

17004.525

الجمهورية الجزائرية الديمقراطية الشعبية

République Algérienne Démocratique et Populaire

Ministère de l'enseignement supérieur et de la recherche scientifique

Université de 8 Mai 1945 – Guelma -

Faculté des Mathématiques, d'Informatique et des Sciences de la matière

Département d'Informatique



**Mémoire de Fin d'études Master**

**Filière : Informatique**

**Option : Informatique Académique**

16 / 9 / 13

**Thème :**

---

**Traitement dynamique des modèles de l'arrière  
plan dans une vidéo**

---

**Encadré Par :**

**Dr.Brahim FAROU**

**Présenté par :**

**Sara AGABI**

**Imane HANNACHI**

**Juin 2016**

# Remerciements

*Tout d'abord, nous tenons à remercier Allah, Le Tout Puissant et Le Miséricordieux, de nous avoir donnés la santé, le courage, la volonté et la patience de réaliser ce travail.*

*C'est avec un grand honneur que nous remercions notre cher enseignant et promoteur, Monsieur FAROU Brahim, Pour nous avoir accepté de nous encadrés et pour nous avoir dirigés tout au long de la réalisation de ce travail, pour ses précieux conseils, sa patience et ses encouragements.*

*Un très grand merci à nos très chers parents, pour leurs soutiens, leurs encouragements, leurs sacrifices, eux qui nous ont guidées vers le chemin de la réussite.*

*Merci aux membres du jury de nous avoir fait l'honneur d'examiner ce travail.  
Sans oublier bien entendu tous les enseignants du département d'informatique.*

*Hannachi Imane & Agabi Sara*



# Dédicaces

*Je dédie ce modeste travail,*

*A mes très chers parents, pour leur soutien, leurs encouragements, leurs sacrifices, eux qui m'ont guidés durant toutes mes années d'étude vers le chemin de la réussite « Papa, Mama merci pour tout ».*

*A Ma Grand-Mère maternelle «Youma», Allah yarhamha,*

*A Ma Grand-Mère paternelle je vous aime,*

*A mes sœurs chéries : Silem, Wafa, Catifa son mari Sami et sa petite Cydia,*

*A tous mes cousins et cousines : Roumaissa, Cheroufa, Said, Minou, Bessouma, Mustapha, Chacha , Anya, Racim, Manel, Amir, Anfal , Anass, Amira et leurs enfants Djad, Miral,*

*A mes chers amis : Amine et leur famille, Hayat,*

*A mes très chères tantes : Tata, Saoty, Fatomty, tata Yamina et leurs familles,*

*A ma chère Sara qui m'a accompagné durant mon chemin universitaire, et leur famille,*

*A toute personne que je n'ai pas citée et qui m'a aidé de près ou de loin, je vous remercie,*

*A tous les enseignants qui m'ont encouragé et soutenu.*

*Hannachi*

*Imane*



# Dédicaces

*Je dédie ce modeste travail,*

*A ma mère la perle la plus chère pour tendresse, sa sympathie, son sacrifice et  
pour son orientation*

*A mes chères grand mères Tata Fatima et Tata Malika je vous aime beaucoup,*

*A mes très adorables frères : Fouzi et son épouse, Fethi, Salah-eldinne, Kheire-  
eldinne et Nasr-eldinne,*

*A mes sœurs chéries : Wahiba et son mari Wahab, Samira et son mari Ismail,  
Doria, Bessma et son mari Abd-elkader, Selma et son mari Salih,*

*A Dina, Joumana, Islem, Abdou, Micha, Jassem, Roudaina, Yaakoub, Rami,  
Ghayt, Mayssa, Maya et Simou,*

*A ma deuxième famille qui est mes amis : Fares et leur famille, Rima et leur  
famille, Marwa, Sihem et Hayat,*

*A ma chère Imane qui m'a accompagné durant mon cursus universitaire, et leur  
famille,*

*A tous les enseignants qui m'ont encouragé et soutenu pendant mon cursus,*

*A tous ceux qui me sont chers, là où ils pourraient se trouver.*



*Agabi*

*Sara*

## Résumé

La vidéo est devenue un support d'information incontournable dans notre quotidien. Le besoin d'extraire les informations contenues dans ces derniers de façon automatique nous oblige à concevoir des algorithmes rapides et robustes pour une bonne exploitation.

La nature dynamique des vidéos nécessite des algorithmes avec des paramètres adaptatifs pour pouvoir exprimer l'évolution des scènes à travers le temps.

Pour cela nous avons proposé un algorithme avec un paramétrage flexible pour exprimer la nature dynamique de la vidéo à travers de modèles de l'arrière plan.

# Sommaire

Résumé .....	i
Sommaire .....	ii
Liste des figures .....	v
Liste des tableaux .....	vi
Introduction générale.....	1

## **CHAPITRE 1 : Les systèmes de vidéo surveillance**

1. Introduction .....	3
2. La vidéo surveillance .....	3
2.1 Définition .....	3
2.2 Présentation .....	4
2.3 Historique .....	4
2.4 Objectifs majeurs .....	7
2.5 Domaine d'application .....	7
3. Le document vidéo .....	8
4. Les caractéristiques d'une vidéo .....	9
4.1 nombre de frame par seconde .....	9
4.2 Entrelacement .....	9
4.3 Résolution d'affichage .....	10
4.4 Proportion .....	10
5. Analyse de contenu de vidéo .....	11
6. Modélisation du contenu vidéo .....	11
7. Les formats de fichiers vidéo numériques.....	11
8. Les images de références .....	12
9. Les images .....	12
9.1 Définition .....	12
9.2 Les pixels .....	13
9.3 Les espaces de couleur .....	13

9.3.1 L'espace RGB.....	13
9.3.2 L'espace HSV .....	14
9.3.3 L'espace XYZ .....	14
9.3.4 L'espace Yuv .....	15
9.4 Le passage d'un espace a un autre .....	15
9.5 Résolution .....	16
10. Conclusion .....	16

## **CHAPITRE 2 : Méthodes d'extraction de l'arrière plan**

1. Introduction .....	18
2. Dérivée temporelle .....	18
3. Flot optique .....	18
4. Modélisation par une seule gaussienne .....	19
5. Mixture de gaussienne .....	19
6. Estimation de la densité basée sur un noyau (KDE) .....	19
7. Estimation de la densité basée sur un noyau adaptatif (KDEA) .....	20
8. Modélisation récursive .....	20
9. Différence de frames .....	21
10. Soustraction par image de référence .....	21
11. Apprentissage sub-spatial basé ACP (SL-PCA) .....	21
12. Filtrage prédictifs .....	21
13. Filtre de Kalman .....	23
14. Dictionnaire de mots visuels .....	25
15. Séparateur a vaste marge (SVM) .....	25
16. Les séparateurs à vaste marge pour la régression (SVR) .....	26
17. Les séparateurs à vaste marge pour la description de données (SVDD) .....	27
18. Simple gaussienne généralisée .....	29
19. Mélange de gaussienne généralisée .....	30
20. Conclusion .....	30

# Liste des figures

<b>Figure</b>	<b>Titre</b>	<b>N° de la page</b>
1.1	Caméras de vidéo surveillance	3
1.2	Système de vidéo surveillance	4
1.3	Système de caméra de surveillance (caméra reliée à un simple écran de contrôle)	6
1.4	Système de caméra de surveillance (caméra analogique)	6
1.5	Système de caméra de surveillance (caméra numérique consultable)	6
1.6	Caméra de vidéosurveillance dans un bureau	8
1.7	Caméra de vidéosurveillance dans l'autoroute	8
1.8	Caméra de vidéosurveillance dans une maison	8
1.9	Les trois composants de l'espace RGB	13
1.10	Les trois composants de l'espace HSV	14
1.11	L'espace XYZ	15
1.12	L'espace Yuv	15
3.1	Architecture du système	32
3.2	Différentes forme de contour	33
3.3	Masques utilisés dans la dilation	36
3.4	Masques utilisés dans l'érosion	37
3.5	Processus de classification	41
4.1	Vidéo fragmentée	46
4.2	Image originale	47
4.3	Image en niveaux de gris	47
4.4	Image avec l'algorithme de canny	47
4.5	Image après l'application des opérations morphologiques	48
4.6	Echantillons d'image de dilation et de séparation	48
4.7	WavingTrees(a) image originale, (b) image segmenté a la main	49
4.8	Résultats des testes avec changement de paramètre (Waving Trees).	49
4.9	TimeOfDay(a) image originale, (b) image segmenté a la main	50
4.10	Résultats des testes avec changement de paramètre (TimeOfDay)	50





---

4.11	Bootstrap (a) image originale, (b) image segmenté a la main	51
4.12	Résultats des testes avec changement de paramètre (Bootstrap)	51
4.13	Camouflage (a) image originale, (b) image segmenté a la main	52
4.14	Résultats des testes avec changement de paramètre (Camouflage)	52
4.15	ForegroundAperture (a) image originale, (b) image segmenté a la main	53
4.16	Résultats des testes avec changement de paramètre (ForegroundAperture)	53
4.17	MovedObject (a) image originale, (b) image segmenté a la main	54
4.18	Résultats des testes avec changement de paramètre (MovedObject)	54
4.19	LightSwitch (a) image originale, (b) image segmenté a la main	55
4.20	Résultats des testes avec changement de paramètre (LightSwitch)	55

# Liste des tableaux

<b>Tableaux</b>	<b>Titre</b>	<b>N° de la page</b>
4.1	Base des vidéos	45

## Introduction générale

Avec la généralisation de l'utilisation des images numériques, l'analyse du mouvement dans les vidéos s'est révélée être un outil indispensable pour des applications aussi diverses que la vidéo surveillance, la compression vidéo, l'imagerie médicale, la robotique, l'interaction homme machine, l'analyse de séquences sportives...etc. En effet, les zones de mouvement d'une séquence d'images correspondent souvent à des événements sur lesquels un système de vision doit se focaliser. Le but de ce travail est la détection et le suivi d'objets en mouvement dans une séquence d'image.

