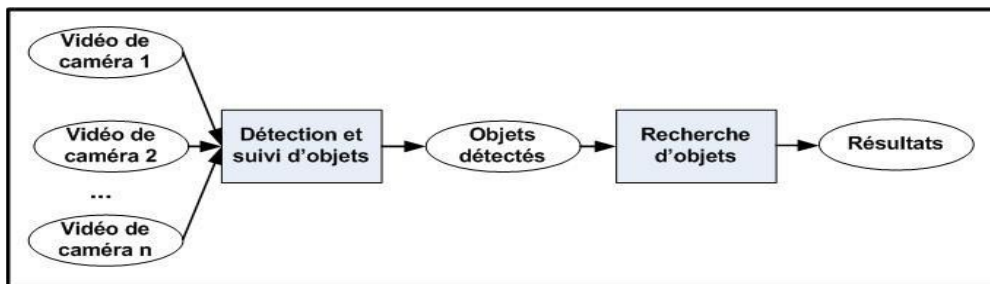
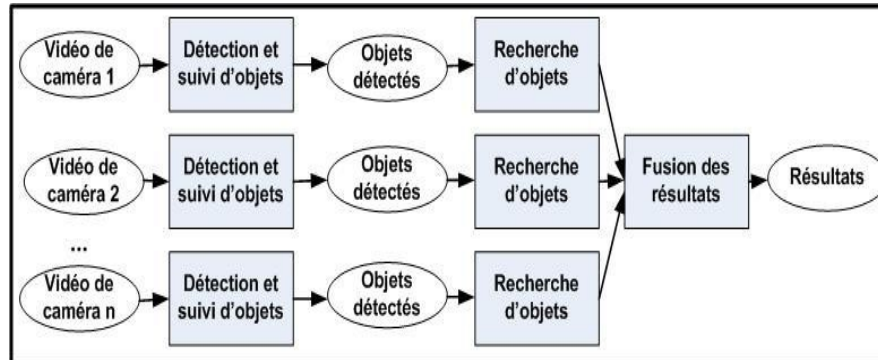


Figure I.6 : les approches basées sur la fusion précoce

- **Les approches basées sur la fusion tardive** : Contrairement aux approches de fusion précoce, celles de fusion tardive font la détection et le suivi d'objets séparément pour chacune des caméras. L'appariement des objets détectés et de la requête se fait tout d'abord indépendamment pour chacune de caméras. Une fusion est alors effectuée sur les résultats

## Chapitre I : les systèmes de vidéosurveillance

obtenus pour avoir les résultats finaux. La figure illustre le processus commun des approches dans cette catégorie.



**Figure I.7 : les approches basées sur la fusion tardive.**

- **Les approches hybrides** : Dans les approches hybrides, la recherche des objets se fait sur les régions détectées avant la fusion et aussi après la fusion. Une région appelée PA (person's appearance) est détectée pour un objet dans chacun des frames. Ensuite un ensemble de régions détectées appelé SCAT (single camera appearance trace) d'un objet dans un intervalle de temps est créé. Puis un MCAT (multicamera appearance trace) est construit pour chacun des objets en appliquant l'algorithme qui permet de relier les SCATs du même objet observé par différentes caméras. Cet algorithme est appelé consistent labeling. Le processus de l'appariement des objets est constitué de deux étapes :
  - Étape 1 : appelée Best PA selection : un PA d'un objet est choisi comme une requête, dans MCAT de cet objet, un SCAT dont la taille d'objet est la plus grande est identifié. Le PA dans le SCAT identifié dont la variation de couleur est la plus grande est déterminé comme une requête intermédiaire.
  - Étape 2 : appelée Similarity-based retrieval : la requête intermédiaire est comparée avec les MCATs indexés en se basant sur la densité de probabilité de couleur. Nous détaillons la deuxième étape. Pour un MCAT dans la base d'indexation : pour le premier PA de MCAT, un histogramme de couleur R, G, B est calculé, dix pixels  $X_t$  dont leurs couleurs correspondent à dix valeurs maximales de l'histogramme sont choisis. Dix gaussiennes sont initialisées par les dix pixels, leurs poids sont également initiés par la même valeur. Pour un PA suivant de MCAT, dix pixels dont leurs couleurs correspondent à dix valeurs maximales de l'histogramme sont déterminés. Pour chacun des pixels  $X_t$ , l'approche vérifie: si toutes les dix gaussiennes ayant la différence de leurs moyennes et les pixels  $X_t$  est supérieur à  $2.5 * \sigma$ , une nouvelle gaussienne va être créée. sinon l'approche met à jour les moyennes et les écarts types des gaussiennes

## Chapitre I : les systèmes de vidéosurveillance

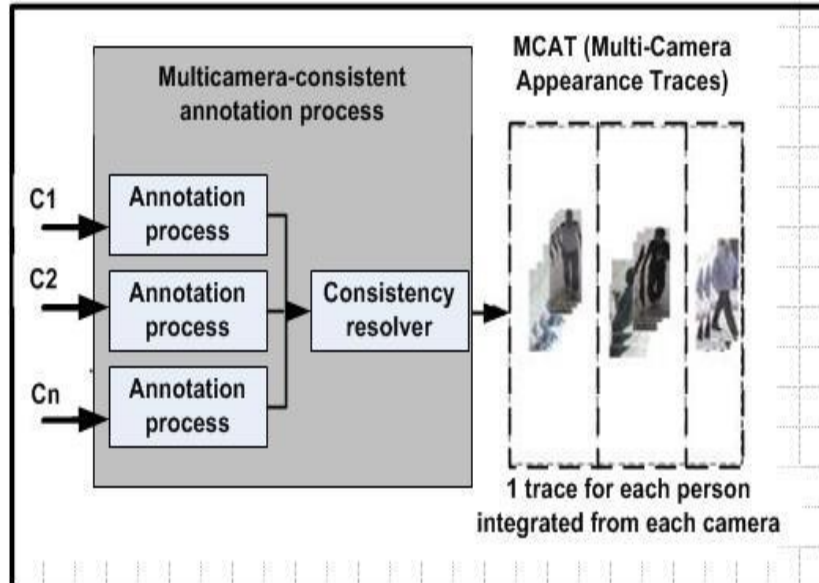


Figure I.8 : Processus de génération PA, SCAT et MCAT .

- **L'approche au niveau événements** : Dans le cadre de l'indexation et de la recherche de vidéos pour la vidéosurveillance au niveau événements. Au lieu de reconnaître tous les événements d'intérêt de l'utilisateur ce qui est dans la plupart des cas impossible, l'approche de Ghannem

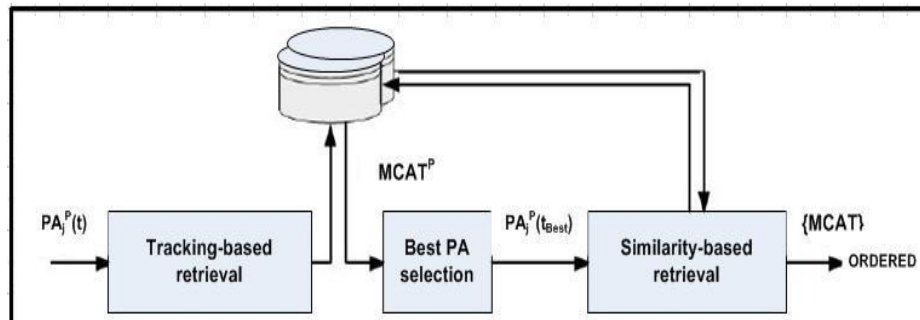


Figure I.9 :Appariement des objets.

Appariement des objets est constitué de deux étapes : la première étape (Best PA selection) permettant de déterminer le PA d'un objet à travers de PAs créés à partir des vidéos provenant de toutes les caméras, la deuxième étape consistant de comparer le PA choisi avec les MCATs de tous les objets détectés, permet aux utilisateurs de définir eux-mêmes des événements d'intérêt à partir d'événements primitifs reconnus à l'aide des relations logiques et temporelles. Cette approche est illustrée dans la figure I.9. Un module de vision consiste à suivre les objets et à détecter les événements primitifs.

- **Approche au niveau composé** : Les approches combinant les objets et les événements font l'objet d'un travail considérable. Une des approches présentée par Hu et al. est dédiée aux enregistrements de routes. Les auteurs ont défini une activité et un modèle d'activité. Un modèle d'activité représente un ensemble d'activités similaires. Les descripteurs de l'activité

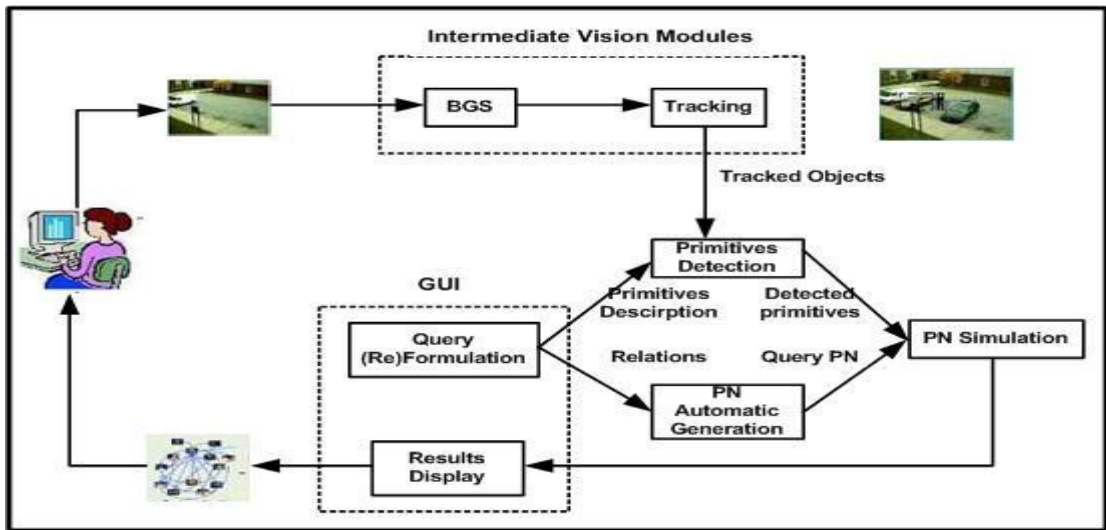


Figure I.10: module vision (Intermediate Vision) pour la détection et suivi des objets.

## IV. Conclusion

Dans ce chapitre, nous avons définis la vidéosurveillance ainsi que ses domaines d'utilisation, ses technologies et son utilisation pour l'indexation et la recherche.

Les différents concepts traités dans ce chapitre nous aiderons à mieux comprendre notre mode d'opération et les notions fondamentales pour mener à bien notre projet.

# chapitre II

*Traitement d'image*

## Chapitre II : traitement d'image

### I. Introduction

Le traitement d'images est une discipline de l'informatique et des mathématiques appliquées qui étudie les images numériques et leurs transformations, dans le but d'améliorer leur qualité ou d'en extraire de l'information. Ce dernier est souvent synonyme d'amélioration des images avec pour but l'obtention d'une plus grande lisibilité. Il n'y a pas création d'informations, mais mise en évidence de l'information pertinente déjà présente.

Dans ce chapitre nous allons parler de quelques concepts de l'image et de traitement d'image puis nous allons introduire la notion de « caractéristique locale et globale » ainsi que la détection et l'extraction de ces dernières, quelques descripteurs (locaux et globaux) seront aussi présentés d'une manière expéditive.

### II. Quelques notions de terminologie

#### II.1 Définition de l'image

Une image numérique est une image (dessin, icône, photographie, ...) créée, traitée, stockée sous forme binaire, est défini par le nombre de pixels ayant chacun comme caractéristique un niveau de gris ou de couleur prélevé à l'emplacement correspondant dans l'image réelle, ou calculé à partir d'une description interne de scène à représenter. La numérisation d'une image est la conversion de celle-ci de son état analogique (distribution continue d'intensité lumineuses dans un plan  $x^*y$ ) en une image numérique représentée par une matrice bidimensionnelle de valeur numérique  $i(x,y)$  ou  $x, y$  sont des coordonnées cartésiennes d'un point de l'image et  $i(x,y)$  : niveau de gris en ce point .

#### II.2 Les attributs de l'image

##### II.1.1 Pixel

Le pixel abréviation de deux mots en l'anglais : **P**icture **e**lement est l'élément de base d'une image ou d'un écran, c'est-à-dire un point, l'ensemble de ces pixels est contenu dans un tableau à deux dimensions constituant l'image, qui la composent en largeur et en hauteur et aussi de l'étendue des teintes de gris ou des couleurs que peut prendre chaque pixel .

1	1	1	1	1	1	1	1	1	1
1	0	0	0	1	1	0	0	0	1
1	1	0	1	1	1	1	0	1	1
1	1	0	1	1	1	1	0	1	1
1	1	0	1	1	1	1	0	1	1
1	1	0	0	0	0	0	0	1	1
1	1	0	1	1	1	1	0	1	1
1	1	0	1	1	1	1	0	1	1
1	1	0	1	1	1	1	0	1	1
1	0	0	0	1	1	0	0	0	1
1	1	1	1	1	1	1	1	1	1

Figure II.1 : Représentation d'une lettre sous la forme du groupe de pixels.

##### II.1.2 Dimension

La dimension est le nombre de points (pixel) constituant l'image (dimension informatique) . Cette dernière se présente sous forme de matrice dont les éléments sont des valeurs numériques représentative de l'intensité lumineuse (pixels). Le nombre de ligne de cette matrice multiplié par nombre de colonne nous donne le nombre total de pixels dans une image .

## Chapitre II : traitement d'image

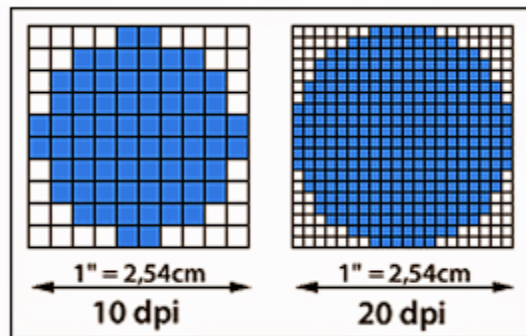


Figure II.2 : Représentation de dimension d'une image.

### II .1.3 Le bruit

C'est un signal qui lors de l'acquisition ou la transmission vient s'ajouter à l'image, il se matérialise de la présence des régions homogènes des valeurs ou moins éloignées l'intensité de la région. Le bruit est le résultat de certains défauts électroniques de la capture et de la qualité de numérisation.



Figure II.3 : image représente les différents types de bruit.

### II .1.4 Résolution

La résolution se caractérise par le nombre de pixels par unité de longueur, c'est en fait la densité de pixels de l'image. Celle-ci n'a rien à voir avec le nombre de pixels. La résolution s'exprime en dpi (dot per inch = point d'encre par pouce) pour une imprimante, ou en ppp (pixel par pouce) pour un fichier image. Évidemment plus la résolution de l'image est élevée (beaucoup de pixels pour un pouce), plus la qualité théorique est importante. Nous parlons de qualité théorique car la notion de qualité est plus fonction de l'œil humain que de chiffres étalés en vrac.

### II .1.5 Histogramme

Est une fonction qui donne la fréquence d'apparition de chaque niveau de gris (couleur) pour diminuer l'erreur de quantification, pour comparer deux images obtenues sous des éclairages différents, ou encore pour mesurer certaines propriétés sur une image. On modifie souvent l'histogramme correspondant. Il permet de donner un grand nombre d'informations sur la distribution

## Chapitre II : traitement d'image

des niveaux de gris (couleur) et de voir entre quelles bornes est répartie la majorité des niveaux de gris (couleur) dans les cas d'une image trip claire ou d'une image trip foncée.

### II .1.6 Conteurs et textures

Les conteurs représente la limite entre deux pixels dont les niveaux des gris représentent une différence significative. La texture dans une image nous est présentée comme un arrangement spatial de couleurs ou d'intensités dans une région de cette image .

### II .1.7 Luminance et contraste

La luminance est tout simplement la sensation visuelle qu'on a de la lumière .De toutes les grandeurs photométrique qui caractérisent la lumière , la luminance est celle qui nous concerne le plus ,car notre organe visuel ne sait (compte )qu'en luminance . Lorsque la vision est gênée dans une forte pénombre , ou lorsqu'un phare nous éblouit, c'est la luminance qui en cause .

## II.3 Les types d'images

### II.3.1 L'image en niveaux de gris

Le niveau de gris, désigne la concentration des points de trame et est donc directement en rapport avec le rendu de l'image. Un niveau de gris va alors varier du blanc au noir . Cela est également valable pour toutes les autres couleurs (cyan, magenta, jaune). On fera varier la densité du cyan du blanc jusqu'au cyan foncé.

Dans une image numérique, le niveau de gris représente la luminosité d'un pixel, lorsque des valeurs de ses composantes de couleur sont identiques. La plupart des formats de fichier image offrent un mode de reproduction en niveaux de gris, qui divise par trois le nombre d'octets nécessaire au codage.



Figure II.4 :image en niveau de gris

### II.3.2 L'image en couleur

Même s'il est parfois utile de pouvoir représenté des images en noire et blanc ou en niveau de gris ,les applications multimédias utilisent le plus souvent des images en couleurs .La représentation des couleurs s'effectue de la même manière que les image monochromes avec cependant quelques particularités .En effet ,il faut tout d'abord choisir un modèle de représentation .Pour cela on utilise un espace de couleur a plusieurs dimensions qui consiste a donner suffisamment de composant numérique pour décrire une couleur .



Figure II.5 : Image en couleur

### II.3.3 Le mode monochrome

Dans ce mode le pixels y est soit allumé (blanc) soit éteint (noire) ,et l'image obtenue n'est pas très nuancée c'est pour cella il faut passer par le mode niveaux de gris pour convertir une image en couleur ou en mode monochrome.

## Chapitre II : traitement d'image



Figure II.6 :image en mode monochrome

### II.4 Les format d'image

#### II.4.1 Les image bitmap

Ce sont les image affichées sur l'écran d'un pc , chaque image est en réalité une matrice de pixels .on distingue plusieurs format bitmap :



Figure II.7 :image matricielle

- **Les format gif ( Graphic Interchange Format )**

Le Gif est le plus vieux format de compression image pour le web (1989). Il est essentiellement utilisé pour les graphiques de faible taille, comprenant des couleurs sans trop de dégradés, pour des dessins au trait ou encore pour des petits dessins animés (GIF animé). Il traite traditionnellement des images indexées en 8 bits, donc à 256 couleurs. La compression utilise un procédé non destructif de données: l'image décompressée est identique à l'image d'origine. Le format GIF 89a, basé sur le format d'origine GIF 87a, peut gérer des zones de transparence dans l'image. Ainsi, le fond peut disparaître: une couleur transparente est une sorte de couleur invisible, qui est remplacée par le fond de la page Web. Le Gif est donc adapté aux logos, à tout ce qui a de grandes zones de couleur homogène. On peut aussi avoir des zones transparentes, et on peut créer une animation minimale avec du GIF appelé « gif animé ». Ce qui est assuré au gif sa longue vie même s'il est un peu archaïque aujourd'hui.

- **Format JPEG (Joint Photographics Expert Group)**

Le format JPEG (JPG) est le plus adapté à la compression d'images photographiques possédant de nombreux dégradés de couleurs. Il n'est pas recommandé pour les dessins au trait ou les images de petite taille. Il est capable de supprimer la plupart des nuances de couleurs que l'œil ne distingue pas. Ce format est un mode de compression à perte. Il est donc recommandé de travailler l'image sous un autre format et de l'enregistrer une unique fois en JPEG pour éviter la perte inutile de données. Il est intéressant d'essayer plusieurs niveaux de qualité sur une même image pour comparer les résultats obtenus. Il existe aussi trois options de compression standard, optimisé et progressif.

- **Format PNG (Portable Network Graphic)**

Le PNG est un format qui a été créé comme réponse à la protection du format GIF. Il étend la possibilité de celui-ci, notamment en permettant une transparence « alpha », c'est à dire partielle. Son utilisation a été freinée sur le web par le refus de Microsoft de l'implémenter dans Internet Explorer . On estime qu'un fichier en PNG est 10 % à 30 % plus léger qu'un fichier GIF. Ce format s'impose comme format universel pour le Web.