Nous allons donc, dans un premier temps, faire une petite présentation de la vidéosurveillance puis nous allons présenter des généralités sur la vidéo, leurs caractéristiques. Nous allons ensuite aborder l'image en traitant les mêmes points.

Dans la deuxième partie nous allons voir les différentes méthodes de modalisation de l'arrière plan. Une étude approfondie est réalisée afin de bien situer notre contribution par rapport à ce qui existe.

Le troisième chapitre présente l'architecture générale de notre système qui se divise en six modules. Chaque module est un processus qui permet de réaliser une tâche spécifique relative à un traitement de la vidéosurveillance. Nous avons d'abord commencé par une fragmentation des vidéos reçues en entrée. Le second module est un filtre de Canny qui permet de détecter les contours afin de minimiser les espaces de recherche. Nous avons appliqué des opérations morpho mathématique basé sur la dilation et l'érosion afin de supprimer les bruits causés par l'extraction de l'arrière plan. La partie extraction est un mélange de loi gaussienne à laquelle, nous avons appliqué des modifications sur le comportement de ces derniers dans un environnement à très forte variations.

Dans le chapitre quatre, nous avons effectué des expérimentations sur un ensemble de vidéos récupérées à partir d'une base de tests publique. Les vidéos utilisées se différencient par rapport au nombre d'objets, à la fluctuation, à la présence de mouvement et à la nature des objets traités. Les résultats obtenus sont très satisfaisants par rapport à l'algorithme de base.

Nous avons terminé ce mémoire par une conclusion et quelques perspectives pour une éventuelle continuité dans le même plan de travail.

# **CHAPITRE I**

## **Les systèmes de vidéo surveillance**

## 1. Introduction :

La vidéosurveillance est omniprésente dans ses jours, et on la retrouve dans de nombreux secteurs d'activité.

La vidéosurveillance consiste à placer des caméras de surveillance dans un lieu public ou privé et de recevoir le flux vidéo sur un PC localement ou à distance, nous allons bien définir cette approche dans ce chapitre.

Les causes de l'installation de systèmes de vidéosurveillance sont diverses, la sécurité publique ainsi que la protection des biens mobiliers ou immobiliers. L'industrie de la vidéosurveillance englobe aujourd'hui toute une variété de systèmes et d'équipements de surveillance et de protection des personnes et des biens.

Dans ce chapitre, nous décrivons tout d'abord c'est quoi une vidéosurveillance ainsi que les objectifs majeurs de ce système et les domaines d'application, ensuite nous allons procéder à la description du support vidéo et les images.

## 2. La vidéosurveillance :

### 2.1 Définition :

La vidéosurveillance ou vidéo-protection est un système de caméras et de transmission d'images, disposé dans un espace public ou privé pour le surveiller à distance ; il s'agit donc d'un type de télésurveillance. Les images obtenues avec ce système peuvent être traitées 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 [S1].



*Figure 1.1 : Caméras de vidéo surveillance.*

## 2.2 Présentation :

La vidéosurveillance est un ensemble de caméras disposées dans une zone publique ou privée pour mieux la surveiller. Pour ce faire, des caméras vidéo sont placées aux endroits à surveiller et ils sont reliés à un terminal qui permet de visualiser les images fournies par toutes les caméras. Aussi, une installation de vidéosurveillance permet d'enregistrer les vidéos filmées par les caméras de surveillance. Deux types de systèmes de vidéosurveillance sont utilisés : le système analogique (le plus ancien utilise des caméras qui enregistrent sur des bandes magnétiques).

Système numérique, ce sont des données binaires qui transitent vers les terminaux et ces données sont stockées sur des disques.

Avec les fonctionnalités des caméras qu'ils sont très prisés dans le domaine de la sécurité, elles deviennent de plus en plus sophistiquées. Aussi la simple capture d'image, la vidéosurveillance permet de compter des individus, de détecter les changements dans un champ, de suivre les déplacements d'un objet du champ, de reconnaître automatiquement des informations comme des codes-barres et des plaques d'immatriculation, etc [S2].



*Figure 1.2 : Système de vidéo surveillance.*

## 2.3 Historique :

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, 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.

- **Première génération : le tout analogique :**

Dans le réseau TVCF (réseau de télévision en circuit fermé) analogique traditionnel, des caméras analogiques sont connectées par câbles coaxiaux (un câble par caméra) aux écrans de surveillance et, pour des fins d'archivage, à un magnétoscope qui enregistre la vidéo sur cassette. Un multiplexeur peut être utilisé pour grouper les flux vidéo de plusieurs caméras en un seul signal composite qui est transmis au magnétoscope ou à un moniteur analogique. Il permet d'afficher quatre, neuf ou 16 signaux vidéo sur un même écran, ou d'enregistrer ceux-ci sur un même système d'archivage. Aucune compression n'est effectuée sur les signaux vidéo. Avec un magnétoscope conçu à cet effet, l'enregistrement à taux de trame (*frame rate*) réduit permet d'économiser de l'espace sur la bande vidéo, selon les besoins de surveillance.

- **Deuxième génération : le système hybride**

Le remplacement du magnétoscope à cassette par un enregistreur numérique (DVR Enregistreur vidéo numérique) représente la première étape de la transition numérique en vidéosurveillance. Apparus dans les années 90, les enregistreurs numériques archivent la vidéo sur des disques durs. Comportant souvent plusieurs entrées vidéo, l'enregistreur numérique remplace à la fois le multiplexeur et le magnétoscope analogique.

Les modèles récents sont équipés d'un port Ethernet, permettant la connexion à un réseau et l'accès à distance à la vidéo, soit en temps réel, soit à partir de l'enregistrement. Les transmissions vidéo se font sous protocole IP à partir de l'enregistreur numérique. Ils permettent la transition vers un système hybride de vidéosurveillance sur réseau IP, tout en conservant les caméras analogiques. L'enregistreur numérique hybride (HDVR Enregistreur vidéo hybride), quant à lui, peut recevoir à la fois les flux vidéo de caméras analogiques et IP. Il représente une solution efficace pour moderniser un système de vidéosurveillance en profitant des avantages des nouvelles caméras IP, tout en conservant les caméras analogiques existantes en place.

- **Troisième génération : le tout numérique IP**

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.



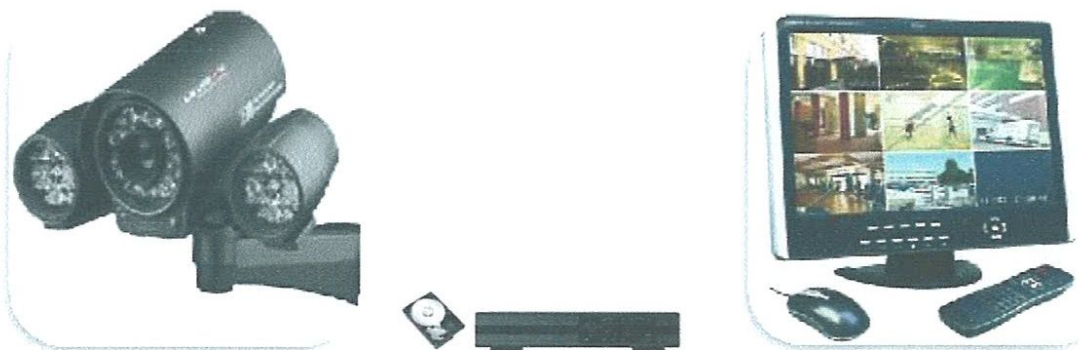
Ces réseaux comprennent donc des caméras réseau, aussi appelées IP. Il s'agit de caméras intégrant leur propre encodeur. Celles-ci sont reliées, via des commutateurs réseau, à des serveurs (ordinateurs personnels), munis d'un logiciel de gestion vidéo. L'enregistrement est fait sur serveur ou sur des 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 transmis 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 [VAL 09].



*Figure 1.3* : Système de caméra de surveillance (caméra reliée à un simple écran de contrôle).



*Figure 1.4* : Système de caméra de surveillance (caméra analogique).



*Figure 1.5* : Système de caméra de surveillance (caméra numérique consultable).

## 2.4 Objectifs majeurs :

L'objectif général d'un système de vidéosurveillance est de contribuer à la sécurité de biens et/ou de personnes [S4].

- Surveiller l'espace en vision large ou en zoom
- Déterminer les situations selon un mode paramétré, de piloter des systèmes en local
- Dissuader par la vision d'équipement de caméras
- Détecter les mouvements et comportements anormaux
- Identifier une présence, un objet, un véhicule, etc.

## 2.5 Domaines d'application :

**Domaine privé :** télésurveillance de résidence exemple les bureaux.

**Domaine public :** amélioration des conditions de transport et de la sécurité des usagers.

- Sur routes et autoroutes :
  - ✓ *Gestion du trafic*, surveillance des zones dangereuses etc.
  - ✓ *Accidentologie* : étude du comportement des automobilistes pour modification éventuelle des voies de circulation et de l'équipement
- Dans les aéroports :
  - ✓ Tour de contrôle : supervision de mouvements des avions sur les pistes, les voies d'accès aux pistes et les parkings.
  - ✓ Bagages, surveillance des parkings d'automobiles.

**Domaine industriel :** associée éventuellement à différents filtres la caméra permet :

- L'observation de matériel ou de structures
- La recherche et l'exploration sous-marine
- L'augmentation de l'efficacité des sites de production, le contrôle des stocks, et l'amélioration de la sécurité dans le cadre des chaînes de fabrication etc...