- **Format WEBP (Web Picture)**

## Chapitre II : traitement d'image

Dans toutes les situations le WebP se révèle meilleur avec jusqu'à 60% de gain. Pour le moment la transparence alpha n'est pas gérée mais sera bientôt ajoutée pour en profiter comme dans le cas du PNG 24 bits. Tout comme le JPEG, le WebP compresse les images avec une perte de données et donc de qualité, même si Google assure que dans la majorité des cas le taux de compression obtenu est supérieur au celui du JPEG à qualité équivalente.

- **Le format PCX (Picture Xchange Image Bitmap Zoft)**

A été créé à l'origine par ZSOFT pour un programme de dessin (paint bruch).

### II.4.2 Les images vectorielles

Etant donné qu'une image vectorielle est constituée uniquement d'entités mathématiques, il est possible de lui appliquer facilement des transformations géométriques (zoom, étirement, ...). De plus, les images vectorielles permettent de définir une image avec très peu d'information, ce qui rend les fichiers très peu volumineux. En contrepartie, une image vectorielle permet uniquement de représenter des formes simples. Les formats d'image vectorielle sont :



**Figure II.8 : image vectorielle**

- **Le format EPS ( Postscript / Encapsulated Postscript)**

Le postscript est un langage de description de page conçu pour imprimer des documents indépendamment du périphérique utilisé. Il contient toutes les commandes nécessaires pour dessiner l'image sauvegardée.

- **Le format CGM ( Computer Graphics Metafile )**

Les fichiers graphiques CGM restent un moyen privilégié d'échange de dessins vectoriels entre applications, mais il ne traite pas les images bitmap.

## III. Le traitement d'image

### III.1 Introduction à la vision par ordinateur

Le domaine de la vision par ordinateur a longtemps consisté à interpréter les objets dans une image d'une scène. Il s'agissait d'un processus de traitement de l'information dont l'entrée est constituée d'une ou plusieurs images. Le système apporte un certain nombre de connaissances, des connaissances physiques tels que la gravitation impliquant qu'un objet doit être posé sur un support horizontal, des connaissances géométriques définissant l'objet en terme de ligne ou de surface et des connaissances de haut niveau décrivant la fonction de l'objet dans la scène ou bien son existence. La présence d'une chaise dans une scène d'intérieur est plus probable que celle d'un arbre. Poussée par le progrès scientifique et technologique, la recherche en vision par ordinateur s'est orientée vers la compréhension de la scène comportant tout type d'objets et, en particulier vers l'analyse de scènes comportant des humains. La disposition de caméras et de cartes d'acquisition vidéo dans tous les domaines a poussé son introduction dans le domaine de l'interaction homme-machine.

### III.2 Acquisition d'une image à partir d'une séquence vidéo

## Chapitre II : traitement d'image

Une séquence vidéo est constituée d'une séquence d'images fortement redondantes, la recherche d'image dans une séquence vidéo comporte des étapes incontournables sont :

- **Découpage en image** : Il s'agit dans cette étape de restituer les images individuelles qui ont formé la vidéo grâce aux algorithmes et aux techniques de programmation ainsi que des logiciels de traitement de vidéo et de capture d'écran existant comme Virtual dub, OSS vidéo de compiler... ; ces images nous permettront par la suite de définir des plans d'images en utilisant des distances de similarité .
- **Segmentation spatiale** : La segmentation d'images constitue un élément clé entre traitement et analyse de l'image, elle permet de passer du support de l'image à une liste d'objets décrits en termes de régions. Dans cette étape il s'agit de segmenter chaque image individuellement (qui est le résultat de l'étape précédente) en utilisant un des algorithmes de segmentation existant.
- **Segmentation temporelle** : La segmentation temporelle est une étape très importante pour les approches d'indexation et de recherche de vidéos par le contenu. La majorité des approches utilisent des critères de bas niveau comme la couleur, la texture ou le mouvement, pour segmenter la vidéo en plusieurs unités de base appelées "*plans*".
- **Sélection de l'image clef et recherche dans la vidéo** : Après avoir déterminé les plans dans l'étape précédente, nous devons extraire dans cette étape les caractéristiques visuelles de chaque plan. Ces caractéristiques sont définies dans une ou plusieurs images appelées "*images clefs*" (appelées parfois images représentatives, ou images caractéristiques). Dans les bases d'images, les recherches consistent à savoir si l'image requête (l'image introduite par l'utilisateur dans la phase on-line) est stockée dans cette base d'images ou bien à extraire de cette base toutes les images similaires à l'image requête au sens d'une distance de similarité.

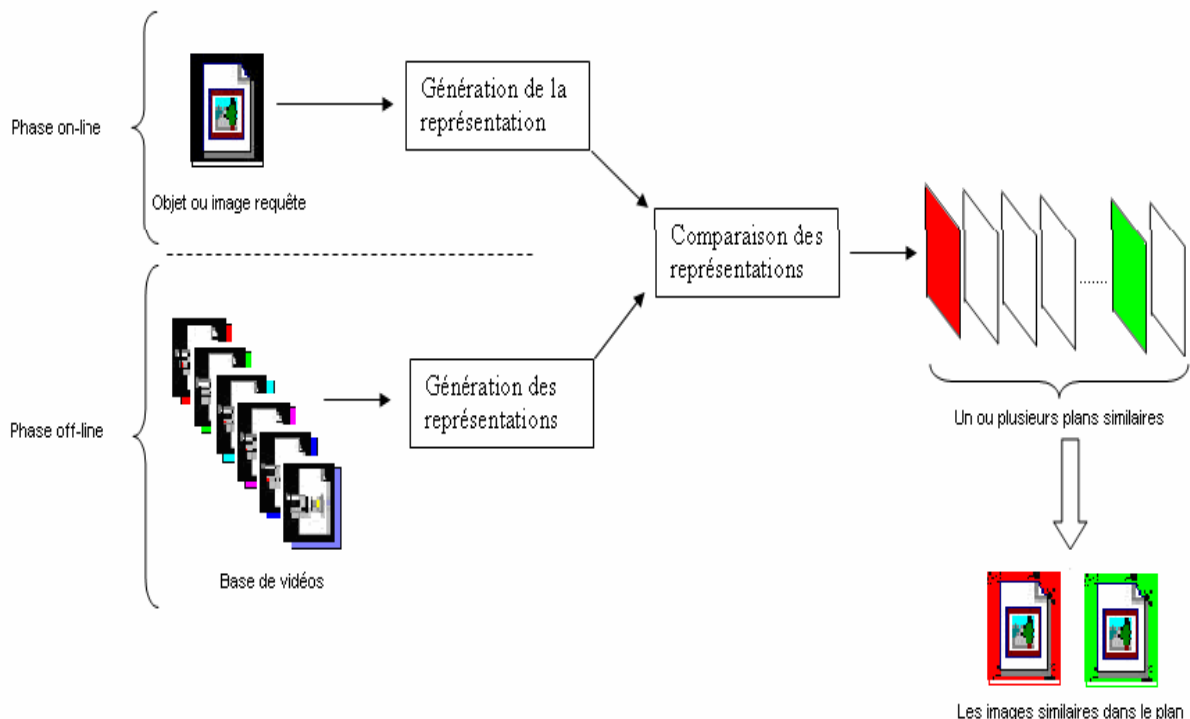


Figure II.9 : Représente l'acquisition d'image à partir d'une séquence vidéo  
III.3 Avantage et inconvénient de traitement d'image

## Chapitre II : traitement d'image

- **Rapidité** : Le traitement d'images accélère beaucoup les travaux traditionnellement réalisés par photo-interprétation, mais qui sont automatisables et donc reproductibles (e.g., vérification de la bonne disposition d'étiquettes sur les bouteilles produites par une usine d'eau minérale). Dans des cas défavorables, l'interprétation visuelle d'image peut être plus rapide et précise que l'interprétation numérique. En effet, l'opérateur humain peut s'adapter à des conditions hors des limites de bon fonctionnement du système d'analyse numérique (e.g., image déformée du fait d'un éclairage défectueux non modélisable). Ainsi, un opérateur peut plus ou moins faire abstraction de la présence de "bruit" dans une partie de l'image lors de l'analyse de celle-ci. De manière intuitive, il arrive à séparer le bruit et l'information recherchée dans l'image. Ceci est le cas d'une image qui contient une route en partie au soleil et à l'ombre. Un photo interprète peut immédiatement réaliser qu'il s'agit d'un même objet alors qu'une interprétation purement numérique réalisée par un logiciel non adapté peut mener à la conclusion qu'il s'agit de deux objets différents. L'extraction numérique d'information dans une image "bruitée" nécessite des algorithmes. Complexes capables d'utiliser des informations annexes qui renseignent sur les caractéristiques du bruit présent.
- **Travaux répétés et fastidieux** : Ces travaux (correction géométrique, étalonnage, gestion de base de données,...) sont souvent difficiles à réaliser, sinon impossibles, de manière non numérique. Ceci explique pourquoi les images analogiques sont souvent numérisées pour être "interprétées".
- **Reproductibilité de l'analyse et de l'interprétation** : L'interprétation humaine (i.e., photo-interprétation) est souvent fastidieuse, longue, coûteuse, difficilement reproductible et très dépendante de l'opérateur. Ainsi, l'interprétation d'une même image par plusieurs photo interprètes donne souvent des résultats très différents, car les êtres humains n'ont pas la même sensibilité aux couleurs, au contexte des objets analysés, etc. Par exemple, un photo interprète peut confondre 2 niveaux de gris alors qu'un ordinateur ne peut confondre 2 valeurs numériques
- **Quantification** : intérêt majeur du numérique est de donner une information quantifiée : Dimensions (aire, périmètre, etc.) de tout ou partie d'un matériau (i.e., nombre de pixels), à partir d'une délimitation numérique ou manuelle. Concentrations chimiques par le biais des caractéristiques spectrales d'objets de l'image. Et grandeurs physiques (rugosité, etc.) dérivées de propriétés optiques comme la réflectance.

### III.4 Exemples d'emploi du traitement d'image

- **Le contrôle des fibres** : Le prix de vente de la laine est très dépendant de sa qualité. En général, la qualité d'une fibre est d'autant meilleure qu'elle est fine. Elle dépend aussi de sa longueur et de sa résistance. Pour beaucoup de raisons (coût, etc.), il est important d'évaluer la qualité des fibres de laine, à la fois pour les produits fabriqués localement et les produits achetés. . Ce contrôle intéresse donc d'une part le producteur de matière (laine, poils de lapin, etc.) qui peut ainsi maîtriser sa production, et d'autre part l'industriel qui s'assure de la qualité de ses fournitures. Le traitement d'image fournit un moyen de contrôle sûr, reproductible et s'affranchissant le plus possible de la subjectivité d'un opérateur.

Déroulement d'un contrôle non automatique. Un échantillon représentatif du lot de fibres à contrôler est tout d'abord prélevé, puis placé sous un microscope afin d'évaluer le diamètre de chaque fibre. Le nombre d'échantillons à observer pour qualifier un jeu de fibres est fonction de la dispersion du diamètre des fibres. Plus elle est faible, plus le nombre d'échantillons peut

## Chapitre II : traitement d'image

être petit. Avant l'introduction de techniques d'automatisation le contrôle était manuel : le manipulateur estimait "à l'œil" la valeur du diamètre de chaque fibre. Cette mesure était donc peu précise. De plus, la nécessité d'analyser un grand nombre de fibres afin d'obtenir des résultats statistiquement fiables rendait cette opération longue et coûteuse. Développement d'un système automatisé : L'Institut Textile de France (Mazamet) évalue la qualité des fibres avec un système de vision industrielle qui comprend un microscope, une caméra, un NS 150001 (micro-ordinateur et logiciel), et une imprimante. Il permet de mesurer le diamètre des fibres. En présence d'un mélange de fibres, le système compte les diverses sortes de fibres par classe. Le contrôle comprend plusieurs étapes :

- Prélèvement d'un échantillon dans le lot à contrôler, puis l'opérateur prépare une lame avec ces fibres et la place sous le microscope .
  - Visualisation de l'image sur le moniteur de l'analyseur après numérisation en temps réel par une caméra (512 x 512 pixels avec 256 niveaux de gris). Le logiciel offre des possibilités d'amélioration de contraste et de compensation des défauts de l'image tels qu'un éclairage non uniforme.
  - Etalonnage éventuel du système par l'opérateur.
  - Binarisation de l'image afin de déterminer les contours de la fibre à l'aide d'opérateurs morphologiques complexes, réalisés avec des mailles hexagonales ou dodécagonales. Le système trie les objets selon leur taille, identifie les objets incomplets, sépare les particules agglomérées, et connecte des lignes interrompues. Les résultats (surface, périmètre, longueur, largeur, coordonnées, etc.) sont rangés dans un tableau qui peut contenir jusqu'à 8000 objets .
  - Marquage de chaque côté des fibres. . Le logiciel trace alors une ligne, détermine les intersections avec les bords de chaque fibre, ce qui permet de calculer le diamètre de chaque fibre intersectée. L'analyse d'un lot de 1200 échantillons, avec plusieurs types de fibres, nécessite environ 2 heures.
- **La mesure de pollution des fluides** : Les performances des systèmes hydrauliques doivent sans cesse s'améliorer du fait de l'évolution des techniques aérospatiales et de la recherche d'un maximum de sécurité, d'efficacité et de fiabilité . Ces 3 propriétés sont très liées à la propreté du fluide hydraulique. La nécessité de maintenir une très faible pollution (norme NF L 41-101) implique un contrôle régulier et fiable. . L'origine de la pollution peut être diverse (air, eau, micro-organismes, particules solides, réaction chimique, etc.). Contrôle non automatique de la pollution d'un fluide.
    - Filtrage du fluide (ex. échantillon de 100cm<sup>3</sup>) à analyser avec une membrane Millipore. Celle-ci est ensuite placée sous un microscope .
    - Comptage par l'opérateur sur un nombre fixe d'échantillons du nombre de particules sur le filtre. Ce comptage manuel est long, fastidieux et parfois incertain. Contrôle automatique de la pollution d'un fluide.

Le système de comptage automatique de contrôle de fluide hydraulique décrit ici est appliqué par plus d'une douzaine de bases aériennes en France. Il permet un contrôle plus systématique et plus rigoureux. En effet, l'analyse est effectuée de manière répétitive et sur une plus grande surface du filtre que lors du comptage manuel. Il comprend 3 parties principales : un microscope électronique et le système informatique NS 15000 (micro-ordinateur + logiciel) qui traite les images. Il identifie et classe selon leurs dimensions les particules emprisonnées dans le filtre Millipore. Le microscope utilisé permet l'observation en réflexion et transmission avec plusieurs grossissements et filtres. . La précision des déplacements est de 0.25 μm. Les déplacements, l'éclairage et l'autofocus sont contrôlés par

## Chapitre II : traitement d'image

l'intermédiaire d'une carte placée dans le système informatique, et pilotée par celui-ci. Cette approche assure l'automatisation de l'opération d'analyse et la reproductibilité des conditions d'observation d'un filtre à l'autre. Plus de 1800 échantillons peuvent être analysés en environ une heure. Les particules sont isolées, classées selon leur longueur.

- **Système de contrôle de laminé** : Ce système de vision industrielle de Delta Technologie détecte et repère recherche en temps réel, la présence de perforations sur du laminé défilant à grande vitesse .

Le système de contrôle comprend deux ensembles d'acquisition identiques implantés en deux endroits du site de production .Une armoire industrielle déportée centralise les signaux, traite et génère si besoin des alarmes (sonore et visuelle) à destination de chaque site. Le paramétrage des sites s'effectue au démarrage et/ou en cours de production.

- **Système de mesure de capacité d'équipement** : Destiné aux industriels de l'électronique, ce système de vision industrielle de *Delta Technologies* mesure le niveau de réglage des équipements de report composant ( $\sim 5^{-1}m$ ), de dépose de points de colle ( $\sim 15^{-1}m$ ) ou les mires de sérigraphie ( $\sim 2^{-1}m$ ). Avant chaque procédure de mesure, le système s'auto-teste et s'auto-étalonne. Toutes les localisations de mesure sont paramétrables, soit par des fichiers de configuration (composants), soit par menu interactif (colle).
- **Système de contrôle du maillage des tissus** : Destiné aux industriels du tricotage, ce système de *Delta Technologies* contrôle la largeur et le maillage des tissus. La vitesse de défilement des tissus est de 50 à 60m/s. Une caméra linéaire fournit des images 1-D à partir desquelles il est possible de mesurer une largeur de laize en 0.2s, avec une précision de  $\sim 2mm$  pour une laize de 1000mm. Une caméra matricielle fournit des images 2-D à partir desquelles il est possible de compter les mailles en rangées et colonnes en 3s (objectif de 1s), avec une précision de  $\sim 1$  maille sur 30mmx30mm inspectés. Ce contrôle non destructif sans contact garantit un suivi continu de la structure du tissu au cours de son traitement. A chaque mise en route du système, une procédure automatique d'inspection des équipements est réalisée. . Avant toute opération, l'opérateur entre dans une base de données des informations sur le tissu (référence, No de lot, largeur, etc.). .). Le contrôle est initié par l'opérateur après engagement du tissu, son arrêt est automatique en fin de passe. . Les mesures effectuées sont stockées dans des fichiers ainsi que le jour et l'heure de réalisation de la mesure en vue d'une exploitation ultérieure.
- **La vision artificielle** : Deux exemples de systèmes dotés d'une vision intelligente sont indiqués ci-dessous :
  - robot capable de se déplacer le long d'un trottoir, de détecter et vérifier les panneaux de circulation aux carrefours, et de vous dire "bonjour" en reconnaissant votre visage. La première, et certainement la plus facile, étape consiste à simuler la vision pour que le robot perçoive son environnement à l'aide de capteurs électromagnétiques (caméra CCD, laser, etc.). Des logiciels spécifiques ont pour but de faire aussi bien et même mieux que les systèmes de visualisation et d'interprétation de l'homme.
  - Une caméra placée le long d'une route et capable de reconnaître les voitures qui roulent trop vite , de les prendre en photo, de détecter la voiture dans l'image, puis la plaque d'immatriculation sur la voiture , de lire automatiquement le numéro de la plaque d'immatriculation ,puis de trouver dans une banque de données le nom et l'adresse du conducteur .

### IV. Indexation et recherche d'image

## Chapitre II : traitement d'image

L'indexation et la recherche d'images deviennent un domaine très actif depuis 1994. Le nombre de publications dans ce domaine est environ 1000 publications chaque année. Un Système d'indexation et de recherche d'images permet aux utilisateurs de retrouver les images qui satisfont leurs besoins dans des bases de données d'images. Son fonctionnement est décomposé en deux étapes : l'étape d'indexation dit, hors ligne (offline) ou le système va extraire les caractéristiques des images de la base et les stocker dans une base de données. La deuxième étape consiste à extraire les descripteurs de l'image requête et la comparer avec les descripteurs existant dans la base de données afin de trouver les images similaires à celle désirée. Nous allons représenter trois aspects importants : la recherche locale vs globale, l'ontologie et le retour de pertinence .

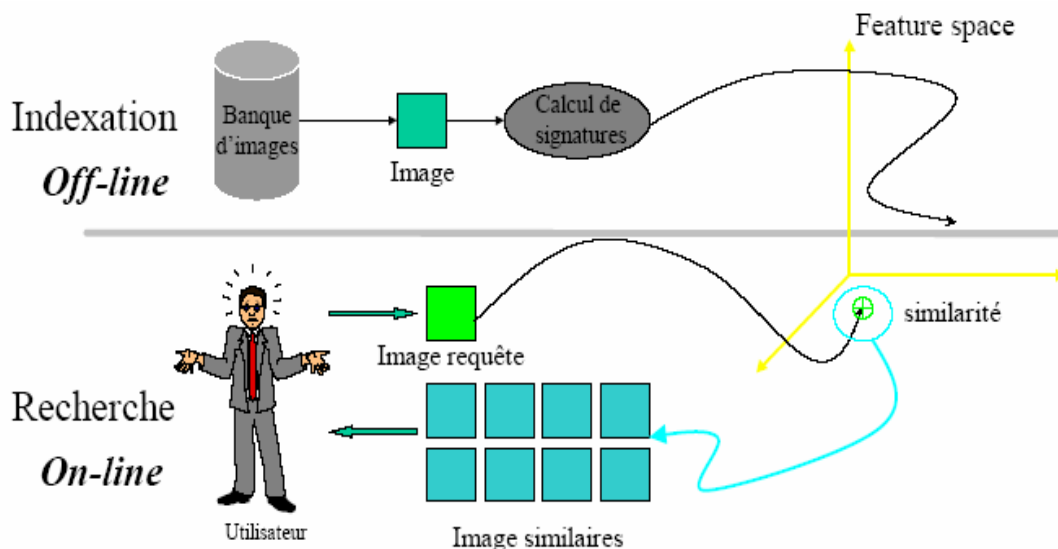


Figure I

### I.10: aperçu d'un système d'indexation et de recherche d'image.

#### IV.1 La recherche locale vs globale

Les premières approches proposées pour l'indexation et la recherche d'images ont utilisé les descripteurs globaux d'images . Les descripteurs globaux tels que l'histogramme des couleurs sont calculées avec la participation de tous les pixels dans une image . La disposition des pixels n'est donc pas prise en compte. Évidemment, de telles approches ne sont pas appropriées pour la recherche d'images ayant une seule ou quelques parties similaires à la requête. La recherche d'images locale doit soit comprendre une segmentation d'images (la catégorie 1) soit utiliser des descripteurs locaux (la catégorie 2). Nous appelons les approches de la première catégorie, les approches d'indexation et de recherche d'images au niveau régions. Pour ces approches, la méthode de segmentation peut être simple ou complexe. L'objectif de cette segmentation est de décomposer une image en quelques régions. Dans le cas idéal, chaque région correspond à un objet réel. Le noyau des approches d'indexation et de recherche d'images au niveau régions est la mise en correspondance entre images. Cette mise en correspondance doit tenir compte de deux caractéristiques. La première est le nombre de régions varie d'une image à l'autre la deuxième est la segmentation n'est pas parfaite, une région peut être appariée à plus d'une région. Dans [Carson 2002], les auteurs ont présenté un système appelé Blobworld pour l'indexation et la recherche d'images au niveau régions. Dans ce système, la requête peut être une ou plusieurs régions. Pour des requêtes contenant une région, soit  $v_i$  le vecteur de descripteurs de la région recherchée, la mise en correspondance entre la région recherchée et une image cible s'effectue en suivant les trois étapes :

- \_ Pour chacune de régions  $v_i$  de l'image cible , la distance de mahalanobis entre  $v_i$  et  $v_j$  est définie par :

## Chapitre II : traitement d'image

$$d_{ij} = \frac{v_i}{(v_i+v_j)^T \sum \dot{c} - v_j} \dots\dots\dots(II.1)$$

La similarité entre deux régions est calculée par :

$$u_{ij} = e^{\frac{-d_{ij}}{2}} \dots\dots\dots(II.2)$$

La similarité entre la région recherchée et l'image cible :

$$u_i = \max_j u_{ij} \dots\dots\dots(II.3)$$

Pour des requêtes contenant plusieurs régions , les auteurs ont appliqué la logique floue pour calculer la similarité entre cette requête et une image de la base en se basant sur celle des requêtes comprenant une seule région. Un exemple d'une requête contenant plusieurs régions est « retrouver des images comportant (une région similaire à la région 1) et (une région similaire à la région 2 ou une région similaire à la région 3) ». la similarité entre cette requête et une image cible est déterminée par :

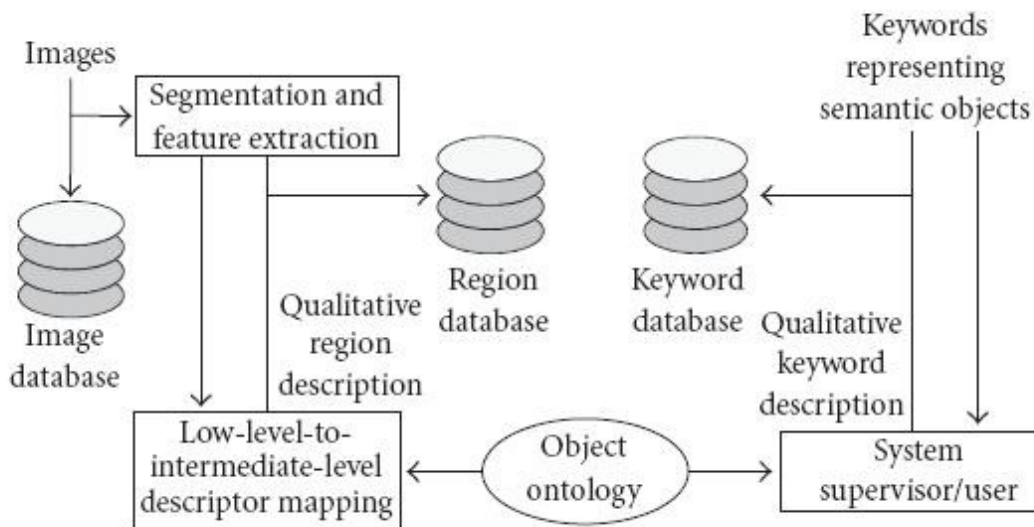
$$\min \{u_1, \max \{u_2, u_3\} \} \dots\dots\dots(II.4)$$

où  $u_1$  ,  $u_2$  et  $u_3$  sont les similarités entre cette image et trois requêtes comprenant les régions 1, 2 et 3.

La méthode de Carson et al. [Carson 2002] tient compte de la première caractéristique (le nombre de régions varie d'une image à l'autre). Cependant, la deuxième caractéristique (la segmentation n'est pas parfaite, une région peut être appariée à plus d'une région) n'est pas prise en compte.

### IV.1.1 Ontologie

Des approches d'indexation et de recherche d'images basées sur l'ontologie cherchent tout d'abord à avoir une représentation formelle sous la forme d'une ontologie pour la connaissance du domaine. Des images sont ensuite indexées et retrouvées en utilisant des concepts de l'ontologie. Trois approches remarquables dans cet aspect sont l'approche de Mezaris et al. celle de Maillot et al. et celle de Lim et al .



**Figure II.11 :Indexation et recherche d'images avec une ontologie.**

Une ontologie qui définit la mise en correspondance entre des descripteurs à bas niveau extraits sur des régions et des termes à intermédiaire niveau est formulée par :

## Chapitre II : traitement d'image

$$O := \dot{\zeta} (D, \leq_D, R, \sigma, \leq_R) \dots \dots \dots (II.5)$$

où :

D : l'ensemble de concepts visuels (p. ex. la forme)

R : l'ensemble de concept de relation (p.ex. la position relative )

$\leq_D$  : un ordre partiel qui décrit l'hiérarchie de concepts (p.ex. la luminance est un sous-concept de l'intensité)

$\sigma$  : une fonction

$$\sigma : R \rightarrow D^{\dot{\zeta}}, \sigma(r) = (\sigma_{1,r}, \sigma_{2,r}, \dots, \sigma_{\Sigma,r}) \text{ avec } \sigma_{i,r} \in D \dots \dots \dots$$

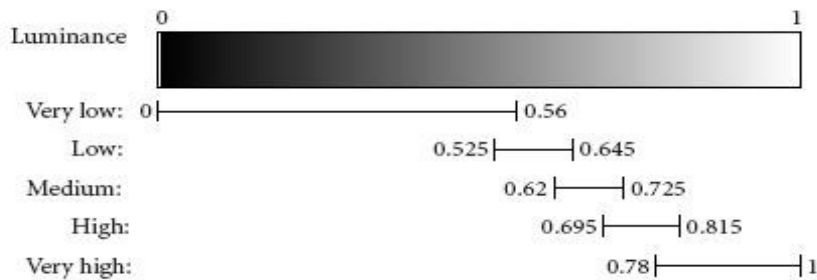
$$(II.6)$$

ex.

$\sigma(r) = ('position', 'position')$

$\leq_R$  : un ordre partiel qui décrit l'hiérarchie de relations

Les requêtes sont exprimées en utilisant cette ontologie. La mise en correspondance entre des descripteurs à bas niveau et des termes à intermédiaire niveau dans [Mezaris 2004] consiste simplement à déterminer un intervalle de valeurs pour chaque terme.



**Figure II.12 : correspondance entre des descripteurs a bas niveau et des termes à intermédiaire niveau pour la luminance.**

### IV.1.2 Retour de pertinence

Les résultats de recherche d'information ne sont pas toujours parfaits en raison du fossé sémantique. La méthode pouvant combler ce fossé doit permettre de communiquer avec l'utilisateur. La recherche d'information interactive est une recherche d'information qui permet à l'utilisateur de faire un retour de pertinence. Le retour de pertinence est un processus qui consiste à apprendre à partir des retours de l'utilisateur et à trouver de nouveaux résultats pour répondre à l'utilisateur en se basant sur la connaissance apprise. L'approche permettant de faire un retour de pertinence comprend deux processus : un apprenant et un sélectionneur. Le sélectionneur effectue la recherche en utilisant la connaissance apprise par l'apprenant et rend les nouveaux résultats à l'utilisateur. L'utilisateur démarre le processus de recherche en exprimant sa requête. Le système effectue la mise en correspondance entre les éléments indexés et la requête. La liste des résultats ordonnés de manière décroissante par leurs similarités avec la requête est rendue à l'utilisateur. Si l'utilisateur est satisfait par la réponse, il finit le processus de recherche (sans avoir démarré le retour de pertinence). Sinon le retour de pertinence démarre. Il demande à l'utilisateur d'annoter les résultats obtenus comme exemples positifs et/ou exemples négatifs. L'apprenant est activé pour apprendre l'intention de l'utilisateur à partir de ses retours.

Les techniques de retour de pertinence peuvent être classées en trois grandes familles : le retour de pertinence basée sur la modification de requête appelé QVM (Query Vector Modification), celui basé sur le rôle du descripteur appelé FRE (Feature Relevance Estimation), et celui basé sur la classification. Les techniques de retour de pertinence basées sur la modification de requête consiste à reformuler des requêtes en utilisant des images positives et des images négatives de manière que des résultats de la prochaine itération comprennent plus de résultats pertinents soit  $X_j^{(i)}$ ,  $X_i^{(j+1)}$

## Chapitre II : traitement d'image

les formulations de requête correspondant à une image recherchée  $i$  à l'itération  $j$  et  $j + 1$ ,  $X_i^{(j+1)}$  est déterminée par  $X_i^{(j)}$  et des images positives et négatives selon la formule suivante :

$$X_i^{(j+1)} = \alpha X_i^j + \beta \sum_{Y_k \in R} \frac{Y_k}{|R|} \dots\dots\dots (II.7)$$

$$\gamma \sum_{Y_k \in N} \frac{Y_k}{|N|} \dots\dots\dots (II.8)$$

où  $R$  et  $N$  sont respectivement l'ensemble d'images positives et celui d'images négatives. Les paramètres  $\alpha$ ,  $\beta$ ,  $\gamma$  décident la participation de chaque composant

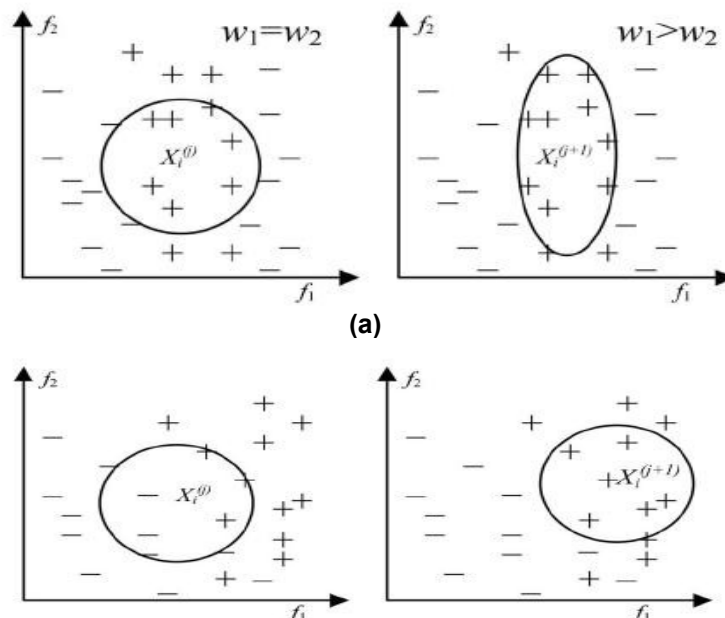
Les techniques de retour de pertinence basées sur le rôle du descripteur visent à déterminer l'importance de chaque descripteur pour la mise en correspondance entre des images. L'importance de chaque descripteur est initialisée par la même valeur. Soit  $X = (x_1, x_2, \dots, x_d)$ ,  $Y = (y_1, y_2, \dots, y_d)$  deux vecteurs de descripteurs de deux images, la distance entre deux image :

$$\text{Dist}(X, Y) = \sum_{i=1}^d (x_i - y_i)^2 \dots\dots\dots (II.9)$$

L'importance de chaque descripteur  $w_i$  est mise à jour tout au long de la session de recherche. Plus le descripteur est pertinent, plus son importance est élevée. Lorsque l'importance de chaque descripteur est déterminée, la distance entre deux images (cf. équation II.9) devient :

$$\text{Dist}(X, Y) = \sum_{i=1}^d w_i (x_i - y_i)^2 \dots\dots\dots (II.10)$$

Les techniques de retour de pertinence basées sur la classification entraînent des classificateurs pour des images positives et/ou pour des images négatives. Des images de la base sont comparées avec ces classificateurs. Les machines à vecteurs de support (SVM) et les classificateurs bayésiens sont largement choisis.



(b)  
**Figure II.13 : Illustration de retour de pertinence basé (a) sur la modification de requête (b) et sur le rôle de chaque descripteur dans le cas où deux descripteurs  $f_1, f_2$  sont utilisés**

### V. Les caractéristique

## Chapitre II : traitement d'image

### V.1 Définition d'une caractéristique

une caractéristique est une entité spécifique caractérisée par divers attribut telles que la couleur, la texture et la forme .

### V.2 Les types des caractéristiques

#### V.2.1 caractéristiques au niveau des pixels (Pixel-level features)

caractéristiques calculées à chaque pixel, par exemple couleur, l'emplacement ...

#### V.2.2 Caractéristiques locales (Local features)

caractéristiques calculées sur les résultats de la subdivision de la bande image sur la segmentation d'images ou de détection des bords. 26 Caractéristiques Une caractéristique locale est un motif d'image qui diffère de son voisinage. Elle est généralement associée à un changement d'une propriété d'image ou plusieurs propriétés en même temps, bien qu'elle ne soit pas nécessairement localisée exactement sur ce changement. Les propriétés d'une image couramment considérées sont l'intensité, la couleur et la texture.



Figure II.14: l'importance des coins et des jonctions dans la reconnaissance visuelle

#### V.2.3 globales (Global features)

caractéristiques calculées sur toute l'image ou tout simplement régulièrement sous-zone d'une image Dans le domaine de la recherche d'images, de nombreuses caractéristiques globales ont été proposées pour décrire le contenu de l'image, avec des histogrammes de couleurs et des variations de ceux-ci. Cette approche fonctionne étonnamment bien, au moins pour les images avec des couleurs distinctives, tant qu'il est en général une composition de l'image dans son ensemble dont l'utilisateur est intéressé, plutôt que l'objet de l'avant- plan. En effet, les caractéristiques globales ne peuvent pas distinguer l'avant- plan de l'arrière-plan, et mélangent l'information des deux parties ensemble Les caractéristiques globales ont été également utilisées pour la reconnaissance d'objets, résultants dans les premières approches fondées sur l'apparence pour relever ce défi.

Turk et Pentland et plus tard Murase et Nayar ont proposé une analyse de composantes principales d'un ensemble d'images et d'utiliser les projections sur les quelques premiers composantes principales comme descripteurs . Une toute nouvelle gamme d'objets naturels pourrait tout d'un coup être reconnue. Cependant, étant basé sur une description globale, l'encombrement d'image et les occlusions forment à nouveau un problème majeur, ce qui limite l'utilité du système à des cas ayant

## Chapitre II : traitement d'image

des arrière-plans propres ou où l'objet peut être segmenté, par exemple, en se fondant sur des informations de mouvement.

### V.3 Caractéristique local contre caractéristique global

les caractéristiques locales permettent non seulement de trouver des correspondances en dépit de grands changements dans les conditions de visualisation, occlusions, et l'encombrement d'image ( appariement large de base), mais aussi donner une description intéressante du contenu d'image pour la récupération de celle-ci et d'objet ou de tâches de reconnaissance de scène (à la fois pour des objets spécifiques, ainsi que les catégories). Pour mettre cela en contexte, nous résumons brièvement quelques stratégies alternatives pour calculer la représentation d'images, y compris les caractéristiques globales, des segments d'image, et échantillonnage exhaustif et aléatoire de caractéristiques.

### V.4 Détection de caractéristique

#### V.4.1 Définition

c'est la première étape dans le traitement de l'image consiste à représenter sous forme d'un vecteur caractéristiques les données issues d'une chaîne de segmentation en régions d'images puis à calculer la distance entre ces images .Dans la recherche d'images similaires par le contenu, diverses statistiques (basés sur la couleur, la texture et la forme) forment la description de l'image. Ces statistiques, que nous appelons ici descripteurs peuvent être mises sous la forme d'un vecteur. La distance entre deux images est alors la distance dans un espace à n dimensions, où n est le nombre de descripteurs choisi. Ce mode de représentation implique un choix définitif dans la nomenclature utilisée pour représenter les images : chaque image doit avoir le même nombre de descripteurs et chaque descripteur possède une place précise au sein du vecteur. Parmi ces descripteurs on distingue deux famille :

#### V.4.2 Les détecteurs de caractéristique local

##### V.4.2.1 Détection des point d'intérêt

La détection de points d'intérêts (ou coins) est, au même titre que la détection de contours, une étape préliminaire à de nombreux processus de vision par ordinateur . Les points d'intérêts, dans une image, correspondent à des doubles discontinuités de la fonction d'intensités. Celles-ci peuvent être provoquées, comme pour les contours, par des discontinuités de la fonction de réflectance ou des discontinuités de profondeur. Ce sont par exemple : les coins, les jonctions en T ou les points de fortes variations de texture. Il ne s'agit pas seulement de trouver des correspondances entre des images mais de tenir compte des changements de points de vue (rotation, déformations locales), de l'échelle, la luminosité des clichés et du bruit. La nature des points d'intérêts détectés dépend du type de détecteur utilisé : détecteurs de coins, d'intersection, de contours, de points isolés ou de zones homogènes.



Figure II.15 : Différents types de points d'intérêts ( coins, jonction en T et point de fortes variations de texture)

- **les approches basées sur les contours** : c'est les points de courbure maximale ou les intersections de droites qui sont détectés dans ces approches.
- **les approches s'appuyant sur des modèles** : les éléments de l'image sont comparés à une bibliothèque de modèles représentant les configurations les plus fréquemment présentes dans

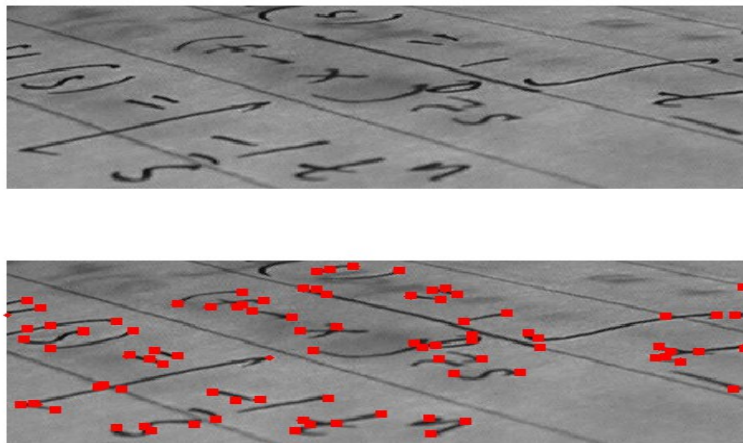
## Chapitre II : traitement d'image

les images. Ces méthodes s'avèrent d'une grande précision mais souffrent d'un manque de répétabilité.

- **les approches basées sur l'intensité des pixels de l'image** : Ces approches sont les plus répandues car elles sont relativement stables. Elles se concentrent sur la détection de coins et de points isolés. Les détecteurs les plus connus sont ceux de Moravec, Harris, SIFT (Scale Invariant Feature transform), SURF (Speeded Up Robust Features), MSER (Maximally Stable External Regions), et FAST (Features from Accelerated Segment Test). Nous présentons ci-dessous les détecteurs les plus couramment utilisés.