**Domaine commercial :**

- Commerce : étude et analyse d'activité, optimisation de l'agencement des magasins
- Loisir et tourisme : aperçu instantané et dynamique des infrastructures et du site

**Domaine médical :**

- Utilisation en salle d'opérations pour assistance avec déport du champ opératoire pour l'anesthésiste, les infirmières, mais aussi pour le télé-enseignement
- Utilisation en salle de réveil pour assistance permanente

Le flux visuel comporte une séquence d'images fixes qui selon l'axe temporel apparaissent animées à une fréquence de 24 à 30 images par seconde [BOU 12].

Le flux sonore est composé d'un ou plusieurs canaux, il est échantillonné entre 16000 et 48000 Hertz.

Un troisième flux d'information associée aux documents vidéo est le texte. Il provient soit d'un flux séparé, soit il est dérivé des sources audio et visuelle.

Une séquence vidéo brute est une suite d'images fixes, se caractérise par trois principaux paramètres :

- La résolution en luminance détermine le nombre de couleurs possibles pour un pixel, elle est généralement de 8 bits pour les niveaux de gris et de 24 bits pour les séquences en couleurs.
- • La résolution spatiale, définit le nombre de lignes et de colonnes de la matrice de pixels.
- La résolution temporelle est le nombre d'images par seconde.

#### 4. Les caractéristiques d'une vidéo :

Les principales caractéristiques de la vidéo sont :

**4.1 Nombre de frame par seconde :** En informatique, les termes anglais de *frame(s) per second* ou *FPS* sont couramment utilisés pour désigner le nombre d'images par seconde.

Le nombre d'images par seconde est une unité de mesure correspondant au nombre d'images affichées en une seconde, varie de 6 à 8 images par seconde pour les anciens appareils mécaniques à 120 ou plus d'images par seconde pour les nouveaux appareils photos professionnels.

**4.2 Entrelacement :** L'entrelacement (en anglais *interlace*), ou balayage entrelacé, est une technique destinée à doubler le nombre d'images par seconde perçue sans augmenter le débit.

Les lignes de balayage horizontal de chaque trame entrelacée sont numérotées consécutivement et divisé en deux champs, une trame paire et une trame impaire.

L'entrelacement est une technique développée pour transmettre les signaux de télévision en utilisant des bandes passantes limitées. Dans un système entrelacé, la moitié des lignes

horizontales de chaque image vidéo est transmise en une seule fois. En raison de la vitesse de transmission, de la rémanence de l'affichage et de la persistance visuelle, le spectateur perçoit chaque image avec une résolution complète. Toutes les normes de télévision analogique utilisent l'entrelacement, alors que Les normes de télévision numérique reposent sur les systèmes entrelacés et non entrelacés [S5].

**4.3 Résolution d'affichage :** La résolution est la "netteté globale de l'image" et il est mesuré comptant les lignes horizontales (résolution horizontale) et des lignes verticales (résolution horizontale) d'une image. Les lignes blanches sont également considérées comme une ligne dans l'image. La résolution quantifie comment étroitement les lignes sont à l'autre et pourtant encore être visiblement comptés [S6].

- *Résolution horizontale :* La résolution horizontale définit la capacité du système à résoudre des lignes verticales. Cela dépend de la caméra et des capacités d'affichage.
- *Résolution verticale :* La résolution verticale est indépendante de la largeur de bande du système et définit la capacité du système à résoudre des lignes horizontales. Il est exprimé en nombre de lignes horizontales distinctes, alternativement noir et blanc, qui peuvent être résolues de façon satisfaisante sur un écran de télévision.

**4.4 Proportion :** décrit les dimensions des écrans vidéo, les formats populaires des vidéos sont rectilignes, donc ont une largeur et une hauteur.

La largeur et la hauteur des vidéos sont généralement mesurées en pixels, si par exemple une vidéo est de 320 pixels de large et 240 pixels de hauteur, il est dit avoir des dimensions de 320 x 240 pixels, et dispose d'un aspect ratio de 4:3.

Le format d'un écran de télévision traditionnelle est de 4 :3, les télévisions haute définition utilise un ratio de 16 :9.

Ratios de plus grande hauteur que de largeur sont rares dans une utilisation quotidienne.

Un pixel est généralement carré (sa largeur est égale à sa hauteur), mais les pixels utilisés dans les vidéos numérique ont souvent des non-carrés aspect ratio, Tel que ceux utilisé dans les variantes PAL,(Phase.Alternating.Line), NTSC (National.Television.System.Committee) de la norme CCIR (Comité consultatif international des radiocommunications) 601 de vidéo numérique, et l'écran large anamorphique format correspondants.

## 5. Analyse du contenu de la vidéo :

Les documents vidéo peuvent avoir des contenus variés (télévisés, documentaires, films, publicité, vidéosurveillance, etc.)

Concernent l'analyse du contenu vidéo on peut distinguer deux niveaux de descriptions :

- *Le niveau signal* ou *bas niveau* : proche de la représentation numérique des documents, il décrit les caractéristiques physiques des segments d'une vidéo comme la couleur, la texture et la forme. Les informations correspondant à ce niveau sont en général de type numérique [BOU12].
- *Le niveau sémantique* ou *haut niveau*: proche de la façon dont les humains représentent le contenu des documents, il vise la description des concepts présents et des relations entre eux. Les informations correspondantes à ce niveau sont en général de type symbolique.

## 6. Modélisation du contenu vidéo :

Comme le document textuel, une vidéo possède une structure spécifique qui peut être soit homogène du fait qu'elle intègre les différents composants du document ou hétérogène lorsqu'il s'agit d'une analyse par média spécifique.

La modélisation d'un document vidéo dans son ensemble, est destinée à accorder une vue synthétique au document de manière de donner à l'utilisateur une idée sur le document.

Elle permet de faciliter l'accès au contenu des documents. Trois méthodes de modélisation des documents vidéo [BOU 12]:

- La modélisation hiérarchique : permet d'associer une description sous d'arborescence du contenu du document.
- La modélisation en strates ou stratification : associe aux parties du document vidéo.
- La modélisation par objets permet de décrire le contenu avec des objets audiovisuel. Elle prend la même structure que la modélisation hiérarchique.

## 7. Les formats de fichiers vidéo numériques :

Il existe plusieurs formats de fichier vidéo numérique, Les principaux sont :

- **Le format MPEG (Motion Picture Expert Group)** : Le principe de ce type de codage repose sur les redondances qui existent entre les images successives. On trouve plusieurs niveaux de la norme MPEG : MPEG1, MPEG2 et MPEG4.

- **Le format AVI (Audio VideoInterleave)** utilisé pour les fichiers Motion Jpeg (Joint Photographic Experts Group), et les fichiers numériques qui peuvent être lus par un logiciel gratuit.
- **Le format MOV (Quick time)** utilisé pour les fichiers vidéo Quick time [BOU 12].

## 8. Les images de références :

On appelle image de référence, l'image qui va être utilisée pour prédire une image et estimer le mouvement à compenser.

Cette image peut être simplement une image précédemment codée dans la séquence. Elle est utilisée sous la forme d'objet mosaïque qui est construit à partir d'une séquence de vues de l'objet considéré. Le principe de construction consiste à recalcr les images (ou portions d'images) traitées dans le référentiel de l'objet mosaïque. Ces images peuvent être utilisées pour représenter sur une seule image la totalité d'une scène [S8].

## 9. Les images :

### 9.1 Définition :

Une image est une représentation visuelle, voire mentale, de quelque chose (objet, être vivant, etc.)

En informatique, on désigne sous le terme d'une image numérique toute image qui a été acquise, traitée et sauvegardée sous une forme binaire, elles peuvent être acquises par des dispositifs comme les scanners, les appareils photos ou caméscopes...

Les images numériques peuvent être créées directement par des programmes informatiques, via la souris, les tablettes graphiques ou par la modélisation 3D et les traitées grâce à des outils informatiques.

Une image numérique est composée d'unités élémentaires (appelé pixel) qui représentent chacun une portion de l'image.

Une image est définie par:

- Le nombre de pixels qui la compose en largeur et en hauteur.
- L'étendue des teintes de gris ou des couleurs que peut prendre chaque pixel.

## 9.2 Les pixels :

Nous avons vu dans la précédente section qu'une image est définie par le nombre de pixels qui la compose et la multiplication de ces deux valeurs donne le nombre global de pixels d'une image, dans ce cas nous allons voir c'est quoi un pixel.

Les pixels sont les plus petits éléments constitutifs d'une image numérique. Le nom de "pixel", abrégé px, provient de l'expression anglaise picture element, qui signifie "élément d'image" ou "point élémentaire".

Un pixel est généralement rectangulaire ou presque carré et présente une taille comprise entre 0,18 mm et 0,66 mm de côté [S8].

Dans le cas d'un écran en noir et blanc, chaque pixel correspond à un point. Par contre, dans le cas d'un écran couleur, chaque pixel est constitué de trois points de couleurs primaires rouge, vert et bleu (RGB).