### A. Le détecteur de harris :

Un des détecteurs les plus utilisés est probablement le détecteur de coins de Harris basé sur les valeurs propres de la matrice de covariance de la région analysée (plus précisément, sur le tenseur de structure) .



**Figure II.16 :détecteur de point de harris**

Néanmoins, le détecteur de coins de Harris n'est pas invariant aux changements d'échelle. C'est-à-dire que cette méthode ne garantit pas que le même objet, représenté à deux échelles différentes, soit détecté de la même manière. Pour parer à ce désavantage, Lindeberg introduit le concept de sélection automatique d'échelle, qui permet de détecter des points d'intérêt d'une image, chacun à leur échelle caractéristique. Le détecteur de Lindeberg est basé sur la recherche des maxima du déterminant de la matrice hessienne. Pour rappel, la matrice hessienne d'une fonction numérique  $f$  est la matrice carrée, notée  $H(f)$ , de ses dérivées partielles secondes. Le principe de cette approche est que les dérivées d'ordre deux permettent de détecter la courbure des valeurs d'intensité des pixels, et donc, au final, les blobs de luminosité de l'image. Afin de détecter ces structures en forme de blobs, Lowe recherche les maxima du Laplacien (qui correspond à la trace de la matrice Hessienne). Mikolajczyk et Schmid améliorent cette méthode et créent un détecteur robuste (Hessian-Laplace) utilisant le déterminant de la matrice hessienne pour détecter la région d'intérêt et le Laplacien pour choisir l'échelle. Le détecteur de Harris-Laplace utilise lui le détecteur de coins de Harris pour détecter les points d'intérêt et le Laplacien pour le choix de l'échelle. On peut voir sur la figure suivante que le détecteur de Harris-Laplace déclenche sur les coins présents dans l'image alors que le détecteur Laplacien déclenche sur les structures de type blobs.

#### A.1 Le détecteur harris et ses variantes :

- **Détecteur de harris Laplace** : Le détecteur de Harris Laplace et Harris invariant affine aux transformations de l'échelle et aux transformations affines. En résumé, le détecteur de Harris Laplace utilise l'échelle adaptée pour localiser des points dans l'espace de l'échelle. Donc il

## Chapitre II : traitement d'image

sélectionne les points pour lesquels le Laplacien de la Gaussienne atteint un maximum sur une échelle donnée, en utilisant l'équation suivante :

$$\log \frac{\Delta L(x, \sigma_n)}{\sigma_n^2} = \sigma_n^2 |L_{xx}(x, \sigma_n) + L_{yy}(x, \sigma_n)| \dots\dots\dots (II.11)$$

Avec  $\sigma_n$  est l'échelle .

La fonction de Harris multi échelle est utilisée pour sélectionner les positions des points d'intérêt . On applique la formule d'échelle suivante :

$$\sigma_n = \varepsilon^n \sigma_0 \dots\dots\dots (II.12)$$

qui donnent les différentes échelles en fonction de  $\sigma_0$  qui permettent de calculer les caractéristiques selon chacune des échelles .  $\varepsilon$  est le facteur de l'échelle entre niveaux consécutifs.

En chaque niveau de la représentation, les points d'intérêt sont extraits en détectant les maximums locaux dans les 8 voisinages d'un point x.

Le moment second ordre  $M(x, \sigma_n)$  est calculé en utilisant l'échelle d'intégration  $\sigma_I$  et l'échelle de différentiation  $\sigma_D$  tel que :

$$\sigma_I = \sigma_D = \sigma_n \dots\dots\dots (II.13)$$

pour chaque point, un algorithme itératif détecte simultanément l'emplacement et l'échelle du point d'intérêt. L'échelle du point d'intérêt est calculée en utilisant la fonction LoG . Ainsi les points pour lesquels la réponse de LoG n'atteint aucun extremum et pour lesquels la réponse est au-dessous du seuil sont rejetés.

- **Détecteur de Harris-Affine** : La matrice second ordre est souvent utilisée pour décrire la structure de l'image locale. Par ailleurs, pour rendre cette matrice indépendante de la résolution de l'image, on utilise une matrice hog svm t adaptée aux changements d'échelle.

$$M(x, y, \sigma_I, \sigma_D) = \frac{1}{\sigma_D g} \begin{pmatrix} L_{xx} & L_{xy} \\ L_{xy} & L_{yy} \end{pmatrix} \dots\dots\dots (II.14)$$

$\sigma_I$  : l'échelle de l'intégration

$\sigma_D$  l'échelle de la différenciation

$L_\alpha$  : La dérivée calculée dans la direction

g : est un noyau gaussien

Les dérivées locales sont calculées avec un noyau gaussien de la dimension déterminée par l'échelle locale.

Les valeurs propres de cette matrice présentent deux principaux changements du signal dans le voisinage d'un point. Cette propriété permet l'extraction de points pour lesquels les deux courbures sont considérables.

$$\text{Corners} = \det(M(x, y, \sigma_I, \sigma_D)) \dots\dots\dots (II.15)$$

Les maximums locaux de la corness déterminent l'emplacement des points d'intérêt.

## Chapitre II : traitement d'image

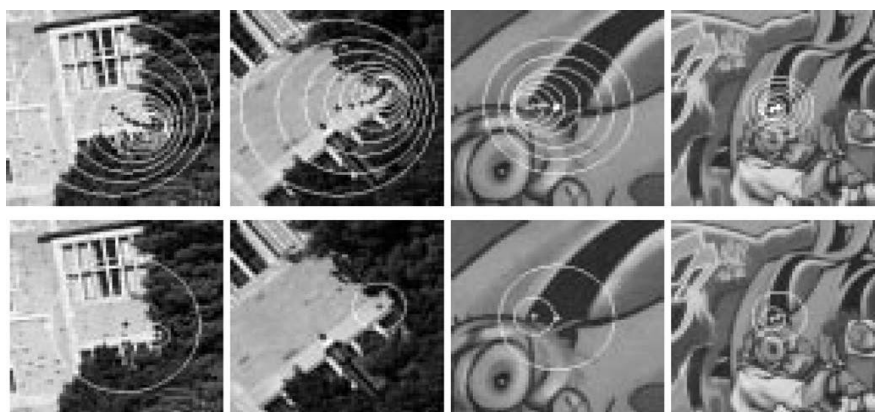


Figure II.17 : Détection de points d'intérêt invariants à l'échelle

### A.2 Détecteur SuperPixel /MSER

Le détecteur SuperPixel est un détecteur de régions nommées superpixels et obtenues par segmentation. L'image est alors considérée comme un ensemble de superpixels ; il y a abstraction de l'unité de niveau qui est le pixel. Ce détecteur présente l'avantage d'avoir un bon compromis entre la localité de la primitive et sa distinctivité (généralement basée sur un large voisinage). D'autre part, il est un des rares à couvrir toute l'image sans chevauchement entre les régions. Le résultat de la segmentation d'une image est illustré sur la figure. Toutefois, il n'est pas invariant à l'échelle et n'est pas adapté pour l'appariement ou la reconnaissance d'objets. Ce détecteur est surtout utilisé du fait que les régions obtenues ont beaucoup plus de sens que les pixels en eux-mêmes, elles présentent donc un intérêt sémantique. Le détecteur MSER (Maximally Stable Extremal Regions) [Matas *et al.*, 2004] est aussi un détecteur de régions. Basé sur l'intensité des pixels, il ne présente pas les mêmes avantages que le détecteur superpixel. En effet, il ne couvre pas toute l'image dans la solution apportée et ne garantit pas que les régions ne se chevauchent pas. Par contre, il est invariant aux transformations affines géométriques et photométriques. Il peut aussi être rendu invariant aux changements d'échelle. Cependant, comme le superpixel, il n'est pas adapté pour l'appariement ou la reconnaissance d'objets.



Figure II.18 : Illustration de la segmentation d'une image en superpixels

### A.3 Le détecteur SIFT

## Chapitre II : traitement d'image

Le détecteur SIFT (Scale Invariant Feature transform) , développé par Lowe est très probablement le plus utilisé. Il présente l'avantage majeur d'être à la fois invariant aux rotations et aux changements d'échelle. De plus, la densité de points détectés est élevée. La particularité de la méthode réside dans le calcul combiné des points d'intérêt et de leurs descripteurs associés. Un descripteur est un vecteur caractérisant le voisinage local en un point d'intérêt. Il caractérise le point d'intérêt de par son unicité. D'autres méthodes ont été développées pour faire évoluer SIFT. Morel et Al, s'intéressent aux déformations de l'image, modélisables localement par une fonction affine en considérant la scène localement plane. Ke et Sukthakar présentent l'évolution PCA-SIFT (Principal Components Analysis SIFT) où la rapidité d'exécution a été optimisée.

### A.4 Détecteur SURF

Le détecteur SURF (Speeded Up Robust Features) est une amélioration du détecteur SIFT. En effet, il pallie le manque d'efficacité de ce dernier en modifiant le calcul de la différence de Gaussiennes long à effectuer. Il est remplacé par une approximation de la matrice Hessienne d'un noyau Gaussien calculée rapidement à partir du principe de l'image intégrale introduite par [Viola et Jones, 2001]. L'enregistrement de l'information pour le calcul du descripteur est aussi différent : l'orientation de la primitive est analysée au travers d'un échantillonnage du voisinage sous forme circulaire. La primitive est ainsi rendue invariante à la rotation. Le résultat est soumis à une transformée en ondelettes de Haar. De manière générale, le détecteur SURF présente les mêmes avantages que le SIFT. L'approximation de la matrice Hessienne n'engendre pas d'erreur importante d'estimation et permet un calcul en temps réel .

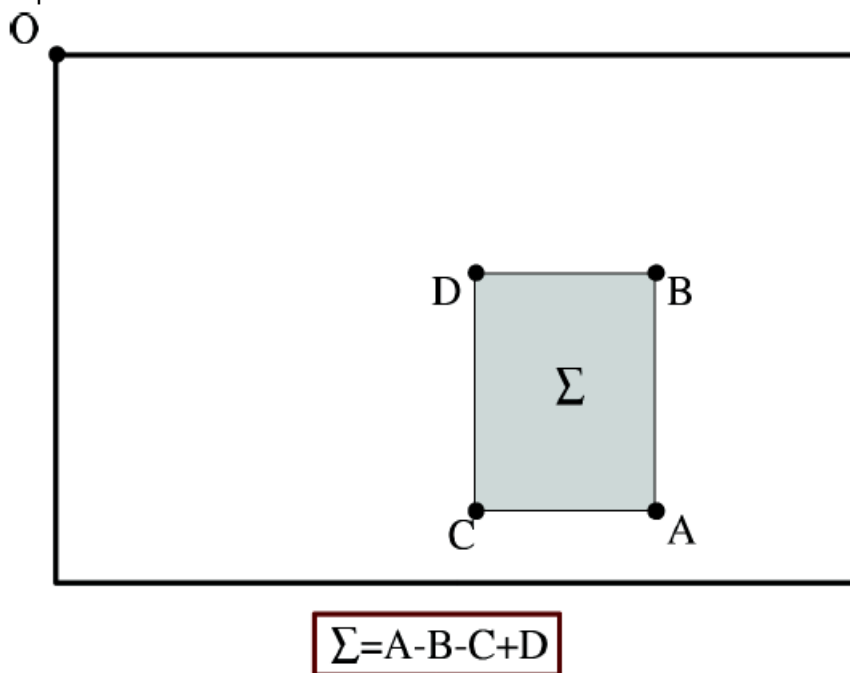


Figure II.19 : Principe de l'image intégrale

### A.5 Le filtre de Canny-Deriche

Ce filtre qui est un filtre passe-bande a un double objectif: sélectionner les fréquences relativement élevées pour mettre en évidence les variations rapides dans l'image, car ces variations rapides correspondent en général aux contours; tout en filtrant autant que possible le bruit dont l'importance relative est plus grande dans les hautes fréquences. Le paramètre  $\alpha$  doit être réglé en fonction de l'application envisagée. L'implémentation séparable est par exemple celle de Deriche (filtre de "Canny-Deriche") sous la forme d'un filtre récursif traitant chaque ligne ou chaque colonne dans un sens puis dans l'autre. Cette forme séparable ne conserve pas l'isotropie: un filtre ne traite pas de la même manière les verticales ou les horizontales que les obliques. Pourtant, si on l'application autorise le traitement par transformée de Fourier, il est possible de réaliser le filtrage

## Chapitre II : traitement d'image

dans le domaine des fréquences bidimensionnelles. L'expression du filtre de Canny dans le domaine des fréquences est C'est sur l'image filtrée ligne par ligne puis colonne par colonne que sont détectées les variations rapides du signal qui sont interprétées comme un contour ou comme du bruit.

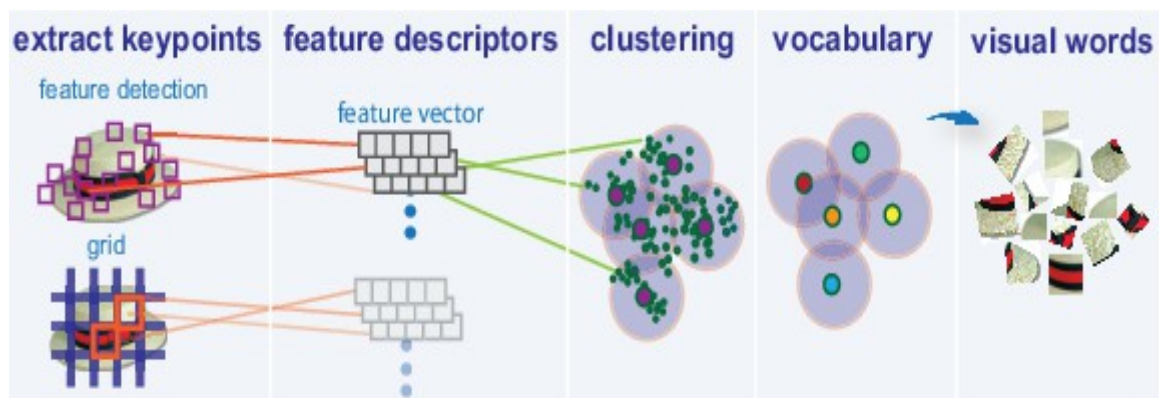
### A.6 Détecteur DAISY

DAISY, est un détecteur mixant les avantages de deux détecteurs très performants : SIFT et GLOH. Ainsi, ce détecteur est basé sur un noyau permettant de créer une information invariante à l'échelle par l'intermédiaire d'une analyse multi-échelle sur un voisinage proche du point étudié. De plus, il est invariant à la rotation et à la translation, plutôt résistant aux occlusions et aux transformations photométriques et géométriques. Il apporte donc un taux de répétabilité élevé adjoint d'une bonne information permettant la distinction tout en conservant la précision. Il s'agit d'un très bon compromis entre toutes les propriétés d'un détecteur. De plus, la précision du détecteur permet la reconstruction de cartes de profondeurs proches de la vérité terrain. Enfin, il améliore l'approche calculatoire du SIFT en utilisant des simplifications par convolution de Gaussiennes, réduisant le nombre d'opérations élémentaires nécessaires et ainsi son temps d'exécution. DAISY peut donc être utilisé en temps réel. A la différence de nombreux détecteurs vus précédemment, DAISY est un détecteur dense. C'est-à-dire qu'il s'agit d'une méthode de calcul d'un descripteur mais qu'il n'y a pas de détecteurs pour sélectionner les primitives dans l'image. Il est dense dans le sens où il est destiné à être calculé pour chacun des pixels de l'image (opération longue). Toutefois, il est possible de le combiner avec un détecteur de primitives locales comme Harris et ensuite de calculer les descripteurs aux localisations des points de Harris.

### V.4.3 Détecteurs de caractéristique global

#### A.1 Bag of Feature

Un sac de Caractéristiques est celui qui représente les images comme une collection de caractéristiques locales désordonnées. Le nom vient du sac de la représentation des mots utilisés dans la récupération textuelle de l'information. Cette section fournit une explication du sac de Caractéristiques d'image. Il y a deux perspectives communes pour expliquer la représentation d'image BoF. D'abord par analogie avec sac de mots, on représente un histogramme normalisé des comptes de mots. Généralement, on compte tous les mots à partir d'un dictionnaire qui apparaissent dans le document. Ce dictionnaire peut exclure certains mots d'information tels que des articles (comme «le»), et il peut avoir un seul terme pour représenter un ensemble de synonymes. Le vecteur d'expression qui représente le document est un vecteur creux où chaque élément est un terme dans le dictionnaire et la valeur de cet élément est le nombre de fois que le terme apparaît dans le document, divisé par le nombre total de mots du dictionnaire. Et aussi par un vocabulaire visuel est construit pour représenter le dictionnaire par des caractéristiques de clustering extraites d'un ensemble de formation images. Les caractéristiques de l'image représentent des zones locales de l'image.



Fig

ure II.20 : les étapes à suivre pour créer bof

#### A.2 Le détecteur GIST

## Chapitre II : traitement d'image

L'idée principale de ce détecteur global est d'extraire l'information la plus importante d'une image, ce qui est aperçu à la première vue, sans un réel souci de précision. Son intérêt réside dans la forte interprétation sémantique de la scène. A partir des informations extraites par le détecteur, il est possible de différencier les lieux, voire scènes, de par leur sémantique plutôt que par des primitives invariantes à un certain nombre de transformations. Il en résulte qu'il est a priori le détecteur idéal pour faire de la localisation topologique. Toutefois, il est montré que le manque de précision de ce détecteur grande information sémantique mais faible distinctivité locale, liée aux primitives engendre des erreurs lors d'une recherche dans une base de données importante ; l'erreur étant qu'il permet d'obtenir une image proche de celle demandée mais pas l'image demandée en elle-même. La signature (descripteur) est simple et de longueur constante quelque soit la taille de l'image et la quantité d'information. Le calcul du descripteur peut être fait en temps réel.

### V.5 Description de caractéristique

la description est la représentation de l'information obtenue à partir du détecteur. Chaque détecteur est généralement élaboré avec son propre descripteur car il est nécessaire lors de sa conception de pouvoir l'enregistrer en mémoire. Toutefois, certains descripteurs sont plus performants que d'autres et présentent un certain nombre d'avantages comme un temps de calcul restreint pour retrouver une primitive à partir de son descripteur. Il existe une multitude de méthodes de description chacune ayant ses avantages et ses inconvénients. Il est donc indispensable d'en connaître les caractéristiques afin de choisir la plus appropriée à la problématique.

#### V.5.1 Descripteur de caractéristique local :

##### V.5.1.1 Descripteur basé sur les moments

La description basée sur les moments a été initialement proposée pour de la reconnaissance d'objets. Ils étaient utilisés pour travailler sur des images contenant l'objet entier. L'avantage des moments, tels que les moments de Hu et de Zernike, est leur invariance aux translations, rotations et changements d'échelle isotrope. Néanmoins, un changement de point de vue, un changement d'échelle anisotrope ou l'ajout d'occultations, provoque une forte diminution de la qualité des résultats. En se basant sur les études comparatives proposées par Choksuriwong et al. en 2005 et 2008, nous proposons d'en détailler les principaux composants.

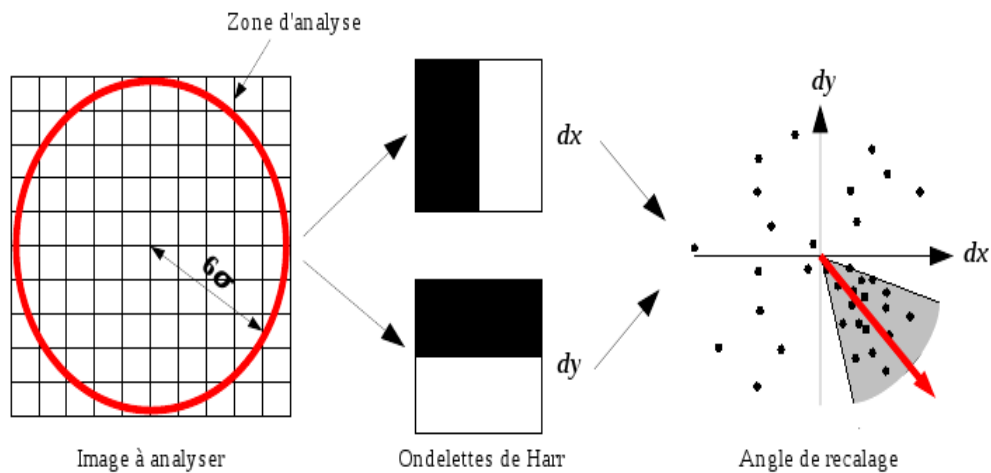
##### A.1 Descripteur SIRF

Lowe propose également en 1999 puis en 2004, de coupler le détecteur DoG à un descripteur local basé sur des histogrammes de gradients orientés. Les données fournies initialement par le détecteur sont les coordonnées des points d'intérêt ainsi que leur échelle caractéristique. Une étape préliminaire consiste à construire l'histogramme des orientations locales définies en chaque point (x)

##### A.2 Descripteur SURF

En 2006, Bay et al. proposent une nouvelle méthode de description locale de points d'intérêt, nommée SURF (Speeded-Up Robust Features). Fortement influencés par l'approche de SIFT, ils couplent une étape de recalage de la zone d'analyse avec la construction d'un histogramme de gradients orientés. La première étape de leur processus est donc de déterminer l'angle de rotation (ou de recalage) à appliquer à la fenêtre de description locale. Pour se faire, les auteurs appliquent des ondelettes de Haar sur l'image intégrale permettant ainsi de diminuer les temps de calculs de façon significative. Ces ondelettes permettent de calculer les dérivées premières de l'image sur un voisinage carré et d'étudier ainsi la répartition des gradients horizontaux et verticaux. Dès lors les réponses des ondelettes permettent de tracer le graphique de distribution des gradients et d'en déduire l'angle de recalage.

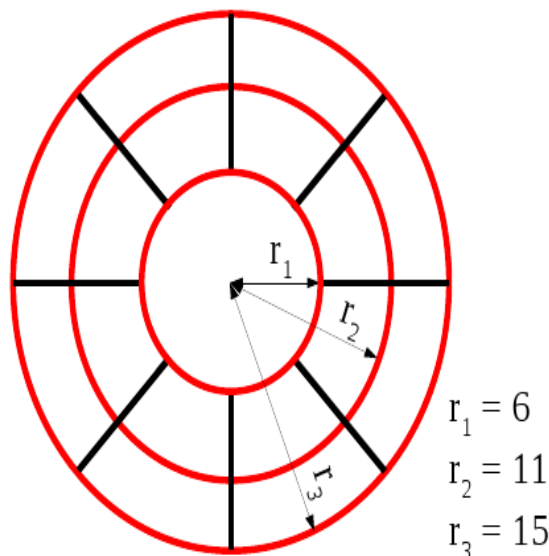
## Chapitre II : traitement d'image



**Figure II.21 : Détermination de l'angle de recalage du SURF**

### A.3 DESCRIPEUR GLOH

Détaillé en 2005 par Mikolajczyk et Schmid puis testé la même année par Dalal et Triggs , le descripteur GLOH (Gradient location-orientation histogram) a été proposé afin d'augmenter les performances du SIFT. Il est repris en 2009 par Chandrasekhar afin d'étudier toutes les possibilités de cette méthode et de l'améliorer. L'idée générale est de construire un histogramme de gradients orientés dans un plan circulaire. Pour ce faire, le GLOH est constitué de dix-sept zones d'analyse suivant trois paramètres radiaux et huit paramètres angulaires. En chacune des zones, un histogramme de gradients orientés est construit suivant seize classes (intervalles de 8°).



**Figure II.22 : Masque d'analyse de la méthode GLOH**

### A.4 Descripteur DAISY

Le descripteur DAISY proposé par Tola et al. en 2008 est une nouvelle approche de description locale s'inspirant des avantages des méthodes SIFT et GLOH. Elle a pour objectif d'accélérer les temps de calculs et d'améliorer la gestion des invariances. L'idée principale est de remplacer les calculs de gradients des méthodes précédemment citées par des filtres de dérivées gaussiennes orientées.

### V.5.1.2 Descripteur basé sur les histogrammes

## Chapitre II : traitement d'image

D'un point de vue général, l'histogramme représente une estimation de la distribution des intensités de l'image. Swain et Ballard utilisent en 1991 des histogrammes ayant pour objectif la reconnaissance d'objets. De nombreuses méthodes ont dès lors été proposées et il est possible de les classer en deux catégories, l'une s'appuyant sur des histogrammes d'intensités lumineuses et l'autre sur des histogrammes de gradients orientés

### A.5 Histogramme de gradients orientés (HOG)

Proposé en 2005 par Dala ET Triggs, les histogrammes de gradients orientés sont utilisés principalement en vision par ordinateur pour la détection d'objets. Leurs utilisations se sont également révélées particulièrement efficaces pour la détection de personnes. D'un point de vue général, un gradient permet de calculer les variations d'une fonction par rapport aux changements de ses paramètres. Dans le cas d'une image, la détermination du gradient consiste à calculer la variation de l'intensité des pixels dans différentes directions. En effectuant une telle analyse sur une image entière ou sur un voisinage, nous construisons un histogramme de gradients orientés permettant d'étudier les orientations des gradients locaux. L'idée directrice d'un HOG est que l'apparence et la forme d'un objet dans une image peuvent être décrites par la répartition de l'intensité du gradient. L'analyse des différentes méthodes existantes permet de différencier deux types d'architecture : une construction à base carrée appelée R-HOG et une autre à base circulaire appelée C-HOG.



Figure II.23 : Exemple de gradients d'une image

### V.5.2 Descripteur de caractéristique global

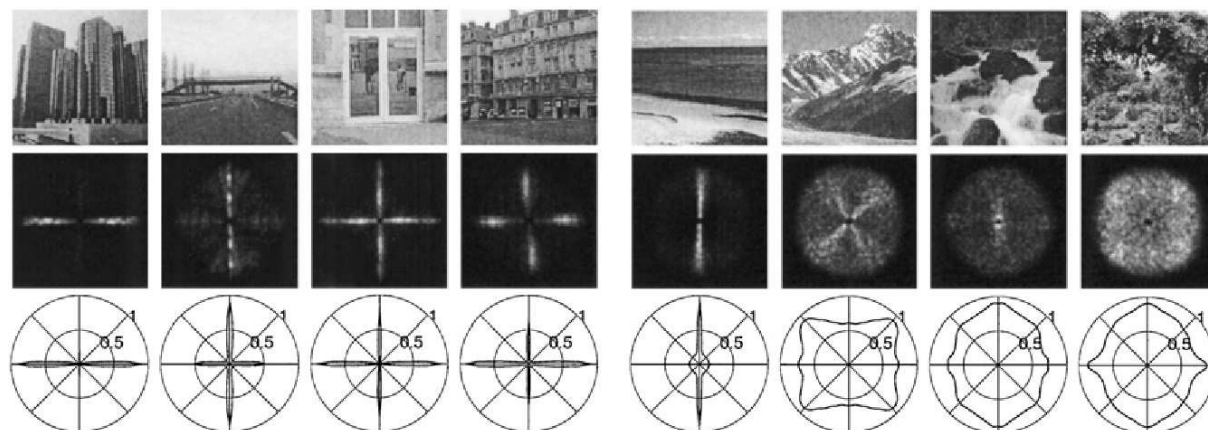
#### V.5.2.1 Le descripteur GIST

Initialement proposé par Oliva et A. Torralba, le but du descripteur GIST était à l'origine de permettre la classification automatique d'images. Partant du constat que l'être humain est capable de classer une image en très peu de temps et quelque soit sa complexité, l'idée était de recréer un système capable d'effectuer la même opération. Il s'agissait alors de capturer l'information essentielle contenue dans l'image. Le GIST est un descripteur global permettant d'extraire l'enveloppe spatiale de l'image qui contient l'essentiel de l'information. L'enveloppe spatiale correspond aux différentes fréquences et orientations contenues dans l'image. En dégradant une image en ne conservant que les fréquences et orientations principales, il est toujours possible d'identifier les paramètres principaux de l'image et de la classer. Étant donné qu'il s'agit d'un descripteur global capturant l'essentiel de l'information, ces paramètres ne peuvent pas concerner les détails de l'image. Les paramètres de classification ont été définis dans les travaux sur le GIST comme étant la possibilité de distinguer un environnement naturel d'un environnement urbain, le degré d'ouverture de l'environnement (par exemple une plage ou une montagne), la rugosité, le caractère accidenté d'un environnement naturel et le degré d'expansion pour un environnement urbain. Le classificateur complet, basé sur une SVM, a été entraîné suivant ces paramètres sur une base de données d'entraînement. La classe d'appartenance d'une image s'obtient simplement en fournissant son GIST au classificateur.

Montrant de bonnes performances dans la classification automatique, le descripteur GIST a été rapidement utilisé pour d'autres tâches. Le GIST est utilisé en tant que descripteur de structure de

## Chapitre II : traitement d'image

l'environnement, dénommée enveloppe spatiale dans les travaux de Oliva et A. Torralba.. L'information essentielle de l'image, est contenue dans les fréquences et orientations de celle-ci. Chercher les fréquences contenues dans une image implique le calcul de sa TFD 2D .Les images visuellement proches possèdent des spectres similaires correspond aux spectres des images.



**Figure II.24: L'extraction de l'enveloppe spatiale d'image avec le descripteur GIST**

Les fréquences et orientations les plus importantes sont facilement déterminables à partir du spectre des images. Le principe du GIST est alors de récupérer ces fréquences et orientations principales. Pour cela, plusieurs filtres de Gabor sont utilisés. La banque de filtres utilisée regroupe plusieurs filtres de Gabor chacun défini par une fréquence centrale et une orientation. Le nombre de filtres est paramétrable en fonction du nombre de fréquences centrales et d'orientations désirées. En pratique, le choix des filtres utilisés est déterminé de manière à ne pas avoir de recouvrement des spectres de chacun des filtres dans les zones de maximum d'amplitude ; il y a seulement recouvrement dans les zones de fortes atténuations. Les fréquences centrales  $F_0$  sont alors automatiquement déterminées en fonction de l'étalement de la Gaussienne dans le domaine fréquentiel et du nombre de filtres souhaités afin de limiter le recouvrement. De même, les orientations sont déterminées de manière à limiter le recouvrement et de façon à orienter le filtre suivant l'axe origine-centre du filtre

Pour appliquer les filtres de Gabor à l'image, cela se fait simplement par multiplication du module du spectre du filtre de Gabor avec le module du spectre de l'image le filtrage correspond à une multiplication dans le domaine fréquentiel. Le résultat obtenu est donc une image de la taille de l'image d'origine représentant le spectre filtré par le filtre de Gabor. Il en résulte alors autant de spectres que de filtres. Pour une image en noir et blanc, le résultat correspond au filtrage direct de l'image par la banque de filtres. Pour les images en couleurs, chaque canal est traité indépendamment ; le résultat correspond à la concaténation du filtrage de chacun des canaux. Le descripteur GIST final est la concaténation de tous ces résultats de filtrage dans un seul vecteur

Le descripteur résultant est de grande dimension mais décrit globalement l'image, il s'agit de la concaténation de l'information essentielle de l'image permettant sa classification efficace suivant les paramètres définis précédemment. Ce descripteur décrit alors la structure de l'environnement.

Il est important de savoir que l'image est pré-traitée. En effet, le spectre n'est pas calculé sur l'image telle qu'elle.

## VI. Conclusion

Dans ce chapitre, nous avons essayé de faire un récapitulatif sur les notions élémentaires de traitement d'image.

nous avons porté également la notion de caractéristique, l'extraction (détection et description des caractéristiques globales et locales). Nous avons cité les détecteurs et les descripteurs les plus connus et utilisés tels que Harris, bof et GIST.

# chapitre III

*Les systèmes de reconnaissance de  
plaque d'immatriculation*

### I. Introduction

La lecture automatique de plaques minéralogiques ou reconnaissance automatisée de plaques d'immatriculation (RAPI) est une méthode de surveillance de masse qui utilise les techniques de traitement d'image et de la reconnaissance optique de caractères sur des images pour lire les plaques d'immatriculation de véhicules. Dans leur version la plus simple, les systèmes de RAPI utilisent, en cascade, deux techniques de traitements d'images : la détection d'objet pour repérer les plaques d'immatriculation des véhicules qui se présentent, par exemple, à l'entrée d'un parking , à partir d'une image fournie par une caméras suivie de la reconnaissance optique de caractères afin d'identifier les caractères alphanumériques de la plaque. Pour obtenir des résultats satisfaisants, ce sont des séries de techniques de traitement d'image qui sont appliquées afin de détecter, normaliser et agrandir l'image de la plaque d'immatriculation,

Les phases de la reconnaissance sont généralement réalisées en appliquant des algorithmes de reconnaissance de forme et de caractères, qui se basent sur des descripteurs idoines de la plaque ou sur la comparaison de modèles ,l'étude décrite dans ce chapitre est dédiée justement à cette dernière .

### II. Présentation de système de reconnaissance automatique des plaques d'immatriculation(RAPI)

#### II.1 Définition d'une plaque d'immatriculation

Une plaque d'immatriculation ou plaque minéralogique est une [plaque](#) portant une combinaison unique de chiffres ou de lettres (pour une zone géographique donnée), destinée à identifier facilement un [véhicule](#) terrestre (automobile, moto, véhicule agricole, etc.). Dans le cas des véhicules circulant sur le réseau routier, ceci permet de lutter contre les infractions au [code de la route](#) et contre le [vol](#).

#### II.2 système de reconnaissance automatique des plaques d'immatriculation(RAPI)

Un système de reconnaissance automatique des plaques d'immatriculation (ou les RAPI )est une technologie qui trouve son essence dans ces 20 dernières années dans le développement des techniques de traitement d'image ainsi que dans les OCR (reconnaissance optique des caractères). Généralement un système RAPI typique est devisé en quatre phases :



Figure III.1 :Etapes d'un RAPI.

## Chapitre III : les systèmes de reconnaissance de plaque d'immatriculation

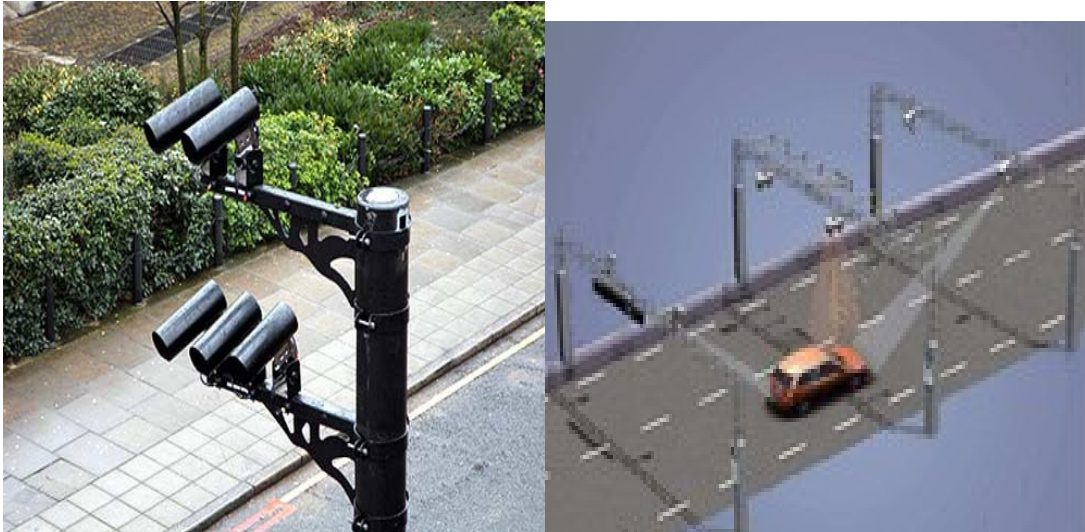
- L'acquisition de l'image à partir d'une séquence vidéo et son envoi vers le système.
- La détection et l'isolement de la plaque : c'est la phase la plus importante et la plus difficile ; elle détermine la rapidité et la robustesse du système. Plusieurs études et recherches sont consacrées à cette phase depuis plusieurs années ainsi différentes méthodes ont été proposées.
- La segmentation de la plaque en caractères : la plaque une fois extraite subira un ensemble de traitements pour être segmenté en séparant les caractères.
- La reconnaissance optique des caractères :

Ces systèmes ont été développés il y'a de ça une vingtaine d'années et leurs performance dépend des conditions environnantes et doivent être capables de gérer un grand nombre de difficultés possibles

### III. Domaines d'utilisation des RAPI

Les systèmes à base de capteur vidéo sont utilisés aujourd'hui dans de nombreuses applications. On peut citer les plus connues sans toutefois garantir l'exhaustivité de cette liste :

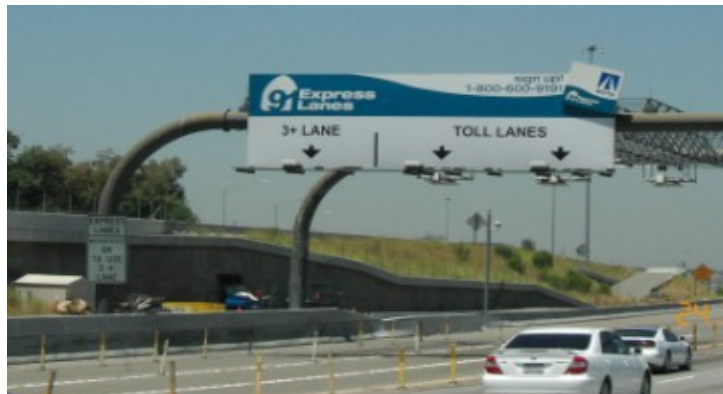
- **Vol de voitures** : le système est déployé sur le bord des routes, et réalise une comparaison en temps réel entre les voitures qui passent et la liste des voitures volées. Lorsqu'une correspondance est trouvée, une alerte est déclenchée pour informer l'agent de police de la voiture détectée et les raisons pour arrêter la voiture.
- **Parking** : le numéro de la plaque d'immatriculation est utilisé pour le paiement du stationnement au parking pour les gens ayant des cartes près-payée pour les parkings, afin de calculer les frais de stationnement en comparant les temps d'entrée et de sortie au parking.
- **Péage urbain** : le numéro de la plaque d'immatriculation est utilisé pour calculer les frais de voyage dans une route à péage, ou utilisé pour vérifier le billet. A l'étranger le système du péage urbain fonctionne grâce à l'analyse des plaques d'immatriculation enregistrées par un réseau de caméras de surveillance disposées en entrée et sortie d'une zone de péage . Il compare ensuite les données recueillies à une base de données centralisée pour identifier le propriétaire du véhicule et vérifier dans les délais son paiement effectif. En Suède , Les équipements étaient installés sur trois portiques , le premier portique indiquait par panneau l'entrée du point de contrôle et le prix à payer ainsi que des caméras pour l'enregistrement des plaques arrières du véhicules ;tandis que le deuxième portique était équipé de récepteurs permettant la lecture des badges à bord du véhicule ; le troisième portique équipé de caméras enregistrait les plaques avant du véhicule.



**Figure III.2 : Gestion des véhicules prioritaires**

Les Etats Unis ont aménagé sur leurs infrastructures routières des voies pour bus et co voiturage appelée « High Occupancy Vehicle » (HOV) ainsi que des voies appelées « High Occupancy Toll » (HOT) autorisant l'accès sur la voie de véhicules avec un seul passager à bord mais s'acquittant d'un péage modulé en fonction de la fréquentation de cette voie.

En Californie sur la SR 91, le dispositif permettant de contrôler l'accès à cette voie spécialisée est composé d'un portique équipé d'une antenne de lecture RFID et de 3 caméras dont une utilisée en lecture automatique des plaques d'immatriculation. De plus, sur le trajet de 17 km de long, plusieurs dizaines de caméras sont également utilisées pour la détection d'incidents.



**Figure III.3 : Péage de la SR91, Californie**

- **Control d'accès** : l'ouverture automatique d'une porte pour les membres agréés dans une zone de sécurité. Ce genre de système est mis en place pour aider les agents de sécurité. Les événements sont enregistrés sur une base de données et peuvent être utilisés pour rechercher l'historique des événements en cas de besoin.
- **Control des frontières** : le numéro de la voiture est enregistré à l'entrée ou à la sortie du pays, et utilisé pour surveiller les passages frontaliers. Chaque véhicule est enregistré dans une base de données centrale et lié à des informations supplémentaires telle que les données relatives aux passeports. Il est utilisé pour suivre tous les passages frontaliers.
- **Code pénal de la route** : le numéro de plaque est utilisé pour produire une amende de violation de vitesse ou de feux rouges. Le processus manuel de préparation d'une amende de violation est remplacé par un processus automatisé qui réduit les surcharges et les délais. Les amendes peuvent être consultées et payées en ligne.