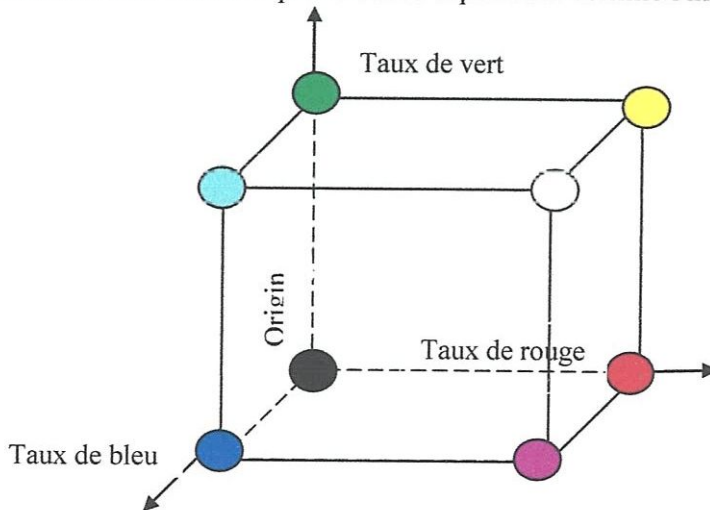
## 9.3 Les espaces de couleur :

Les couleurs trouvent leur origine dans la séparation de la lumière blanche naturelle, un mélange quelconque de ces couleurs donne une autre couleur.

### 9.3.1 L'espace RGB :

L'espace **RGB** est un espace vectoriel défini par la commission internationale de l'éclairage (CIE) en 1931, Il fait partie de la famille des espaces Rouge Vert Bleu.

L'ensemble des couleurs produites se représente comme l'intérieur d'un cube :



*Figure 1.9* : Les trois composants de l'espace RGB.

### 9.3.2 L'espace HSV :

Le modèle TSV pour Teinte Saturation Valeur (en anglais HSV pour Hue Saturation Value), le principe de cet espace est de caractériser les couleurs de façon intuitive, conformément à la perception naturelle des couleurs [MAN 12].

- **La teinte (H : HUE) :**  
intuitivement, c'est le nom qu'on utilisera pour désigner la couleur, "vert", "mauve", "orange", etc. Idéalement associé à une longueur d'onde, donc à une position sur le cercle de newton tel que :  
0° ou 360° : rouge;  
60° : jaune ;  
120° : vert ;  
180° : cyan ;  
240° : bleu ;  
300° : magenta.
- **La saturation (S) :**  
La saturation est l'« intensité » de la couleur, elle est parfois appelée « pureté ». Elle varie entre la pureté maximale (couleur éclatante) et l'achromatisme (niveau gris).
- **La valeur (V) :**  
La valeur est la « brillance » de la couleur, c'est la mesure de l'intensité lumineuse de la couleur qui doit varier entre le noir absolu et le blanc.

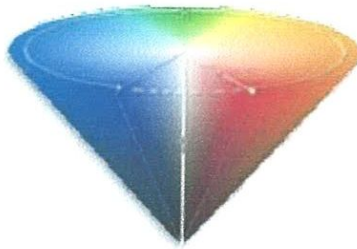


Figure 1.10 : Les trois composants de l'espace HSV.

### 9.3.3 L'espace XYZ :

L'espace XYZ (CIE 1931) est défini à partir d'une transformation linéaire sur l'espace RGB telle que toutes les couleurs du spectre visible soient contenues dans le triangle XYZ.



Les coordonnées XYZ des couleurs naturelles ont ainsi toujours des valeurs positives [MAN, 12].

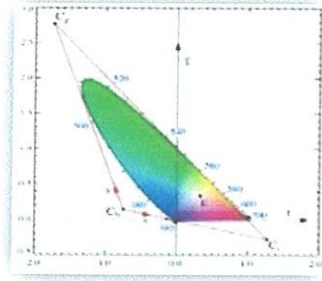
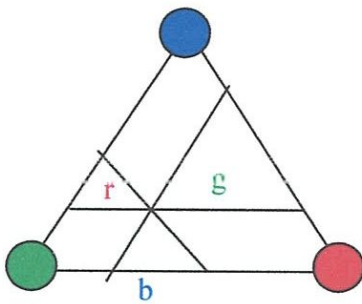


Figure 1.11 : L'espace XYZ.

### 9.3.4 L'espace Yuv :

Le principe de l'espace Yuv est de représenter les couleurs en utilisant une composante de luminance Y, et deux composantes de chrominance (u,v) correspondant aux composantes bleue et rouge dans les coordonnées chromatiques réduites [MAN 12] :



$$C_b + C_r + C_g = C_{te}$$

$$C_g = C_{te} - C_b - C_r$$

Figure 1.12 : L'espace Yuv.

## 9.4 Le passage d'un espace a un autre :

- **Le passage de RGB à HSV :**

Le passage de RGB à HSV se fait par une transformation non linéaire. Plusieurs opérateurs ont été proposés pour la conversion, nous proposons une transformation de l'image originale (RGB) en image HSV avec les équations suivantes [BEL 12]:

$$v = \frac{r+g+b}{3} \tag{1.1}$$

$$s = 1 - \frac{3\min(r,g,b)}{r+g+b} \tag{1.2}$$

$$h = \begin{cases} \theta & \text{si } b \leq g \\ 2\pi - \theta & \text{si } b > g \end{cases} \quad \theta = \arccos\left(\frac{(r-g)+(r-b)}{2\sqrt{(r-g)^2(r-b)(g-b)}}\right) \quad (1.3)$$

- **Le passage de RGB à Yuv:**

Le passage du système RGB au système Yuvse faite selon le système suivant [BEL 12] :

$$\begin{matrix} Y \\ u \\ v \end{matrix} = \begin{bmatrix} 0,299 & 0,587 & 0,114 & 0 \\ -0,1687 & -0,3313 & 0,5 & 1 \\ 0,5 & -0,41874 & -0,0813 & 1 \end{bmatrix} \times \begin{bmatrix} R \\ G \\ B \\ 128 \end{bmatrix} \quad (1.4)$$

### 9.5 Résolution :

La résolution d'une image est défini par le nombre de pixels par unité de longueur de la structure à numériser s'exprime en dpi (dots per inches) ou ppp( point par pouces).

Plus le nombre de pixels est élevé par unité de longueur de la structure à numériser, plus la qualité d'information qui décrit cette structure est importante et plus la résolution est élevée. La résolution d'une image numérique définit le degré de détail qui va être représenté sur celle-ci.

Les phénomènes de numérisation dépendent des deux équations suivantes :

$$X \times \text{résolution} = x \text{ pixels} \quad (1.5)$$

$$Y \times \text{résolution} = y \text{ pixels} \quad (1.6)$$

Avec X et Y représentent la taille géométrique en pouce ou en centimètres de la structure à numériser, résolution représente la résolution de numérisation, x et y représentent la taille en pixels de l'image [BEL 12].

### 10. Conclusion :

Dans ce chapitre nous avons présenté un petit résumé sur la vidéosurveillance avec un petit historique ainsi que les applications de ce système, nous avons présenté aussi le document vidéo et ses caractéristiques avec les différents formats d'une vidéo numérique, et une brève présentation sur les images avec les différents espaces de couleur,

# **CHAPITRE II**

## **Méthodes d'extraction de l'arrière plan**

## 1. Introduction :

La modélisation de l'arrière-plan est une étape essentielle dans la majorité des systèmes de vidéosurveillance, Il existe de nombreuses méthodes de modélisation qui utilisent à chaque instant l'ensemble des observations pour construire un modèle de l'ensemble de l'arrière-plan. Nous allons décrire quelques une au court de ce chapitre.

## 2. Dérivée temporelle :

Pour détecter les zones en mouvement dans un champ de vision on mesure le changement d'apparence des pixels entre deux trames consécutives, on suppose la dérivée temporelle en tout point.

L'intensité lumineuse d'un pixel étant un signal discret à une dimension, l'estimation de la dérivée temporelle instantanée du signal au temps  $t$  est donnée par :

$$\forall (x, y) \in \mathbb{E} \forall t > 0 I_t(x, y, t) \approx |I(x, y, t) - I(x, y, t - 1)| \quad (2.1)$$

Avec  $I(t)$  l'image présente dans le flux d'entrée au temps  $t$  et par  $I(x, y, t)$  le pixel de coordonnées  $x = (x, y)$  dans cette même image.

Cette méthode se montre peu robuste face à des phénomènes tels que les mouvements lents ou saccadés [VER 07].

## 3. Flot optique :

La dérivée temporelle quantifie la variation de l'aspect de chaque pixel considéré individuellement, le flot optique est un champ de vecteurs à deux dimensions représentant la projection sur le plan image du mouvement réel observé. Le temps de calcul du flot optique est une opération très coûteuse.

L'algorithme de Lucas et Kanade, 1981 s'agit de calculer au temps  $t$ , le déplacement  $d = (dx, dy)^T$  du point  $x = (x; y)^T$ . L'hypothèse de la conservation de l'intensité lumineuse [VER 07].

$$\forall x \in \mathbb{E} \forall t > 0 I(x + d, t + 1) - I(x, t) = 0 \quad (2.2)$$

La minimisation de la fonction d'erreur quadratique  $\xi$  calculée sur un voisinage  $N(x)$  se fait par le vecteur  $d$  (du flot optique au point  $x$ ).

$$\xi(d) = \sum_{x' \in N(x)} [I(x' + d, t + 1) - I(x', t)]^2 \quad (2.3)$$

#### 4. Modélisation par une seule gaussienne :

En probabilité, on dit qu'une variable aléatoire réelle  $X$  suit une loi normale (ou loi normale gaussienne) d'espérance  $\mu$  et d'écart type  $\sigma$  (strictement positif) si cette variable aléatoire admet pour densité de probabilité la fonction  $f(x)$  définie, pour tout nombre réel  $x$ , par :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2.4)$$

Avec

- **$\mu$  son espérance** : correspond à la moyenne de l'intensité du pixel.
- **$\sigma$  son écart type** : indique la dispersion de la couleur de pixel autour de ce centre exprimant la moyenne.