On cite aussi :

### Chapitre III : les systèmes de reconnaissance de plaque d'immatriculation

- les stations-service (enregistrement quand un client part sans payer) .
- un outil de [marketing](#) pour enregistrer les modes de consommation.
- comparer les plaques d'immatriculations au [Système d'information Schengen](#).
  
- Reconnaissance Adaptative de la Hongrie (Budapest Hongrie) pour le contrôle de trafic.
- Asia Vision Technology Ltd (Hong Kong).
- Les technologies AutoVu Inc (Amérique du nord).
- Systèmes de reconnaissance Zamir Ltd (Jérusalem Israël).
- Mesure de la longueur de queue d'une congestion
- Mesure inter-véhiculaires.
- distance trop proche.
- Détection de contre-sens.
- Contrôle des voies spécialisées (bus, TC, taxis ...).
- Détection des vulnérables : 2RM, vélos et piétons.
- Mesure de la densité du trafic (uniquement avec la vidéo).
- Détection Automatique d'Incidents (DAI).
- Détection Automatique de Bouchons (DAB).
- Détection Transport de Matières Dangereuses (TMD).

#### **IV. Caractéristiques des Plaques d'immatriculation de Véhicule algérien**

Nous allons dans un premier lieu avant de proposer les méthodes de reconnaissance de plaques d'immatriculation, énumérer les caractéristiques des plaques d'immatriculation de véhicule algérien

- deux couleurs de la plaque sont disponibles . Certaines plaques sont blanches, et d'autres sont jaunes ,tandis que la couleur des caractères est noire.
- La plaque peut avoir une forme rectangulaire ou carrée, selon le type de véhicule (la plaque dans une moto par exemple est carrée) et la répartition des chiffres sur la plaque.
- La plaque est composée de dix chiffres divisé en trois parties, chacune a une correspondance (comme illustré dans la figure 3.4).

**On commence par la droite:**

**-15 correspond à la wilaya**

**-14 correspond à l'année de fabrication du véhicule**

**-1 correspond au type de véhicule.**  
**(1 est pour les voiture, 2 pour un camion**  
**3 pour les express ...)**

**-05430 signifie que cette voiture est la 5430ème qui**  
**est entré à la wilaya 15 (tizi ousou) en 2014**

**Figure III.4 :la plaque d'immatriculation algérienne**

## **V. Quelques méthodes existantes pour la reconnaissance des plaques d'immatriculation**

La reconnaissance d'objets dans les images représente une tâche très difficile dans la vision par ordinateur. La tâche de Reconnaissance des plaques d'Immatriculation des Véhicules (RPV) représente un grand défi des images de véhicule, à cause :

- des différents formats des plaques,
- les différents styles,
- la prise de vue
- la variation des conditions d'éclairage lors de l'acquisition des images de véhicules

Les différents systèmes RPIV proposés sont basés sur des propriétés différentes. Certaines techniques utilisent des méthodes basées sur l'information globale d'image tandis que d'autres utilisent des méthodes basées sur l'information locale.

### **V.1 Utilisation des informations globales**

#### **V.1.1 méthode basée sur la CCA(Connected Component Analysis )**

CCA est une technique importante appliquée dans le traitement de l'image binaire. elle est largement utilisée dans les problèmes de reconnaissance de texte, en particulier, dans la reconnaissance de plaques d'immatriculation. CCA numérise une image binaire et donne à chaque pixel une étiquette basée sur la connectivité de pixels, chaque étiquette symbolisant un composant. les pixels qui partagent la même étiquette font un composant connecté. les mesures spatiales, comme la superficie et l'aspect ratio, sont couramment utilisés pour la détection de la plaque d'immatriculation. Cette méthode a été utilisée dans plusieurs ouvrages.

## Chapitre III : les systèmes de reconnaissance de plaque d'immatriculation

Dans [1] par exemple, un système d'extraction de caractères à partir de plaques d'immatriculation a été présenté .Le système peut extraire avec succès des caractères de plaques d'immatriculation dans une partie quelconque d'une image capturée sans prendre en considération la luminance de l'environnement ou la taille et de l'inclinaison de la plaque d'immatriculation ou de la couleur du véhicule . Par ailleurs, le temps de traitement de ce système est satisfaisant pour plusieurs applications. Le système proposé ne dépend que de l' exhaustivité de un caractère et est donc tolérant des plaques inclinées et milieux complexes . Dans un parking réel, un taux de 95,6 % d'extraction a été obtenu en appliquant le système à 228 véhicules.

Ce travail présente un algorithme d'étiquetage de composant connecté . La structure de données de l'algorithme est dite liaison équivalente. L'avantage de la structure de données est qu'elle ne nécessite peu ou pas plus de mémoire pour le stockage que ne l'est requis dans les travaux antérieurs. L'algorithme explore une région adjacente à chaque pixel une seule fois. Les opérations de fusion et de recherche sont plus efficaces que d'autres.

Le système décrit dans ce travail passe par les étapes suivantes :

- Une étape de prétraitement.
- Une étape d'étiquetage des composants connectés et modifiés.
- Une étape de détermination des composants des caractères

Dans [ 2] , les auteurs présentent une approche distinguée par chiffre de reconnaissance effectuée par un réseau neurone , formé avec une transformée de Radon de quatre angles de lettres et chiffres échantillonnés . L'algorithme proposé est en mesure de reconnaître correctement les caractères sur la plaque d'immatriculation avec une probabilité de 94,1 % . Le système a été testé sur 51 images contenant des plaques; le taux de reconnaissance de 357 chaînes de caractères est de 99,15 % .les résultats sont les suivants :

- Extraction de plaque d'immatriculation :100%
- Reconnaissance de plaque d'immatriculation :94.1%
- Reconnaissance de caractères : 99.1%

### V.1.2 méthode basée sur la PCA(L'Analyse en Composantes Principales )

Le système proposé dans [3] comporte les étapes suivantes:

Etape 1: Le fichier vidéo est converti en images isolées par un programme mmread matlab. L'image requête est filtrée en utilisant les filtres Wiener et médian

- Etape 2: La plaque est détectée en utilisant le détecteur Prewitt , normalisée en utilisant **PCA** et extraite en utilisant un histogramme vertical et un autre horizontal déterminés au par avant.
- Etape 3: L' image de la plaque déterminée en utilisant les étapes ci-dessus est convertie en image binaire . L'image binaire est segmentée en deux parties: une partie arabe et une partie anglais par un histogramme profil de projection horizontale . Ensuite, chaque partie est segmentée en caractères isolés par un histogramme de projection verticale

### Chapitre III : les systèmes de reconnaissance de plaque d'immatriculation

- Etape 4 : Le caractère segmenté est représenté par un vecteur caractéristique basée sur les caractéristiques **PCA** de valeurs Eigen de la matrice de coefficients d'ondelettes et les caractéristiques invariants de contour.
- Etape 5 : La phase de reconnaissance comporte deux étapes. La première étape est de former le réseau neurones (**RN**) sur le de vecteur de caractéristique pour les échantillons des caractères utilisés dans l'apprentissage pour obtenir les poids optimaux. La deuxième étape est de tester le système en utilisant un nouveau caractère d'entrée .

#### V.1.3 Utilisation des méthode d'extraction de caractéristiques **GLCM** (Gray Level Co-occurrence Matrix) et **EDMS**(edge direction matrix)

La méthode Matrice Co -occurrence au Niveau de gris ( **GLCM** ) est une méthode d'extraction de caractéristiques statistiques pour l'extraction de caractéristiques globales . Cette méthode a été appliquée dans divers extraction de caractéristiques de texture. **GLCM** est basée sur une matrice qui montre la distribution des occurrences dans l'image sélectionnée. Cette méthode implique une approche statistique avec une matrice de co -occurrence qui est en mesure de décrire des statistiques de deuxième ordre pour des images de texture.

Le problème le plus important de **EDMS** est le nombre de caractéristiques produit avec cette méthode qui est a seulement 18 caractéristiques et ne suffit pas à des fins d'extraction de caractéristiques et il provoque la réduction du taux de reconnaissance. Le but de cette recherche décrite dans[4] est d'améliorer le taux de reconnaissance de **EDMS** en le combinant avec une méthode d'extraction de caractéristiques globales afin d'augmenter le nombre de caractéristiques extraites et produire de meilleurs taux de reconnaissance . La méthode proposée est une combinaison de **GLCM** et **EDMS** avec et sans méthode de sélection de caractéristiques.

Dans cette recherche, l'explication globale des systèmes de reconnaissance de caractères optiques et de l'extraction de caractéristiques est donnée et le nombre de nombreuses techniques importantes et pertinentes d'extraction de caractéristiques sont discutées. La méthode d'extraction de caractéristiques proposée est expliquée et une méthode d'extraction de caractéristiques spatiale qui est le filtre de Robinson est introduit. Les résultats expérimentaux qui sont concernés à **GLCM**, **EDMS** et la méthode proposée avant et après la sélection des caractéristiques appliquée à la plaque d'immatriculation est représenté , la comparaison de la performance est proposée aussi. De plus, certaines expériences ont été réalisées afin d'indiquer les performances des techniques d'extraction de caractéristiques spatiales et globales pour la reconnaissance de caractère et d'objet. Les résultats de la méthode proposée comme une technique d'extraction de caractéristiques globales sont comparées avec le filtre Robinson comme une technique d'extraction de caractéristiques spatiales pour l'ensembles de données objets et caractères qui sont, respectivement, de grandes images binaires de forme et des images de plaque d'immatriculation automatiquement recadrées. Les résultats expérimentaux montrent que la méthode proposée avant et après la sélection des caractéristiques est plus performante que les autres méthodes avec 90,83% et un taux de précision de 92,5%, respectivement, avec **RN**(réseaux neurones) comme classificateur. Une étude comparative de l'application des méthodes d'extraction de caractéristiques spatiales et globales dans la reconnaissance des objets et des caractères est effectuée. A cet effet, le filtre de Robinson utilisé comme une méthode d'extraction de caractéristique spatiale appliquée à un ensemble de grandes images binaire

## Chapitre III : les systèmes de reconnaissance de plaque d'immatriculation

de forme pour la reconnaissance d'objets et un ensemble d'images de plaque d'immatriculation recadrées automatiquement pour la reconnaissance de caractères ainsi que la méthode proposée avant et après la sélection des caractéristiques comme une méthode d'extraction de caractéristiques globales. Les résultats expérimentaux montrent la meilleure performance de la méthode proposée comme une méthode d'extraction de caractéristiques globales pour des fins de reconnaissance d'objet avec un taux de précision de 92,5% avec RN .

### V.1.4 Méthodes basées sur les couleurs et les formes des plaques d'immatriculation

Dans [5] une nouvelle approche pour la détection de la plaque d'immatriculation , sur la base des caractéristiques de couleur dans des plaques d'immatriculation iraniennes utilisant un balayage vertical périodique ou un balayage de pente , a été présenté . La méthode proposée

détecte l'emplacement des plaques d'immatriculation en reconnaissant sa teinte et sa forme . La localisation est l'échelle et la rotation invariant .

Le système de localisation proposé détecte toute les plaque d'immatriculation candidates disponibles dans une scène . Après la détection de la plaque d'immatriculation candidates, il les extrait et transmet l'image à la phase de reconnaissance pour reconnaître les huit caractères dans chaque plaque d'immatriculation iranienne . Le système de détection proposé localise le minimum des candidats et détecte le candidat réel (à savoir, la plaque d'immatriculation réelle) directement dans 98,73 % des images

Après une détection réussie , la qualité de l'image de la plaque d'immatriculation est améliorée, en particulier dans les contours des caractères, en appliquant des filtres de traitement d'image, tels que l'égalisation d'histogramme et le filtre de Laplacien. Tous les composants de surplus sont éliminés par l'application de certaines règles et fonctions morphologiques. Les caractères dans la plaque d'immatriculation sont extraits par le composant connecté ou l'analyse de projection et transmis au classificateur hybride pour la reconnaissance. Le classificateur hybride comprend un arbre de décision, et un SVM.

Les images en couleur pour le système proposé peuvent être générées par

Un équipement de surveillance photographique normale; donc un équipement photographique professionnel, tels qu'une caméra infrarouge, n'est pas nécessaire. Les types de plaques d'immatriculation iraniennes, telles que les plaques d'immatriculation publique avec un fond jaune ou des plaques d'immatriculation de gouvernement avec fond rouge, ont été identifiés par un histogramme de teinte dans la phase de détection.

La méthode proposée a été largement analysé; De plus, elle a été entièrement mise en œuvre et a prouvée pour être pratique. La figure suivante montre l'interface graphique d'utilisateur d'un logiciel LAPi Iranien basée sur la détection proposée et les méthodes de reconnaissance.

L'application principale qui a été programmé via Java enregistre le numéro de la plaque, la vitesse du véhicule, les coordonnées de l'emplacement de la plaque d'immatriculation, le pourcentage de précision pour chaque caractère reconnu, la date, l'heure, et l'adresse et l'emplacement de l'image d'origine redimensionnée dans une base de données.

## V.2 Utilisation des informations locales

### V.2.1 La reconnaissance des caractéristique a laide la méthode surf

Dans [6], la méthode de SURF est utilisé pour la détection et de l'extraction de caractéristiques sur la base de la matrice hessienne. Pour l'utilisation de la Hsian, on utilise une fonction continue de deux

### Chapitre III : les systèmes de reconnaissance de plaque d'immatriculation

variables avec la valeur de la fonction en  $(x, y)$  est donnée par  $f(x, y)$ . Hessienne matrice,  $H$ , est désignée par la matrice des dérivés partiels de la fonction  $f$ .

$$H(f(x,y)) = \begin{bmatrix} \frac{\partial^2 n}{\partial x^2} & \frac{\partial^2 n}{\partial x \partial y} \\ \frac{\partial^2 n}{\partial x \partial y} & \frac{\partial^2 n}{\partial y^2} \end{bmatrix} \quad (III.1)$$

$$\text{Det}(H) = \frac{\partial^2 n}{\partial x^2} \frac{\partial^2 n}{\partial y^2} - \left( \frac{\partial^2 n}{\partial x \partial y} \right)^2 \quad (III.2)$$

Cette valeur discriminante est utilisée pour classer le minimum et le maximum de la fonction à l'aide du test de la dérivée du second ordre. le déterminant est le produit des valeurs propres de la Hessien nous pouvons classer les points selon le signe du résultat.

Pour faire de cette méthode un travail avec des images, la fonction  $f(x, y)$  a été remplacé par les intensités de pixel d'image et pour calculer la dérivée partielle de second ordre, nous avons utilisé la convolution avec des noyaux dérivés. Les noyaux Gaussien appropriés peuvent être construits en  $x, y$  et combinés en calculant quatre entrées de la matrice Hessien. Gaussienne permet de faire varier la quantité de lissage au cours de l'étape de convolution de telle sorte que le déterminant est calculé à différentes échelles. En outre, comme Gaussien est une fonction isotrope (symétrique à savoir circulairement), la convolution avec le noyau permet une rotation invariante. La matrice Hessien  $H$  est calculée, en fonction de l'espace à la fois  $x = (x, y)$  et de l'échelle. Après avoir trouvé les points d'intérêt, on doit trouver les valeurs aberrantes pour une détection précise.

Dans le processus de segmentation, on a segmenté les lettres dans les différentes parties et on les a stockées dans un dossier. Pour cela, des opérations morphologiques ont été utilisées telles que la dilatation, à proximité de trouver les composants connectés et supprimé toutes les composantes inférieures à une certaine valeur de seuil et tous les composants sous forme de caractères ont été extraits et sont ensuite introduits dans le module de reconnaissance de ANN.



**Figure III.5 :la segmentation des caractères en utilisant des opérations morphologiques**

Après que la segmentation est terminée et chaque caractère a été extrait individuellement vient l'étape de correspondance. Pour l'étape de correspondance, une base de données de 26 alphabets et nombres de taille 24x42 pixels contient plusieurs échantillons d'un caractère a été créée afin de former le réseau de neurones. Un réseau stable a été construit pour reconnaître le caractère de la plaque d'immatriculation par la formation d'échantillons de caractères de base de données. Cette formation peut prendre beaucoup de temps qui est aussi son inconvénient mais formé une fois que le réseau peut facilement être appliquée aux caractères identifier des plaques segmentées

## Chapitre III : les systèmes de reconnaissance de plaque d'immatriculation

la méthode a le meilleurs résultats par rapport aux méthodes utilisées précédentes avec un taux de détection de 95% dans l'ensemble selon la base de données

### V.2.2 Maximally Stable Extremal Region (MSER) :

MSER est un algorithme efficace qui permet la détection des plaques d'immatriculation en images. Dans [7], un système de détection roman qui est motivé par le travail de Matas et al. a été proposé. L'approche proposée ne nécessite aucun système d'apprentissage et est capable de détecter différents types de plaques d'immatriculation internationales sans adaptation. L'algorithme de détection est basé sur l'analyse des résultats d'un taux MSER). Les résultats de la détection MSER peuvent être utilisés pour détecter des plaques d'immatriculation. Dans ce travail, on a utilisé le principe suivant :

MSER + trouve la plaque d'immatriculation, alors que MSER- identifie l'individu caractères sur elle. L'idée sous-jacente de ce système est de détection de la plaque d'immatriculation roman est d'analyser les résultats de détection MSER + et MSER-. On recherche d'une plus grande région MSER + (plaque d'immatriculation) qui contient un ensemble de petites régions MSER- (caractères). les régions MSER- doivent approximativement être de même taille. Après vérification, le MSER + est retourné comme résultat la plaque d'immatriculation de localisation et en outre segmentations des caractères sont fournis par les détections de MSER- correspondantes Bien que le processus de détection est simple, il est efficace et permet une détection précise et stable de la plaque d'immatriculation dans les scènes complexes

Après détection, un tracker robuste est appliquée pour augmenter le nombre de caractères les représentations de l'étape de reconnaissance ultérieure. une version modifiée du suivi MSER a été introduit comme tracker qui présente certains avantages Contrairement à d'autres systèmes de repérage.

La segmentations des caractères individuels sur la plaque, suivi ainsi et la segmentation sont traitées simultanément. La première étape de l'algorithme est de propager le point central de la région à l'image suivante et de recadrer une région d'intérêt tout autour de l'image. Ensuite, un arborescence des composants de données de la structure est construit. Chaque nœud de l'arbre composant contient une région candidate pour le suivi et l'algorithme de recherche du nœud qui est le plus similaire à la région. Le meilleur ajustement est identifié en comparant les vecteurs caractéristiques qui sont construits pour chacun des nœuds de l'arborescence des composants et la région d'entrée. Le candidat, avec la plus petite distance euclidienne entre son vecteur de caractéristiques et celui de la région est pris comme représentation suivi.

Les caractéristiques calculées sont la valeur moyenne de gris, la taille de la région, le centre de masse, la largeur et la hauteur de la zone de délimitation et de stabilité. Toutes ces caractéristiques sont calculés lors de la création de l'arborescence des composants. Ainsi, aucun supplément de temps de calcul n'est nécessaire. Après la détection de la nouvelle représentation du décrit les étapes peuvent être répétées pour le suivi de la région à travers la totalité de la séquence.

Le suivi est considéré comme valide tant que la valeur de distance est minimale est inférieure à un seuil fixe. En répétant les étapes présentées la plaque d'immatriculation est suivi par la séquence et les régions détectées MSER- au sein de la plaque sont fournies à titre de segmentations des caractères individuels. En outre, le système de suivi est également utilisé pour rejeter les faux positifs de l'étape de détection en supprimant les pistes de plaques instables.

Pour la reconnaissance des caractères individuels sur les plaques détectées ils se sont basés sur des machines à vecteurs de support SVM (*Support Vector Machine*). Puisque la reconnaissance de caractères est un problème multi-classe, une méthode basée sur la combinaison de plusieurs SVMs binaires a été appliquée. La stratégie est appelée un-contre-un où pour chaque paire de classes de sortie d'un SVM individuel est formé, pour un total nombre  $n * (n - 1) / 2$  de classificateurs. Ensuite, tous les classificateurs sont évalués, les votes se résument et la classe avec le nombre maximum est choisie. D'abord, les segmentations de caractères fournis sont alignés, puis l'approche d'un-contre-un est utilisée pour classer chaque caractère indépendamment de tous les autres. L'approche de suivi présenté fournit plusieurs représentations de plaque d'immatriculation pour chaque véhicule. Un système de vote à la majorité est ensuite utilisé pour déterminer le résultat de la reconnaissance de caractères finale pour chaque véhicule.