#### 5. Mixture de gaussienne :

Mixture de gaussienne ou mélange de gaussienne est une combinaison pondérée des distributions de probabilité. Elle sert à estimer la distribution de variables aléatoires en les modélisant comme une somme de plusieurs gaussiennes.

Il s'agit de déterminer l'écart type et la moyenne de chaque gaussienne.

La fonction de densité de probabilité est définie par une combinaison linéaire des fonctions. Un modèle avec  $M$  composantes est défini comme suit :

$$P(x) = \sum_{c=1}^M P(c)P(x|c) \quad (2.5)$$

Avec  $P(c)$  est appelé le coefficient de mélange et  $P(x|c)$  varie selon  $c$ .

#### 6. Estimation de la densité basée sur un noyau (KDE) :

L'estimation de densité basée sur un noyau (KernelDensity estimation, KDE) est une technique d'estimation de densité de probabilité à partir d'un ensemble d'échantillons.

Soit un ensemble d'échantillons,  $S = \{x_i\}$ ,  $i = 1 \dots N$  distribué selon la loi de densité  $p(x)$ . Une estimation de  $p(x)$  peut être évaluée par :

$$P(x) = \frac{1}{N} \sum_{i=1}^N K_{\sigma}(x - x_i) \quad (2.6)$$

Avec  $K_\sigma$  est une fonction noyau de largeur  $\sigma$  [BRU 12].

### 7. Estimation de la densité basée sur un noyau adaptatif (KDEA) :

L'estimation de densité par noyau adaptatif permet de s'adapter localement à l'échelle des données. Deux approches sont possibles [GRE 05]:

- L'estimation « ballon » : consiste à rendre H fonction de x.

$$\hat{f}(x) = \frac{1}{n|H(x)|^{\frac{1}{2}}} \sum_{i=1}^n K((H(x))^{-\frac{1}{2}}(x - x_i)) \quad (2.7)$$

- L'estimation « sample point estimator » : consiste à associer une matrice H à chaque  $x_i$ , la matrice H est remplacée par n matrices notées  $H(x_i)$  ou plus rapidement  $H_i$ .

$$\hat{f}(x) = \frac{1}{n} \sum_{i=1}^n |H_i|^{-\frac{1}{2}} K(H_i^{-\frac{1}{2}}(x - x_i)) \quad (2.8)$$

Avec H est la matrice de paramètres d'échelles, K fonction noyau.

### 8. Modélisation réursive :

La modélisation de l'arrière-plan est une étape importante dans les systèmes de vidéosurveillance.

Une méthode réursive: l'image d'arrière-plan est mise à jour pour chaque nouvelle image de la séquence d'apprentissage.

Supposons que la séquence d'apprentissage n'est pas soumise à des changements de luminosité et que les objets se déplacent continuellement dans la scène.

L'approche employée consiste en une estimation réursive de l'arrière-plan à l'aide d'un modèle statistique représentant la distribution de couleur pour chaque pixel de l'image.

Chaque pixel de l'image est représenté à l'aide d'une distribution gaussienne paramétrée par sa moyenne  $\mu$  et son écart-type  $\sigma$ .

Chaque pixel est modélisé à l'aide d'un vecteur contenant les caractéristiques couleurs RGB (red, green, bleu) s'écrivant  $[\mu_R, \mu_G, \mu_B, \sigma_R, \sigma_G, \sigma_B]$ . Pour chaque nouvelle observation x, les paramètres sont mis à jour réursivement :

$$\begin{cases} \mu = (1 - \alpha_t)\mu + \alpha_t x \\ \sigma^2 = \max\left((1 - \alpha_t)\sigma^2 + \alpha_t(x - \mu)^2, \sigma_{min}^2\right) \\ \alpha_t = \max(\alpha_{min}, 1/t) \end{cases} \quad (2.9)$$

Avec x la caractéristique couleur RGB d'un pixel et  $\alpha(t)$  un paramètre appelé taux d'apprentissage.

Le paramètre  $\alpha$  est rendu variable et décroissant pour les premières images de la séquence d'apprentissage jusqu'à une valeur limite  $\alpha_{\min}$ . Le choix de cette valeur est défini empiriquement et dépend de la séquence vidéo ainsi que de la fréquence d'acquisition [BRU 12].

### 9. Différence de frames :

La différence de frames ou bien la différence entre deux images consécutives consiste à la soustraction d'une image acquise au temps  $t_n$  d'une autre au temps  $t_{n+k}$ , ou  $k$  est habituellement égal à 1.

Si aucun mouvement ne s'est produit pendant l'intervalle de temps observé alors, l'image résultante sera vide parce que l'intensité et la couleur des pixels seront presque identiques. Mais, s'il ya du mouvement, les pixels frontières des objets en déplacement devraient changer drastiquement de valeurs, révélant alors la présence d'activité dans la scène.

Comme, cette technique nécessite très peu de ressources, car aucun modèle n'est nécessaire, alors cela implique donc qu'il n'y a pas de phase d'initialisation obligatoire avec une scène statique, ce qui procure une très grande flexibilité d'utilisation. En plus, une opération de soustraction d'images requiert très peu de puissance de calcul, lui conférant un avantage supplémentaire [LME 03].

Par contre, les résultats obtenus avec cette méthode ne sont pas éloquents que ceux générés en utilisant un modèle statistique de l'arrière-plan. En effet, certains traitements supplémentaires sont nécessaires afin de déterminer la zone en mouvement, car l'information disponible ne concerne que les contours des régions en déplacement.

### 10. Soustraction par image de référence :

La soustraction de l'arrière plan à l'aide d'une image de référence est utilisée pour affiner les résultats. Supposant, l'arrière-plan consiste en une image  $B$  dans laquelle la scène est représentée sans objet. La segmentation consiste à étudier chaque pixel de l'image et à les comparer avec l'image de référence  $B$ . Si la différence est supérieure à un seuil  $\tau$ , le pixel est classé en tant que l'avant-plan (foreground), sinon il est classé en tant que l'arrière-plan (background). Le résultat de la segmentation est une image binaire  $M$  donnée par :

$$M_t(u, v) = \begin{cases} 1 & \text{Si } |I_t(u, v) - B(u, v)| \geq \tau \\ 0 & \text{sinon} \end{cases} \quad (2.10)$$

$I_t$  est l'intensité lumineuse de la  $t$ -ième image et  $(u, v)$  les coordonnées d'un pixel de cette image.

Il est souvent difficile d'obtenir l'image d'arrière-plan et il est nécessaire de la mettre à jour régulièrement pour prendre en compte les changements de luminosité. L'estimation de l'image  $B$  peut être effectuée de différentes façons. Elle peut être représentée par la valeur moyenne estimée de façon incrémentale

$$B_t(u, v) = \frac{t-1}{t} B_{t-1}(u, v) + \frac{1}{t} I_t(u, v) \quad (2.11)$$

De façon générale, la valeur moyenne réursive permet d'estimer la moyenne réursive à l'aide du filtrage adaptatif suivant :

$$B_t(u, v) = \alpha B_{t-1}(u, v) + (1 - \alpha) I_t(u, v) \quad (2.12)$$

$\alpha$  est le taux d'apprentissage et permet de contrôler la vitesse de mise à jour de l'image d'arrière-plan [BRU 12].

### 11. Apprentissage sub-spatial basé ACP (SL-PCA) :

L'utilisation de méthodes d'analyse de données telles que l'ACP (Analyse en Composantes Principales) a également été appliquée pour la modélisation de l'arrière-plan.

L'objectif consiste à créer une matrice de données à partir d'un ensemble d'images d'apprentissage afin d'y appliquer une ACP et d'en dégager une base de vecteurs propres.

En pratique, la construction du modèle d'arrière-plan est réalisée à partir d'un ensemble de  $N$  images d'apprentissage prises à des instants différents dans la vidéo. Afin d'obtenir une plus grande représentativité de l'arrière-plan, ces images sont prises à des instants non consécutifs.

À partir de ces images est construite une image moyenne  $I_\mu$  et une matrice de covariance  $C$  calculée sur le vecteur  $X_I$  représentant le réarrangement des images d'entraînement sous la forme d'un vecteur une dimension. Comme  $w$  est la largeur,  $h$  est la hauteur et  $c$  le nombre de canaux couleur, alors chaque image de la base d'apprentissage est représentée sous forme d'un vecteur 1 dimension de taille  $(w \times h \times c)$  et la matrice contenant l'ensemble des  $N$  images est de taille  $(w \times h \times c \times N)$ . Une fois calculée, la matrice de covariance est diagonalisée pour obtenir une base de vecteurs propres  $\Phi$  et une matrice diagonale  $\Lambda$  comportant les valeurs propres associées (seuls les vecteurs propres associés aux  $K$  plus grandes valeurs propres sont conservés) [BRU 12].



Une fois l'apprentissage terminé, chaque nouvelle image  $I$  est projetée dans l'espace de dimension réduite. Les objets en mouvement sont extraits en calculant la distance entre l'image d'entrée  $I$  et l'image reconstruite à partir de sa projection notée  $I_\phi$  et donnée par :

$$I_\phi = (I - I_\mu)\Phi\Phi^T + I_\mu \quad (2.13)$$

## 12. Filtrage prédictifs :

La configuration d'un objet est représentée par son vecteur d'état contenant sa position  $(x, y)$  et son vecteur vitesse  $(v_x, v_y)$ . A l'instant  $t$ , le vecteur d'état  $X_{(i)}^t$  pour un objet  $i$  s'écrit  $X_{(i)}^t = [x_{(i)}^t y_{(i)}^t \dot{x}_{(i)}^t \dot{y}_{(i)}^t]^T$ .

Nous omettons de préciser l'indice  $(i)$  lorsqu'aucune ambiguïté ne se présente pour faciliter la lecture. A l'aide des relations de Newton, possible d'écrire les équations d'évolution du vecteur d'état par :

$$\begin{cases} x_t = x_{t-1} + \dot{x}_{t-1}\Delta t + \frac{1}{2}\ddot{x}_{t-1}\Delta t^2 \\ y_t = y_{t-1} + \dot{y}_{t-1}\Delta t + \frac{1}{2}\ddot{y}_{t-1}\Delta t^2 \\ \dot{x}_t = \dot{x}_{t-1} + \ddot{x}_{t-1}\Delta t \\ \dot{y}_t = \dot{y}_{t-1} + \ddot{y}_{t-1}\Delta t \end{cases} \quad (2.14)$$

Avec  $(x, y)$  la position prédite de l'objet,  $(\dot{x}, \dot{y})$  le vecteur vitesse des objets,  $(\ddot{x}, \ddot{y})$  le vecteur accélération des objets et  $\Delta t$  représente l'intervalle de temps entre deux prédictions [BRU, 12].

## 13. Filtre de Kalman :

Le filtre de Kalman est un filtre bayésien qui fournit une solution optimale et exacte. L'utilisation du filtre de Kalman nécessite la description du modèle sous la forme d'un système linéaire, et les perturbations appliquées au système sont de natures gaussiennes. On peut traduire ça par les équations suivantes :

$$x_t = F_t x_{t-1} + q_t z_t = H_t x_t + r_t \quad (2.15)$$

Avec

$$q_t \sim N(0, Q_t) r_t \sim N(0, R_t) \quad (2.16)$$

- $F$  est appelée matrice de transition du système, est connue et peut éventuellement évoluer au cours du temps.

- $q_t$  est un vecteur aléatoire de même dimension que  $x_t$ , inconnu et inaccessible à la mesure, et est appelé bruit d'état, Ce vecteur est supposé gaussien, centré en zéro et de matrice de covariance  $Q$ .
- $H_t$  est appelée matrice d'observation décrit la relation entre le vecteur d'état et l'observation.
- $r_t$  est un vecteur aléatoire de même dimension que  $z_t$  également inconnu et appelé bruit de mesure. Ce vecteur est supposé gaussien, centré en zéro et de matrice de covariance  $R$ .

Ce modèle est défini par sa densité de transition  $p(x_t, x_{t-1})$ , la vraisemblance des observations  $p(z_t, x_t)$  et la densité initiale  $p(x_0)$  toutes supposées gaussiennes avec  $m$  la moyenne et  $P$  la matrice de covariance :

$$p(x_t, x_{t-1}) = N(x_t | F_t x_{t-1}, Q_t) \quad (2.17)$$

$$p(z_t, x_t) = N(z_t | H_t x_t, R_t) \quad (2.18)$$

$$p(x_0) = N(x_0 | m_0, P_0) \quad (2.19)$$

L'algorithme de Kalman comporte les étapes suivantes [BRU 12] :

- Initialisation du vecteur d'état  $X$  et de sa matrice de covariance  $P$ .

$$m_{0/0} = m_0 \quad (2.20)$$

$$P_{0/0} = P_0 \quad (2.21)$$

- Calcul de l'estimation de l'état du système  $\hat{X}_t$  à l'instant  $t$  à partir des mesures à l'instant précédent  $t-1$ .

$$m_{t/t-1} = F_t m_{t-1/t-1} \quad (2.22)$$

- Mise à jour intermédiaire de la matrice de covariance de l'état  $P_{t/t-1}$

$$P_{t/t-1} = F_t P_{t-1/t-1} F_t^T + Q_t \quad (2.23)$$

- Calcul du gain du filtre de Kalman  $K_t$ . Ce gain ne dépend pas des données mesurées et tient compte uniquement des caractéristiques statistiques du bruit de mesure.

$$K_t = P_{t/t-1} H_t^T (H_t P_{t/t-1} H_t^T + R_t)^{-1} \quad (2.24)$$

- Mise à jour de la matrice de covariance de l'état.

$$P_{t/t} = (I - K_t H_t) P_{t/t-1} \quad (2.25)$$

- Correction de l'estimation de l'état.

$$m_{t/t} = m_{t/t-1} + K_t (Z_t - H_t m_{t/t-1}) \quad (2.26)$$

#### 14. Dictionnaire de mots visuels :

Pour modéliser l'arrière-plan A par exemple on utilise un dictionnaire de mots visuels, Ce dictionnaire consiste à construire pour chaque pixel un modèle représenté par un ensemble de variables appelées mots visuels caractérisant son état actuel [BRU 12].

Dans le modèle W4 (Minimum-Maximum filter), chaque pixel est caractérisé par un jeu de 3 valeurs, le minimum d'intensité (Min), le maximum d'intensité (Max) et la différence maximum d'intensité entre 2 images consécutives (D). Ces valeurs sont initialement estimées durant une phase d'apprentissage et mises à jour régulièrement au cours du temps.

Un pixel est considéré comme étant en mouvement si une des deux conditions suivantes est remplie :

$$|Min_t - I_t| > D_t \quad \text{ou} \quad |Max_t - I_t| > D_t \quad (2.27)$$

#### 15. Séparateur à vaste marge (SVM) :

Les séparateurs à vaste marge appelés aussi les machines à vecteurs de support sont des techniques d'apprentissage utilisé pour résoudre des problèmes de classification.

Les séparateurs à vaste marge sont une classe très spécifique d'algorithmes caractérisés par l'utilisation de noyaux, les séparateurs à vaste marge ont été inventés par Vapnik et al. en 1992 [VAP 92].

Pour la classification, les SVM travaillent en déterminant un hyperplan dans un espace de fonction dimensionnelle élevée pour séparer les données d'apprentissage en deux classes.

Le meilleur hyperplan peut être dérivé en minimisant la marge qui représente la plus petite distance entre l'hyperplan et les données.

L'utilisation de cette technique pour la modélisation de fond a été proposée par Lin et al [LIN 02].

Qui proposent une version probabiliste des SVM vu que ces derniers génèrent uniquement des sorties binaires. La transformation des sorties est assurée par une fonction sigmoïdale selon l'équation suivante :

$$P\left(y = \frac{1}{f}\right) = \frac{1}{1 + \exp(Af + B)} \quad (2.28)$$

Où y est une étiquette de classe binaire et f est un score de sortie de la fonction de décision des SVM. Les deux paramètres A et B sont adaptés en utilisant une estimation du maximum de vraisemblance de l'ensemble (f, y), et dérivé en minimisant la fonction log-vraisemblance négative.

Pour éviter le sur apprentissage et d'en tirer la formation impartiale pour la minimisation, un ensemble de retenue est générée à partir des données en divisant respectivement chaque ensemble d'apprentissages en 80% utilisés pour la formation des SVM et 20% utilisés pour la minimisation des deux paramètres.

Dans le contexte de la modélisation de l'arrière-plan, Lin et al. ont utilisé 100 images de taille  $160 \times 120$  avec des arrière-plans connus. Chaque image est divisée en blocs de taille  $4 \times 4$  et deux caractéristiques ont été utilisées : le flux optique et la différence temporelle. Un bloc de l'image est classifié comme arrière-plan si la probabilité de sortie est supérieure à un certain seuil  $T$  fixé de manière empirique.

### 16. Les séparateurs à vaste marge pour la régression (SVR) :

Les séparateurs à vaste marge pour la régression appelés aussi les machines à vecteurs supports séparés SVR (*Support Vector regression* en anglais) pour modéliser chaque pixel d'arrière plan en fonction de l'intensité.

L'utilisation des Séparateurs à vastes marges étendue au cas de la régression constitue une alternative prometteuse pour la modélisation des systèmes complexes.

Étant donné un ensemble de données d'apprentissage, le SVR adopte une fonction en spécifiant une limite supérieure sur une fraction des données d'apprentissage autorisées à se situer en dehors d'une distance  $s$  à partir de l'estimation de la régression. Pour le cas de la régression linéaire, la variable à prédire  $y$ , est un nombre réel. L'utilisation des SVM pour la régression consiste à utiliser hyperplan séparateur optimal pour prédire  $y$ .

Pour chaque pixel appartenant à l'arrière-plan, un SVR distinct est utilisé pour le modéliser en tant qu'une fonction d'intensité. Pour classer un pixel donné comme arrière-plan ou non, Wang et Al [WAN 08]. Proposent de fournir la valeur de l'intensité de chaque pixel au SVR associé, ensuite d'appliquer un seuil pour la sortie du SVR.

Soit l'ensemble  $D$  de  $N$  paires de données d'un pixel  $p$  de l'image, ayant un vecteur de données  $X$  comme entrée et l'étiquette  $Y$  de ce vecteur comme sortie.

$$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_i, y_i), \dots, (x_N, y_N)\},$$

$$\text{pour } i = 1, \dots, N \text{ avec } x_i \in R^p, z_i \in R \quad (2.29)$$

Où  $x_i$  correspond à la valeur d'intensité du pixel  $p$  dans le frame  $i$ , et  $y_i$  correspond au degré de confiance du pixel  $p$  étant un pixel d'arrière-plan. Une fois l'apprentissage du SVR est achevé, le degré de confiance du pixel  $p$  dans un nouveau frame  $i$ ,  $f(x_i)$ , est calculé selon la fonction de régression linéaire suivante :

$$f(x_i) = \sum_{j=1}^N (a_i - a'_j) k(x_i, x_j) + \varepsilon \quad (2.30)$$

Où  $k(x_i, x_j)$  est une fonction noyau. Les paramètres  $a$ ,  $a'$  et  $s$  appelés multiplicateurs de Lagrange, sont obtenus par la résolution d'un problème d'optimisation en utilisant le procédé des multiplicateurs de Lagrange. Compte tenu du modèle de l'arrière-plan à base de SVR, l'intensité de chaque pixel dans un nouveau frame est utilisée pour la construction du vecteur d'entrée du SVR. La sortie de la SVR représente le degré de confiance qu'un pixel donné appartient à l'arrière-plan.