Pour évaluer la qualité du travail proposé 109 voitures ont été utilisées pour la reconnaissance des plaques d'immatriculation. En raison de la faible résolution et l'évolution des conditions d'éclairage, le taux de reconnaissance moyenne indépendante classement de tous les caractères de chaque plaque d'immatriculation détectée est seulement 80.74, le pourcentage de classifications totalement correctes

## Chapitre III : les systèmes de reconnaissance de plaque d'immatriculation

atteint plus de 90%. En utilisant toutes les représentations disponibles pour reconnaître la plaque de chaque voiture (4880 plaques ce travail classe 94,65% de toutes les voitures totalement correctes qui est un résultat prometteur pour un tel ensemble de données difficile.

### V.2.3 basé sur SOFT (La méthode de Viola et Jones )

La méthode SOFT appelée aussi La méthode de Viola et Jones est une méthode de détection d'objet dans une image numérique, proposée par les chercheurs Paul Viola et Michael Jones en 2001. Elle fait partie des toutes premières méthodes capables de détecter efficacement et en temps réel des objets dans une image. Inventée à l'origine pour détecter des visages, elle peut également être utilisée pour détecter d'autres types d'objets comme des voitures ou des avions. La méthode de Viola et Jones est l'une des méthodes les plus connues et les plus utilisées, en particulier pour la détection de visages et la détection de personnes.

Le système dans [8] fonctionne sur la plupart des images génériques .une méthode de détection des bords Sobel et la méthode de Otsu pour le seuillage ont été proposées afin de localiser la plaque. L'étape de détection est suivie par la l'enlèvement de frontière , la suppression du bruit, l'extraction de caractère et la séparation et enfin le modèle correspondant à la reconnaissance des caractères. L'algorithme est expérimenté avec succès sur la variété des images réelles et représenté dans le document.

Une fois que le rectangle provisoire est sélectionné sur l'image, l'algorithme cherche pour 3 lignes continues de pixels blancs vers le haut et vers le bas à partir des limites inférieures et supérieures entre le contenu du rectangle . une procédure similaire est appliquée sur des colonnes pour trouver les limites à gauche et à droite pour les cultures. Ces régions de pixels blancs correspondent à un caractère historique . et quand ladite condition est remplie, le rectangle défini résultant par les sauts nouvellement trouvés est finalisé et la région correspondante est rognée .

Après l'étape de l'isolement et la suppression du bruit ,pour la détection de chaîne, la zone correspondante de l'image originale a dû être coupée. Le contraste élevé entre les caractères et le fond dans une plaque, est une caractéristique forte qui a été pris en compte dans la segmentation des caractères . Par conséquent, le contraste d'étirement a été appliqué sur la plaque extraite.

La sortie de l'image numérisée est complétée et à partir de celle-ci, la sélection des caractères est simple. Les composantes connexes de l'image binaire ont été marquées. Un algorithme adaptatif itérativement analysé pour la présence d'ordre le plus bas de composant connecté. En tenant compte de la première et la dernière ligne et la colonne de leur occurrence, la partie contenant le caractère alphanumérique a été recadrée sélectivement de l'image et affichée sur une interface utilisateur graphique. Le rognés pixels de la plaque ont été remplacés par des pixels noirs sur elle .

La reconnaissance a été effectuée grâce à l'enregistrement sur la base de la zone.

Les concepts de codage utilisés sont impartiaux envers tout notamment le format de compression d'image. JPEG, TIFF, BMP, etc. peut être utilisé. L'entrée du système peut être soit en couleur soit en niveaux de gris. Une base de données de 70 images de véhicules de la ville de Bhopal et Indore a été pris en considération et 95,32% été le taux de reconnaissance obtenu.

### V.3 Les méthodes globales vs les méthodes locales pour la RPIV

Le tableau 1 présente les résultats obtenus à cet effet pour chacune des méthodes décrites ci-dessus .

	Ref	Techniques utilisées			résultat s
		Détection de plaque	Segmentation de caractères	Reconnais sance de caractères	
I n f o r m	[2]	Détection des contour , opérations morphologiques ,	CCA	Les Réseaux de Neurones RN) avec	94.1%

### Chapitre III : les systèmes de reconnaissance de plaque d'immatriculation

a t t r i b u t s  g l o b a l e		CCA et Otsu		une transformé e Radon	
	[3]	Détection des contour avec Prewitt PCA	Histogramme de profil de projection horizontale et verticale	(RN),	97.6%
	[4]	EDMS GLCM	-----	RN	92.5%
[5]	Un algorithme modèle de correspondance	seuillage multi niveaux	SVM	94%	
I n f o r m a t i o n  L o c a l e	[6]	Détection avec SURF	des opérations morphologiques telles que la dilatation	RN	95%
	[7]	A Laide de MSER		SVM	94,65%
	[8]	La méthode de Viola et Jones (SOFT)		l'enregistrement sur la base de la zone	95,32%

**Tableau III.1 : tableau de comparaison des résultats des méthodes présentées**

L'évaluation de ces méthodes basée sur le taux de reconnaissance de plaques d'immatriculation dévoile le fait que les méthodes basées sur l'information locale d'une image sont plus performantes que celle basé sur l'information globale de celle-ci avec une moyenne de reconnaissance 94.99%.

## VI. Conclusion

### Chapitre III : les systèmes de reconnaissance de plaque d'immatriculation

Nous avons abordé dans ce chapitre le système de reconnaissance automatique des plaques d'immatriculation (RAPI) , ses étapes et son domaine d'utilisation .

nous avons recensé aussi quelques méthodes existantes pour la reconnaissance des plaque d'immatriculation en utilisant soit l'information locale d'image ou l'information globale de celle-ci. Une comparaison de ces méthodes basée sur le taux de reconnaissance a été effectué pour conclure la performance des méthodes basées sur l'information locale par rapport à celle basées sur l'information globale .

le prochain chapitre sera consacré à l'implémentation des méthodes proposées après avoir présenter le principe de chaque méthodes.

# chapitre IV

*Implémentation*

## Chapitre IV : Implémentation

### I. Introduction

Après avoir étudié les différentes méthodes de reconnaissance de plaque d'immatriculation soit celles basées sur les caractéristiques locales ou celles basées sur les caractéristiques globales, dans le chapitre précédent, et après avoir lancé notre hypothèse en se basant sur les résultats expérimentaux de ces méthodes, nous allons dans ce chapitre ratifier cette hypothèse par l'implémentation des méthodes que nous jugeons performantes pour cela nous avons mis en œuvre deux méthodes l'une basée sur MSER et l'autre basée sur GLCM et EDMS.

### II. Implémentation des systèmes de reconnaissance des plaques d'immatriculation de véhicules

#### II.1 La reconnaissance basée sur les caractéristiques locales (avec MSER)

##### II.1.1 Le principe de MSER

MSER est une méthode de détection de blob dans les images. L'algorithme MSER extrait à partir d'une image un nombre de régions co-variantes, appelées MSER: un MSER est un composant connecté stable de certains ensembles en niveaux de gris de l'image.

MSER est basé sur l'idée de prendre les régions qui restent à peu près la même à travers une large gamme de seuils.

- Tous les pixels en dessous d'un seuil donné sont blancs et tous ceux qui sont au dessus ou égale à ce seuil sont noirs.
- Si on a montré une séquence d'images  $I_t$  à qui on a effectué un seuillage avec une trame  $t$  correspondante au seuil  $t$ , on verrait d'abord une image noire, puis des taches blanches qui correspondent au minima d'intensité locale qui apparaîtront puis croîtront.
- Ces taches blanches finiront par être fusionnées, jusqu'à ce que l'ensemble de l'image est devenue blanche.
- l'ensemble de tous les composants connectés dans la séquence est l'ensemble de toutes les régions extrémales.

En option, les cadres elliptiques sont fixés aux MSERs en ajustant les ellipses aux régions. Ces descripteurs de régions sont gardés comme caractéristiques

Le mot extremal renvoie à la propriété que tous les pixels à l'intérieur du MSER ont soit l'intensité la plus supérieure (régions extrémales lumineuses) ou la plus inférieures (régions extrémales sombres) que tous les pixels sur sa frontière extérieure

Cette opération peut être effectuée en premier tri de tous les pixels par valeur de gris, puis en ajoutant par incrémentation les pixels à chaque composant connecté tant que le seuil est modifié. La zone est surveillée. Les régions ainsi que leur variation où le seuil est minimale sont définies stable au maximum:

- soit tous les pixels en dessous d'un seuil, d'une couleur blanche. Les autres, noirs.
- soit Une séquence d'images à qui on applique un seuillage avec des seuils de balayage du noir au blanc, nous passons d'une image noire à des images où des taches blanches apparaissent et « grossissent » en se fusionnant, jusqu'à l'image finale.

## Chapitre IV : Implémentation

- A travers une large gamme de seuils, la binarisation locale est stable et montre une certaine invariance à la transformation affine des intensités d'image et mise à l'échelle

Le MSER met en œuvre les étapes suivantes :

- balayer le seuil d'intensité du noir au blanc , effectuer un seuillage simple de luminance de l'image ,
- Extraire les composants connectés ( « Régions extrémales » )
- Trouver Un seuil quand une région extrémale est " Stable Au maximum " , à savoir minimum local de la croissance relative de son carré. En raison de la nature discrète de l' image dans la région au-dessous / au-dessus peut être confondue avec la région effective , dans le cas où la zone est toujours considérée maximale.
- Garder Ces descripteurs de régions comme caractéristiques

Cependant , même si une région extrémale est au maximum stable , il pourrait être rejetée si :

- il est trop grand ( il y a un paramètre MaxArea ) ;
- il est trop petit ( il y a un paramètre MinArea ) ;
- il est trop instable ( il y a un paramètre MaxVariation ) ;
- il est trop similaire à son parent MSER

### II .1.2 Multi - résolution MSER

Méthode:

- Etape 1: Au lieu de détecter les caractéristiques que dans l'image d'entrée , on construit une pyramide à l'échelle avec une octave entre les échelles.
- Etape 2: Détecter les MSERs séparément à chaque résolution.
- Etape 3: Les MSERs en double sont supprimés en éliminant les MSERs à échelle fine avec les mêmes emplacements et tailles que les MSERs détectés à l'échelle la plus grossière .

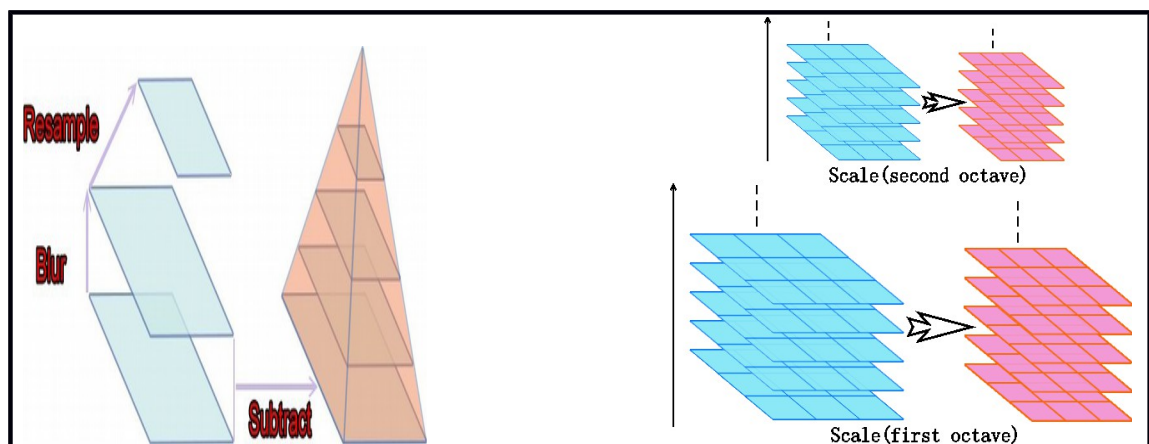


Figure IV.1 : Pyramide à l'échelle avec une octave entre les échelles.

## Chapitre IV : Implémentation

### II .1.3 Implémentation

Dans cette partie, nous allons décrire l'ensemble d'étapes à suivre pour la reconnaissance des plaques d'immatriculation . Le système mis en place détecte des plaques d'immatriculation en utilisant un algorithme qui est basée sur l'analyse des résultats de Détection MSER . Le concept, présenté ci-dessous , ne nécessite aucun système d'apprentissage et est capable de détecter les différents types de plaques internationales. Après la détection d'une plaque, une segmentation de celle-ci est requis et pour se faire un détecteur de MSER est utilisé encore une fois dans cette étape . l'étape qui suivra sera dédiée à la reconnaissance de caractères dans la plaque détectée.

#### II .1.3.1 Détection de plaque d'immatriculation

Le système de détection que nous proposons est motivé par le travail de Matas et al, qui ont présenté une méthode pour apprendre les régions extrémales d'une catégorie spécifique , plus exactement, les caractères sur une plaque d'immatriculation, pour effectuer la détection. Contrairement à leur approche le système que nous présentons ne nécessite aucun système d'apprentissage et est capable de détecter différents types de plaques d'immatriculation internationales sans adaptation.

L'algorithme de détection est basé sur l'analyse des résultats de détection MSER .les MSERs désignent un ensemble de régions distinguées et se sont révélés être l'un des meilleurs détecteurs de point d'intérêt en vision par ordinateur . Toutes ces régions sont définies par une propriété extrémales de la fonction d'intensité dans la région et sur sa limite extérieure. Des propriétés spéciales forment leurs performances supérieures comme détecteur local stable. L'ensemble des MSERs est fermé sous des transformations géométriques continues et est invariant à des changements d'intensité affines. En outre, MSERs sont détectés dans toutes les échelles. Nous exploitons principalement ces propriétés à des fins de segmentation.

En général, deux variantes de détection de MSER peuvent être distinguées, désignées comme MSER + et MSER-. Alors que MSER + détecte les régions lumineuses avec limite plus sombre, MSER- trouve des zones sombres avec limite plus lumineuse. La figure 1 (a) représente une image à partir d'une séquence vidéo de la circulation et la figure IV.2 (b) et la figure 1 (c) illustrent les résultats de détection de MSER correspondants sous forme d'images binaires. Comme on peut le voir, la la plaque d'immatriculation elle-même est identifiée comme MSER +, tandis que les caractères sur la plaque sont détectés comme MSER-.



Figure IV.2 : Les résultats de la détection MSER

## Chapitre IV : Implémentation

L'idée sous-jacente de notre système de détection de la plaque d'immatriculation est d'analyser les résultats de détection MSER+ et MSER-. Nous sommes à la recherche d'une plus grande région MSER+ (plaque d'immatriculation) qui contient un ensemble de petites régions MSER- (caractères).

Une telle combinaison est considérée comme résultat de détection de plaque d'immatriculation. En outre, on vérifie la détection en vérifiant si les régions MSER- sont approximativement de même taille, si leur point de centre s'étend approximativement sur une ligne et si la hauteur de MSER+ est dans la gamme de la hauteur moyenne de MSER-. Après vérification, le MSER+ est retourné comme résultat de localisation de la plaque d'immatriculation et en outre les segmentations des caractères sont fournies par les détections de MSER- correspondantes.

Bien que le processus de détection est simple, il est efficace et permet une détection stable et précise de la plaque d'immatriculation dans les scènes complexes. Un exemple de résultat est représenté sur la figure IV.3.



Figure IV.3 : résultats de détection de la plaque d'immatriculation

### II .1.3.2 Segmentation de la plaque d'immatriculation

La plaque d'immatriculation détectée, est donnée à un détecteur MSER pour exploiter des caractères possibles. Le MSER a été initialement proposé pour trouver les correspondances à travers les points de vue. Il peut rendre les bordures persistantes autour des objets comme des changements d'éclairage, il a prouvé son efficacité dans la segmentation des caractères qui révèlent souvent que les bords sont robustes pour les variations d'éclairage. L'extraction de MSER considère l'ensemble de tous les seuils possibles qui sont capables de binariser l'image d'intensité  $I(x)$  en image binaire  $E_{t_M}(x)$  comme suit:

$$E_{t_M}(x) = \begin{cases} 1 & \text{si } I(x) \leq t_M \\ 0 & \text{sinon} \end{cases} \dots\dots\dots (IV.1)$$

## Chapitre IV : Implémentation

Tous les pixels en dessous d'un seuil donné sont «noir» et tous ceux au dessus ou égale sont 'blanc" .Les taches "noires" correspondant au minima d'intensité locale apparaîtront puis croîtront . L'image entière va devenir noire en fusionnant tous les taches "noires». Le gradient de Sobel et la détection du contour d'images sont réalisées pour la segmentation de la plaque.

### Algorithme de la Segmentation

1. Entrée: la plaque d'immatriculation détectée ( de la partie détection )
2. Binariser I a une  $E_t$  avec un seuil variant d'un minimum au maximum M.
3. Trouvez tous les bords de l'image extraite en utilisant le gradient sobel .
4. Déterminer la zone de délimitation en utilisant la région de propriétés.
5. Sortie : régions segmentées pour la reconnaissance de caractères .

### II .1.3.3 RECONNAISSANCE DE CARACTÈRES

La reconnaissance de caractères est la traduction mécanique ou électronique des images de texte manuscrit ou de dactylographié en un texte modifiable avec la machine. La normalisation est d'affiner les caractères dans un bloc qui ne contient pas d'espaces blancs supplémentaires (pixels) dans tous les quatre côtés des caractères. Les caractères extraits coupés de la plaque et les caractères de la base de données sont maintenant de taille égale .Les caractères segmentés ont beaucoup de variation dans la taille. Dans cette phase , tous les caractères sont normalisées à la hauteur prédéfinie ( longueur verticale ) en pixels . Étant donné que les caractères ont toujours une largeur variable (longueur horizontale ), chaque caractère est normalisé à une taille de 42 x 24 . Les caractères extraits coupés de la plaque et les caractères de la base de données sont maintenant de taille égale et la prochaine étape est le template matching (appariement modèle) . Le template matching est un algorithme efficace pour la reconnaissance de caractères . Les meilleures similitudes entre les caractères sont reconnues en comparant les caractères de la base de données déjà stockées dans un modèle.

### Algorithme de reconnaissance de caractères

1. Entrée: Un caractère segmenté I est stocké dans une matrice 3 × 3 .
2. Créez des fichiers de modèles ( 0-9 )
3. Comparer avec les caractères dans la bibliothèque de modèles .
4. Obtenir le maximum de pertinence du modèle correspondant .
5. Sortie de la valeur de modèle correspondant

La présence d'un seul pixel fournit des informations sur la correspondance exacte d'une image de modèle avec une image d'entrée segmentée . Une fois le processus d'adaptation de modèle est complété un numéro de plaque particulier peut être correctement reconnu . A

## Chapitre IV : Implémentation

partir de cette information de numéro les détails sur un véhicule peuvent être récupérés à partir de la base de données .

### II .1.3.4 Résultats et caractérisation

Nous avons appliqué ce système sur les 222 exemples de plaques d'immatriculation réelles. Il identifie correctement 93.75 % des exemples; ce qui correspond au taux de reconnaissance en micro-précision.



Figure IV.4 : Quelques exemples de plaques d'immatriculation utilisées dans le système proposé

- Le temps d'exécution

Les temps d'exécution des différentes étapes du système sont présentées dans le tableau suivant :

Etapes	Détection	Segmentation	reconnaissance
Temps d'exécution	70 ms	5 ms	6 ms

Tableau IV.1 temps d'exécution par image

- La performance

Etapes	Détection	Segmentation	reconnaissance
Performance	95.2%	97%	93.75%

Tableau IV.2 : la performance par étape

## II.2 Reconnaissance des plaques d'immatriculation avec GLCM et EDMS (approche global)

### II.2.1 La méthode GLCM :

## Chapitre IV : Implémentation

GLCM (Gray level Dependency Matrix) est une matrice au niveau de gris elle est définie de deux histogramme tridimensionnel pour une paire de pixels, qui sont séparés par une relation spatiale fixe. GLCM d'une image est calculée en utilisant un vecteur de déplacement D, défini par son rayon  $\delta$  et de l'orientation  $\theta$ .