Donc, un pixel est étiqueté comme arrière-plan si son degré de confiance est compris entre deux seuils  $S_L$  et  $S_H$  selon l'équation suivante :

$$I_f = \begin{cases} 0 & \text{Si } S_L < f(x_i) < S_H \\ 1 & \text{sinon} \end{cases} \quad (2.31)$$

Le résultat de cette opération est une image binaire *I* des objets en mouvement. La mise à jour du modèle est effectuée selon l'algorithme d'apprentissage en ligne du SVR.

## 17. Les séparateurs à vaste marge pour la description de données (SVDD):

Les séparateurs à vaste marge pour la description de données (appelée en anglais support vector data description) c'est une approche qui permet de déterminer la frontière autour d'un ensemble qui peut être utilisées pour détecter de nouvelles données n'appartenant pas à l'ensemble de départ.

Les séparateurs à vaste marge pour la description de données sont utilisés pour la modélisation de l'arrière plan (proposée par Tavakkoli et Al) [TAV 06].

La représentation d'une frontière autour d'une donnée se fait par l'utilisation d'une sphère qui est caractérisé par un centre  $\alpha$  et un rayon  $R > 0$ .

La minimisation du volume de la sphère est assurée par la réduction de la valeur du rayon  $R^2$  tous en gardant la tonalité des données d'apprentissage  $x_i$  à l'intérieure de la sphère.

$$F(R, \alpha) = R^2 \quad \text{Avec} \quad \|x_i - \alpha\|^2 \leq R^2, \forall i \quad (2.32)$$

La distance entre la valeur  $x_i$  et le centre  $a$  ne doit pas être strictement inférieur à  $R^2$ , pour trouver les valeurs atypiques dans l'ensemble d'apprentissage, l'introduction de la variable négligeable  $\varepsilon_i \geq 0$  permet de changer le problème de minimisation avec l'équation suivante :

$$F(R, a) = R^2 + C \sum_i \varepsilon_i \quad (2.33)$$

Avec

$$\|x_i - a\|^2 \leq R^2 + \varepsilon_i, \varepsilon_i \geq 0 \forall i \quad (2.34)$$

$C$  est le paramètre de confiance qui permet de jouer le rôle d'un arbitre entre la simplicité du système.

On peut écrire l'équation précédente de la manière suivante :

$$L(R, a, a_i, \gamma_i, \varepsilon_i) = R^2 + C \sum_i \varepsilon_i - \sum a_i \left\{ R^2 + \varepsilon_i - \left( \|x_i\|^2 - 2 \cdot x_i + \|a\|^2 \right) \right\} - \sum \gamma_i \varepsilon_i \quad (2.35)$$

Avec des multiplicateurs de Lagrange  $a_i \geq 0$  et  $\gamma_i \geq 0$ ,  $L$  doit être minimisé en respectant les paramètres  $R$ ,  $a$ ,  $\varepsilon_i$  et maximiser suivant les paramètres  $a_i$  et  $\gamma_i$ .

Avec la supposition que  $0 \leq a_i \leq C$ , le multiplicateur de Lagrange  $\gamma_i$  peut être facilement supprimé.

Pour résoudre le problème d'optimisation nous ramène à l'équation suivante :

$$L = \sum_i a_i (x_i \cdot x_i) - \sum_{i,j} a_i a_j (x_i \cdot x_j) \quad (2.36)$$

Avec

$$0 \leq a_i \leq C, \forall a_i \quad (2.37)$$

Si une nouvelle donnée satisfait l'équation précédente, alors son multiplicateur de Lagrange est exprimé par  $a_i \geq 0$  sinon il est mis à 0.

$$\|x_i - a\|^2 < R^2 \rightarrow a_i = 0 \text{ et } \gamma_i = 0 \quad (2.38)$$

$$\|x_i - a\|^2 = R^2 \rightarrow 0 < a_i < C \text{ et } \gamma_i = 0 \quad (2.39)$$

$$\|x_i - a\|^2 > R^2 \rightarrow a_i = C \text{ et } \gamma_i > 0 \quad (2.40)$$

Le centre de la sphère est une combinaison linéaire des objets  $x_i$ , alors nous avons besoin uniquement des données avec des  $a_i \geq 0$  la description des données, Et c'est pour ça que l'on appelle des SVM de description des données.

$$\|z - a\|^2 = (z \cdot z) - 2 \sum_i a_i (z \cdot x_i) + \sum_{i,j} a_i a_j (x_i \cdot x_j) \leq R^2 \quad (2.41)$$

Tavakkoli et Al ont utilisé cette méthode pour construire une descriptive des frontières de chaque pixel dans l'apprentissage pour la construction du modèle de l'arrière plan.

L'une des avantages de cette modélisation est ce modèle consomme moins de mémoire par rapport aux autres méthodes non paramétriques.

### 18. Simple gaussienne généralisée :

Parmi les méthodes de modélisation de l'arrière plan le modèles simple gaussienne généralisée qui est proposé par Kim et Al.

L'objectif de ce modèle est de proposer une solution pour les problèmes liés aux changements dans l'arrière plan et les ombres [KIM 08].

L'idée derrière cette approche est de prendre en considération que la variance des pixels suit la loi de Laplace au lieu d'une gaussienne.

La variance des pixels est plus proche d'une modélisation Laplace qu'une modélisation gaussienne dans les scènes avec de longues périodes statiques de l'arrière plan.

L'équation suivante définit la variation d'un pixel :

$$P(x_t) = \frac{\rho\gamma}{2\Gamma(\frac{1}{\rho})} e^{-(\gamma^\rho |x-\mu|^\rho)} \quad (2.42)$$

Avec

$$\gamma = \frac{1}{\sigma} \left( \frac{\Gamma(\frac{3}{\rho})}{\Gamma(\frac{1}{\rho})} \right) \quad (2.43)$$

$\Gamma(\cdot)$  est une fonction gamma et  $\sigma^2$  est la variance de la distribution.

Si  $\rho=1$  alors il représente une distribution Laplace, Si  $\rho=2$  alors il représente une distribution gaussienne.

Les modèles sont décidés pour chaque pixel en calculant le coefficient d'apprentissage  $g_2$  des premiers m frames.

Les paramètres optimaux du modèle d'arrière plan sont estimés par la maximisation de la vraisemblance de la valeur observée :

$$g_2 = \frac{N \sum_{i=1}^N (x_i - \mu)^4}{(\sum_{i=1}^N (x_i - \mu)^2)^2} - 3 \quad (2.44)$$

### 19. Mélange de gaussienne généralisée :

Allili et al ont proposé un modèle basé sur les mixtures de gaussienne généralisée (GGMM) pour une meilleure segmentation dans les environnements des bruits [ALL 07].

Chaque pixel est caractérisé par son intensité dans l'espace couleur RGB (red, green, bleu).

La probabilité d'observer la valeur du pixel courant est considérée dans le domaine multidimensionnel avec :

$$P(x_t) = \sum_{i=1}^k \omega_{i,t} \cdot \eta(x_t, \mu_{i,t}, \sigma_{i,t}, \gamma_i) \quad (2.45)$$

$$\eta(x_t, \mu_{i,t}, \sigma_{i,t}, \lambda_i) = \prod_{j=1}^d A(\lambda_i) \exp\left(-B(\lambda_i) \left|\frac{x_j - \mu_j}{\sigma_j}\right|^{\lambda_j}\right) \quad (2.46)$$

$$A(\lambda) = \frac{\left(\frac{\Gamma(\frac{3}{\lambda})}{\Gamma(\frac{1}{\lambda})}\right)^{1/\lambda}}{2\sigma\Gamma(\frac{1}{\lambda})} \text{ et } B(\lambda) = \left(\frac{\Gamma(\frac{3}{\lambda})}{\Gamma(\frac{1}{\lambda})}\right) \quad (2.47)$$

K est le nombre de distributions attribuées à chaque pixel,  $\omega_{i,t}$  est le poids associé à  $i^{\text{ème}}$  gaussienne dans le temps t,  $\mu_{i,t}$  est la moyenne,  $\sigma_{i,t}$  est l'écart type et  $\eta$  est la fonction de densité de probabilité.

De même dans une simple gaussienne généralisé, le paramètre  $\lambda$  permet de désigner le modèle utilisé.

Si  $\lambda_i=0$  alors la distribution est de type gaussienne, alors que si  $\lambda_i=3$  alors la distribution est de type Laplace.

La mise à jour des paramètres de la GGMM se faite de la même manière que dans un modèle GMM.

## 20. Conclusion :

Dans ce chapitre nous avons présenté les différentes méthodes de la modélisation de l'arrière, nous avons commencé par la dérivée temporelle qui a la réputation d'être peu robuste face à des phénomènes tels que les mouvements lents ou saccadés, nous avons également présenté le calcul du flot optique qui présente l'inconvénient d'être très couteuse en temps de calcul. La dernière partie a été dédiée à la modélisation de l'arrière-plan par un mélange de lois gaussiennes.

*Mais le système des techniques -  
Analyse Inconvénient*



# **CHAPITRE III**

## **Traitement dynamique des modèles de l'arrière plan**

## 1. Introduction :

Après avoir présenté un état de l'art sur les systèmes de vidéosurveillance et sur les différents techniques et algorithmes utilisés pour aboutir à un traitement automatique des informations, nous allons, dans ce chapitre, mettre les projecteurs sur les contributions apportées à ce domaine à travers une architecture détaillée d'un système de détection et de suivi des objets en mouvement dans une vidéo. Nous avons proposé des solutions au niveau de la modélisation de l'arrière-plan en proposant une amélioration des modèles de mixtures de gaussiennes. Nous avons également proposé l'utilisation du filtre de Canny pour réduire l'espace de recherche. La figure suivante représente l'architecture de notre système.