### Choix d'un rayon $\delta$

Les applications de grande valeur de déplacement à une texture fine donnerait un GLCM qui ne tient pas compte des informations de texture détaillée. D'après les études antérieures, il a été conclu que la précision globale de classification avec  $\delta = 1, 2, 4, 8$  sont acceptables avec les meilleurs résultats pour  $\delta = 1$  et 2. Cette conclusion est fondée, en tant que le pixel est plus susceptible d'être corrélée a un autre pixel situé près que celui situé loin.

### Choix de l'angle $\theta$

Chaque pixel a huit pixels voisins cela permet le choix entre huit  $\theta$ , qui sont  $0^\circ, 45^\circ, 90^\circ, 135^\circ, 180^\circ, 225^\circ, 270^\circ$  ou  $315^\circ$ , les paires de la matrice Co\_occurrences obtenues en choisissant  $\theta$  égal à  $0^\circ$  seraient similaires à ceux qui obtenus en choisissant  $\theta$  égal à  $180^\circ$ . Ce concept se prolonge à  $45^\circ, 90^\circ$  et  $135^\circ$ . Par conséquent, on a quatre choix pour sélectionner la valeur de  $\theta$ .

## II.2.2 Implémentation

### II.2.2.1 Détection de la plaque d'immatriculation

#### A. Création de la matrice Co-occurrence :

GLCM est une méthode d'extraction de caractéristiques statistiques pour l'extraction mondiale de fonctionnalité. Cette méthode a été appliquée dans divers extraction de caractéristiques de texture. Elle est basée sur une matrice qui montre la distribution des occurrences dans l'image sélectionnée. Cette méthode implique une approche statistique avec une matrice de co-occurrence qui est en mesure de décrire un deuxième ordre de statistiques pour la texture d'image. (GLCM) implique un histogramme à deux dimensions dans I et J indique la fréquence de i qui se produit avec l'événement j. P (i, j, d,  $\theta$ ) indique la fréquence de la matrice co-occurrence et d représente la distance d'une paire de pixels. La direction est spécifiée par  $\theta$  qui peut avoir la valeur  $0^\circ, 45^\circ, 90^\circ$ , et  $135^\circ$ . La zone de la texture est considérée comme rectangulaire et elle a  $N_c$  pixels dans la direction horizontale et  $N_r$  dans la direction verticale. Dans chaque Résolution de cellule la tonalité est définie par des niveaux Ng. Le domaine horizontal spatial est défini par  $L_c = \{1,2, \dots, N_c\}$  et domaine vertical spatial est défini par  $L_r = \{1,2, \dots, N_r\}$  et l'ensemble de Ng quantifiant la tonalité de gris est défini par  $G = \{1,2, \dots, Ng\}$ . L'ensemble des  $(L_r \times L_c)$  est définie comme l'ensemble des cellules de résolution de l'image qui sont classés par leur description ligne-colonne. La représentation des fréquences non normalisées avec des angles différents sont les suivants:

$$P(i, j, d, 0) = \{((k, l), (m, n)) \in (L_r * L_c) * (L_r * L_c), |I - n| = |d, I(K, l) = i, I(m, n) = j\}, Km = 0$$

.....(IV.2)

$$(K, l), (m, n) \in (L_r * L_c) * (L_r * L_c) (K - m = d), |n| = (-d) \vee (K - m) - d, |I - n| = d, I(K, l) = i, I(m, n) = j$$

$$P(i, j, d, 45) = \{$$

.....(IV.3)

$$P(I, j, d, 90) = \{((K, l), (m, n)) \in (L_r * L_c) * (L_r * L_c), \forall I - n \vee i, 0, I(k, l) = I(m, n) = j\}, K = d$$

.....(IV.4)

## Chapitre IV : Implémentation

$$P(I, j, d, 135) = \sum_{\{(K,l), (m,n)\} \in (L_r * L_c) * (L_r * L_c), (K-m=d), |I-n|=d} \dots \dots \dots (IV.5)$$

on appliquant les équations de IV.2 jusqu'à IV.5, la matrice Co-occurrence est construite ce qui explique l'apparition de toute les paire de pixels qui sont l'échelle de gris.

### B. Extraction de caractéristique

Le nombre total de caractéristiques qui est extrait en basant sur la matrice de co-occurrence est 36. Cette méthode est l'une des méthodes les plus courantes dans l'extraction mondiale de fonctionnalité tout en produisant des caractéristiques sans distinction qui réduisent le taux de reconnaissance. Une autre technique globale d'extraction de caractéristiques EDMS le problème qu'on produit avec cette méthode est que seulement 18 caractéristiques produit et cela ne suffit pas à des fins d'extraction de caractéristiques et il provoque la réduction du taux de reconnaissance, mais comparé à d'autres méthodes d'extraction de caractéristiques existantes, il produit un taux de reconnaissance considérable. Une matrice qui implique huit noyaux de quartier a été appliqué afin de produire la matrice de direction Edge. A noter que la deuxième valeur de pixel (EDM1 (2,2)) est égal au nombre de pixels noirs dans l'image de bord et cette valeur est portée définie en tant que second pixel pour EDM2 (2,2). La première matrice EDM1 chaque cellule contient une position basée sur pixel association de quartier et la position est 0-315 degré. En calculant les valeurs THE- occurrence

$EDM_1$ , la relation entre les valeurs de pixels peuvent être déterminées sous-image de bord. La

deuxième matrice d'ordre qui est  $3 \times 3$  considérée comme une direction matrice contient la présentation de la relation de chaque pixel. En cas d'avoir plus d'un angle avec le même numéro d'occurrence le plus petit angle est choisi tout d'abord et les plus petit pixel sont choisi respectivement. On appliquant ces deux matrices, différentes fonctions peuvent être extraits tels que la direction du bord, l'homogénéité avec des angles 0, 45, 90, 135, poids, pixel régularité et bord de régularité avec des angles 0, 45, 90, 135, 180, 225, 270 et 315.

#### Algorithme

Creation de la matrice co-occurrence :

`glcms = graycomatrix(I)` crée une matrice de co-occurrence niveau de gris (GLCM) à partir de l' image I

`glcms = graycomatrix(I,Name,Value,...)`

Retourne une ou plusieurs matrices de co-occurrence de niveaux de gris, en fonction des valeurs paires facultatives

`[glcms,SI] = graycomatrix(____)`envoie l'image à l' échelle, SI utilisé pour calculer la matrice de co-occurrence de niveaux de gris. Les valeursSI sont comprises entre 1et NumLevels.

### II.2.2.2 Segmentation de la plaque d'immatriculation

techniques d'extraction de caractéristiques spatiales extraire couramment les informations spatiales telles que la texture, la taille, la forme, l'orientation, la position et ainsi de suite. Ce type de dispositif d'extraction considère la relation spatiale entre les pixels et est basé sur une fenêtre ou le noyau [23] en mouvement. Robinson masque de la boussole est l'un des filtres de détection de bord de gradient qui détermine la direction d'un pixel en appliquant le gabarit qui correspond à la meilleure zone locale du pixel et peut être considérée comme une technique d'extraction de caractéristiques spatiales [24]. Ces opérateurs se compose de huit noyaux pour chaque pixel qui a produit par la rotation d'un noyau sur la base de ses coefficients circulairement., Les coefficients de filtre de Robinson sont constitués de

## Chapitre IV : Implémentation

0, 1 et 2, et tandis que les zéros correspond à la direction de la ligne. masque Robinson est défini en prenant un seul noyau et le tournant vers les huit directions qui sont du Nord, du Nord-Ouest, Ouest, Sud-Ouest, du Sud, du Sud-Est, l'Est et du Nord-Est, respectivement:

$$\begin{aligned}
 N &= \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} &
 W &= \begin{bmatrix} 0 & 1 & 2 \\ -1 & 0 & 1 \\ -2 & -1 & 0 \end{bmatrix} &
 S &= \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} &
 E &= \begin{bmatrix} 2 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & -1 & -2 \end{bmatrix} \\
 NW &= \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} &
 SW &= \begin{bmatrix} 0 & -1 & -2 \\ 1 & 0 & -1 \\ 2 & 1 & 0 \end{bmatrix} &
 SE &= \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} &
 NE &= \begin{bmatrix} -2 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix}
 \end{aligned}$$

**Figure IV.5: Masque Robinson de huit directions**

La figure 1 montre le masque Robinson de huit directions. Si GS, GSE, GE, GNE, GN, GNW, GW, et GSW représente gradients le long du Sud, Sud-Est, Est, Nord-Est, du Nord, du Nord-Ouest, Ouest et Sud-Ouest directions respectivement alors le maximum de calcul des gradients est appelée bord amplitude qui est calculée comme suit:

$$G_{max} = \max \{GS, GSE, GE, GNE, GN, GNW, GW, GSW\} \dots\dots\dots (IV.6)$$

Ce bord amplitude est la valeur maximale au niveau de chaque pixel qui est calculée par la convolution de chaque masque à l'image de sorte que la direction de bord locale est calculée avec le noyau qui produit l'amplitude maximale [25-29]. Aux fins de l'extraction de caractéristiques à partir du bord détecté par les filtres Robinson, le pixel droit est considéré que les caractéristiques de cette recherche alors qu'il a été normalisé en divisant à 255, parce que l'intensité d'un pixel pour une image d'échelle de gris est compris entre 0 à 255.

$$Feature = G_{max} / 255 \dots\dots\dots (IV.7)$$

Dans cette tendance le nombre total de caractéristiques extraites est dépendent du nombre de pixels dans l'image. Par exemple, si l'image est de 10 × 10, le nombre total de caractéristiques extraites pour cette image par le filtre Robinson serait de 100 caractéristiques.

**Algorithme de segmentation**

- 1 . d'entrée : la plaque d'immatriculation détectée ( de la partie détection )
- 2 .convertie image en niveaux de gris
- 3.éliminer le bruit en utilisant la médiane
4. le filtrage à partir de l'image d'entrée.
5. le masque 3 × 3 est utilisé pour obtenir huit voisins d'un pixel et de leurs valeurs de gris correspondantes.
6. sortie : régions segmentées pour la reconnaissance de caracteres

**II.2.2.3 Reconnaissance de la plaque :**

Pour chaque caractère détecté puis normalisé en taille, les paramètres de forme (surface, périmètre, élongation, compacité, nombre de trous, etc.) doivent être calculés . L'évaluation du classifieur a montré que les paramètres de forme étaient peu voire pas discriminants. Les paramètres conservés

## Chapitre IV : Implémentation

ont été les 128 paramètres de répartition et les surfaces des deux trous inclus dans le caractère (elles sont mises à 0 lorsqu'il n'y a pas de trou). La base d'apprentissage a été constituée à partir de caractères simulés, en utilisant les polices rencontrées sur les plaques d'immatriculation, auxquels des transformations de type rotation et perspective ont été appliquées afin de prendre en compte les conditions d'acquisition.

### II.2.3 Résultats et caractérisation

Nous avons appliqué ce système sur les 222 exemples de plaques d'immatriculation réelles. Il identifie correctement 92.5 % des exemples; ce qui correspond au taux de reconnaissance en micro-précision.

Methode	Etapes	Détection	Segmentation	Reconnaissance
GLCM	Taux d'exécution	92.5%	/	92.5 %
EDMS		90.83%	/	/

Tableau IV.3: taux d'exécution

### III. comparaison des résultats des deux méthodes

La reconnaissance de caractères donne des conséquences satisfaisantes dans les deux cas global et local (plus de 90% de réussite sur les images de bonne qualité). Le taux de reconnaissance plus faible des plaques s'explique par l'échec de reconnaissance lorsqu'un des caractères est incorrectement reconnu, ce qui augmente la probabilité d'échec, et par la détection imparfaite des caractères (caractères non détectés).

Méthode	ETAPES	Détection	Segmentation	Reconnaissance
Local MSER	Taux	95.2%	97%	93.75%
Global GLCM	d'exécution	92.5%	/	92.5%

Tableau IV.4 : comparaison des résultats des méthodes proposées

Les résultats expérimentaux montrent que la méthode basée sur MSER est plus performante que celle basée sur GLCM et EDMS avec un taux de reconnaissance de 93.75%, ce qui homologue notre hypothèse supposé à la fin de chapitre III.

### IV. Conclusion

Nous avons présenté dans ce chapitre en premier lieu les deux méthodes de reconnaissance des plaque basées sur les caractéristiques locales en utilisant MSER et sur les caractéristiques globales en utilisant GLCM ( leurs systèmes de détection et de segmentation) et ensuite nous avons donné les résultats d'exécution de chacune des méthode afin de les comparée.

## Chapitre IV : Implémentation

# Conclusion générale

## Conclusion générale

Le travail présenté ci-dessus est un modeste travail dans le quel nous avons essayé de réaliser une comparaison des méthodes basées sur les caractéristiques locales ou globales pour la reconnaissance des plaques d'immatriculation et cela en abordant des concepts très intéressants dans le domaine de l'audio-visuel et de la vision par ordinateur.

Pour la réalisation de cette dernière nous avons jugé utile de décomposer notre travail en deux parties :

- ❖ La première partie consistait à présenter une étude bibliographique composée de deux chapitres :
  - Le premier présente les notions de bases relatives à la vidéo surveillance
  - Le deuxième présente les notions de bases relatives au traitement d'image où nous avons abordé également la notion de caractéristique (extraction et description).
  
- ❖ La deuxième partie est la base de notre travail, elle est composée de deux chapitres :
  - Le premier est destiné au système de reconnaissance automatique des plaques d'immatriculation (RAPI) ( ses étapes et son domaine d'utilisation) .Quelques méthodes existantes pour la reconnaissance des plaque d'immatriculation en utilisant soit l'information locale d'image ou l'information globale de celle-ci ont été présentées . Une comparaison de ces méthodes basée sur le taux de reconnaissance a été effectuée .
  - Le deuxième présente l'implémentation des méthodes de reconnaissance proposées

Les objectifs assignés dans le cadre de ce projet ont été globalement atteints. La réalisation de ce travail nous à permis d'acquérir de nouvelles connaissances sur le traitement d'image, d'approfondir nos connaissances dans la programmation java.

Une des perspectives qui apparaît à l'issue de ce travail est la possibilité de l'utiliser comme première démarche vers un système RAPI algérien puis qu'il est considéré comme une aide à la discision, Il est également intéressant d'intégrer ces méthodes dans un système embarqué à temps réel.

## Conclusion générale

## Conclusion générale

### Bibliographie

- [1] B.-F. Wu, S.-P. Lin, and C.-C. Chiu, "Extracting characters from real vehicle license plates out-of-doors," *IET Comput. Vis.*, vol. 1, no. 1, pp. 2–10, 2007.
- [2] D. Ventzas, Vehicle's License Plate Recognition System based on a Neural Network Radon Transform Method, International virtual conference, Advanced Research in Scientific Areas, pp. 2097-2104,2012.
- [3] An Intelligent License Plate Recognition System Zyad Shaaban University of Tabuk Tabuk, Kingdom of Saudi Arabia
- [4] CHARACTER AND OBJECT RECOGNITION BASED ON GLOBAL FEATURE EXTRACTION MARYAM NAEIMIZAGHIANI<sup>1</sup> , SITI NORUL HUDA SHEIKH ABDULLAH
- [5] An Iranian License Plate Recognition System Based on Color Features Amir Hossein Ashtari, *Graduate Student Member, IEEE*, Md. Jan Nordin, and Mahmood Fathy
- [6] Automatic License Plate Recognition System using SURF Features and RBF Neural Network
- [7] Detecting ,Tracking and Reconizing License Plates
- [8] AUTOMATIC VEHICLE REGISTRATION PLATE RECOGNITION SYSTEM USING SOFT COMPUTING TECHNIQUE
- [9] MEMOIRE VIDEOPROTECTION – CONTROLE ACCES - INTRUSION
- [10] S.K. Kim, D.W. Kim and H.J. Kim, "ARecognition of Vehicle License Plate Using a Genetic Algorithm Based Segmentation",*Proceedings of International Conference on Image Processing*, 1996.
- [11] S. Kim, et al., "A Robust License-plate Extraction Method under Complex Image Conditions", *Proceedings of 16th International Conference on Pattern Recognition*, 2002.
- [12] W. Jia., H. Zhang, X. He and M. Piccardi, "Mean Shift for Accurate License Plate Localization",*Proceedings of International Conference on Intelligent Transportation Systems*, 2005.
- [13] S.L. Chang, et al., "Automatic License Plate Recognition", *IEEE Transactions on Intelligent Transportation Systems*, 2004, 5(1): p. 42-53.
- [14] An Evaluation of License Plate Recognition Algorithms Hadi Sharifi Kolour Department of Information Technology Local Enhancement of Car Image for License Plate Detection *Vahid Abolghasemi and Alireza Ahmadyfard*
- [15] Unconstrained Licence Plate and Text Localization and Recognition Jiri Matas Karel Zimmermann Czech Technical University Prague Center for Machine Perception
- [16] M. Shridhar and J. Miller. Recognition of license plate images: Issues and perspectives. In Fifth International Conference on Document Analysis and Recognition, pages 17–21, 1999.

## Conclusion générale

- [17] Paul Viola and Michael Jones. Robust real-time object detection. *International Journal of Computer Vision* - to appear, 2002.
- [18] Efficient Maximally Stable Extremal Region (MSER) Tracking Michael Donoser and Horst Bischof Institute for Computer Graphics and Vision Graz University of Technology {donoser,bischof}@icg.tu-graz.ac.at
- [19] Licence Plate Localization And Recognition Using Mser And Swt Algorithms Annie J. Zenath M.S, S. Shriram
- [20] Ahmad Radmanesh, June 2005, "A Real Time Vehicle's License Plate Recognition System", Proceedings of the IEEE Conference
- [21] S.Kranthi, K.Pranathi, A.Srisaila "Automatic Number Plate Recognition (ANPR) International Journal of Advancements in Technology, ISSN 0976-4860
- [22] Detecting, Tracking and Recognizing License Plates Michael Donoser, Clemens Arth, Horst Bischof
- [23] Matas, J., Zimmermann, K.: Unconstrained licence plate and text localization and recognition. In: Proceedings of the IEEE Conference on Intelligent Transportation Systems (ITSC), Vienna, Austria (2005) 572–577
- [24]. Rahman, C., Badawy, W., Radmanesh, A.: A real time vehicle's license plate recognition system. In: Proceedings of the IEEE Conference on Advanced Video and Signal Based Surveillance (AVSS). (2003) 163–166
- [25] Matas, J., Zimmermann, K.: Unconstrained licence plate detection. In: Proceedings of International Conference on Intelligent Transportation Systems (ITSC). (2005) 572–577
- [26] Extracting characters from real vehicle licence plates out-of-doors B.-F. Wu, S.-P. Lin and C.-C. Chiu
- [27] The efficient real- and non-real-time make and model recognition of cars Remigiusz Baran · Andrzej Glowacz · Andrzej Matiolanski
- [28] A Study of Automatic License Plate Recognition Algorithms and Techniques Nima Asadi Intelligent Embedded Systems
- [29] Système de Reconnaissance Automatique des Plaques Minéralogiques Ameer CHHAYDER et Imene BELHADJ MOHAMED *IPEIS BP 3018 Sfax Tunisie*
- [30] Text Extraction from image using MSER approach Geetanjali Adlinge<sup>1</sup>, Shashikala Kashid<sup>2</sup>, Tejasvini Shinde<sup>3</sup>, Virendrakumar Dhotre<sup>4</sup>
- [31] THÈSE DE DOCTORAT Hinde ANOUAL Détection et Localisation de texte dans les images de scènes naturelles : Application à la détection des plaques d'immatriculation marocaines
- [32] Efficient Maximally Stable Extremal Region (MSER) Tracking
- [33] Local Invariant Feature Detectors: A Survey Tinne Tuytelaars<sup>1</sup> and Krystian Mikolajczyk<sup>2</sup>

## Conclusion générale

[34] IMAGE QUALITY ANALYSIS USING GLCM

[35] Développement d'une méthode de représentation d'image sur base d'un vocabulaire visuel

[36] Object detection ,recognition and re-identification in video footage

[37] AQUISITION E et TRAITEMENT D'IMAGE NUMERIQUE

[38] THÈSE DE MAGISTER djazia ZEROUAL IMPLIMENTATION D'UN ENVIRONNEMENT PARALLELE POUR LA COMPRESSION D'IMAGE A L'AIDE DES FRACTALES