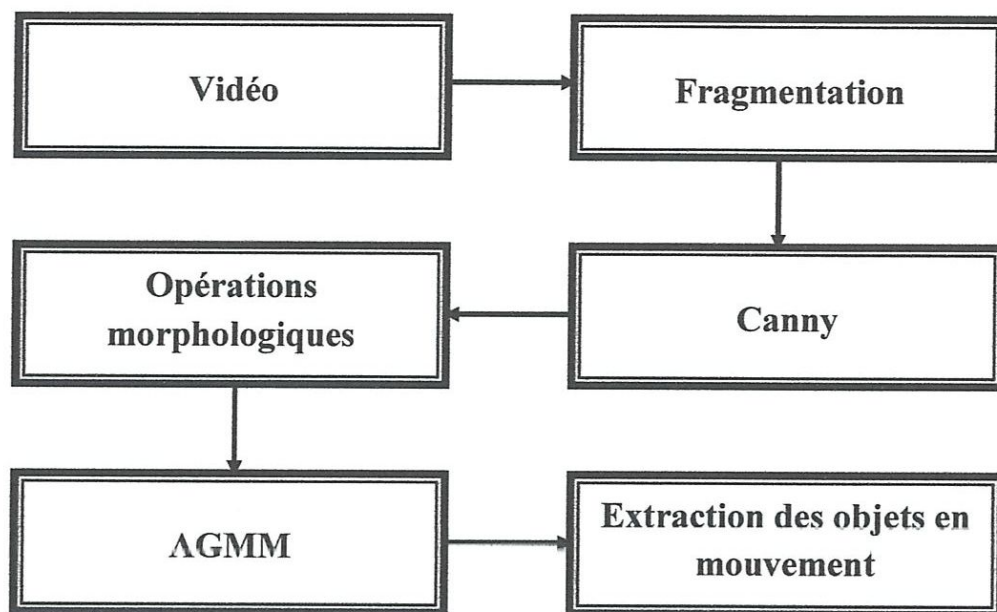


Figure 3.1 : Architecture du système.

## 2. Vidéo :

La vidéo regroupe l'ensemble des techniques permettant l'enregistrement ainsi que la restitution d'images animées, accompagnées ou non de son, sur un support électronique et non de type photochimique.

## 3. Fragmentation :

La fragmentation d'une vidéo consiste à découper la vidéo en séquence d'image, plutôt que de travailler sur le temps exploité par la vidéo. Cela consiste à convertir le temps exploité par la vidéo au nombre d'images correspondantes pour aboutir finalement à une base d'images fixes RGB extraites à partir de la séquence vidéo ou

chaque image a son temps correspondant. La fragmentation se fait à l'aide de l'algorithme suivant :

**Fragmenter vidéo**

NB\_Image = temps exploite par la vidéo / temps exploite par une image.

Tant que ( $i \leq \text{NB\_Image}$ ) faire

Capturer l'image media N i.

Mettre l'image dans un buffer média.

Convertir le buffer média en une image.

Enregistrer l'image.

**4. Canny :**

L'algorithme de Canny est un algorithme de détection du contour, le but de la détection de contours est de repérer les points d'une image numérique qui correspondent à un changement brutal de l'intensité lumineuse.

Un contour peut être défini comme une marche d'escalier si le contour est net, comme il peut être rampe si le contour est plus et comme un toit s'il s'agit d'une ligne sur un fond uniforme, voir figure 1.

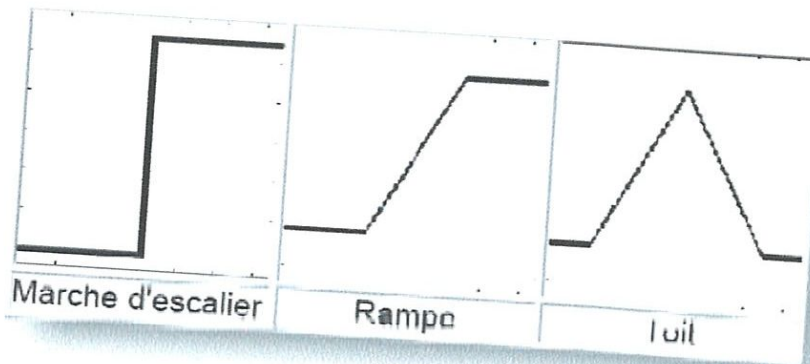


Figure 3.2 : Différentes forme de contour.

Le principe général de la méthode de Canny implémente une estimation du gradient de l'image à l'aide du filtre de Sobel, suivi d'un seuillage par hystérésis du module de gradient. Un seuil haut et un seuil bas sont à définir.

Cependant, ces contours peuvent être dus au bruit restant ou encore aux variations de couleur.

Le moyen le plus rapide et le plus efficace pour se débarrasser de ces faux contours est d'utiliser un double seuillage.

- tous les pixels supérieurs au seuil maximal défini sont des contours forts.
- tous les pixels inférieurs au seuil minimal défini sont supprimés.
- tous les pixels situés entre ces deux valeurs sont considérés comme des contours faibles.

- **Détection de contour par hystérésis:** Les contours sont détectés par la suppression de tous les contours qui ne sont pas liés à au moins à un contour très visible (contour certain). Durant cette étape, nous ajoutons tous les contours forts définis lors du double seuillage à l'image finale. Ensuite nous parcourons un à un les contours faibles:
  - Si un contour faible est relié à un contour fort, il est ajouté à l'image finale.
  - Sinon le contour faible est supprimé [S9].

### ❖ Les avantages de l'algorithme de Canny :

Cet algorithme a été conçu pour être optimal suivant trois critères :

- Bonne détection : faible taux d'erreur dans la signalisation des contours.
- Bonne localisation : minimisation des distances entre les contours détectés et les contours réels.
- Clarté de la réponse : une seule réponse par contour et pas de faux positifs [BEL 12].

## 5. Opération morphologie mathématique :

La morphologie mathématique est une théorie et technique mathématique et informatique d'analyse de structures qui est liée avec l'algèbre, la topologie et les probabilités.

Le développement de la morphologie mathématique est inspiré des problèmes de traitement d'images, domaine qui constitue son principal champ d'application. Elle fournit en particulier des outils de filtrage, segmentation, quantification et modélisation d'images.

Les opérations morphologies mathématiques travaillent sur le voisinage local de chaque pixel, la forme de ce voisinage est appelée élément structurant.

Parmi les opérations morphologies mathématiques on distingue la dilatation et l'érosion :

- **La dilatation** : après l'application de l'algorithme de canny (détection de contour) le contour obtenue est mal formés contient plusieurs coupures et pour lier ces coupures on utilise la dilation, on parcourt l'image pixel par pixel on cherche les pixels qui ont la valeur 0 si ce pixel a 2 voisins (1) il va changer sa valeur a 1 si non il reste le même,

Voila des exemples :

1 : signifie la couleur noir.

0 : signifie la couleur blanche.

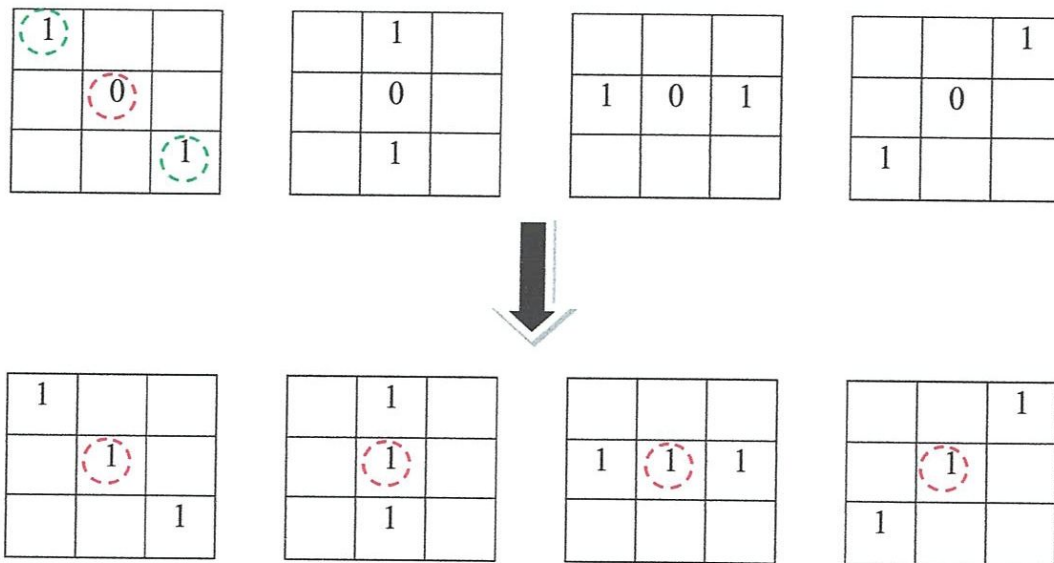


Figure 3.3 : Masques utilisés dans la dilation.

- **L'érosion** : après la dilation certains pixels auront plusieurs voisins (1), on parcourt l'image pixel par pixel on travaille sur les pixels qui vaut 1 si ce pixel a deux voisins (1) le contour est bien formé si il a plus de deux ont les élimine et on laisse juste deux.

Voila un exemple :

1 : signifie la couleur noir.

0 : signifie la couleur blanche.

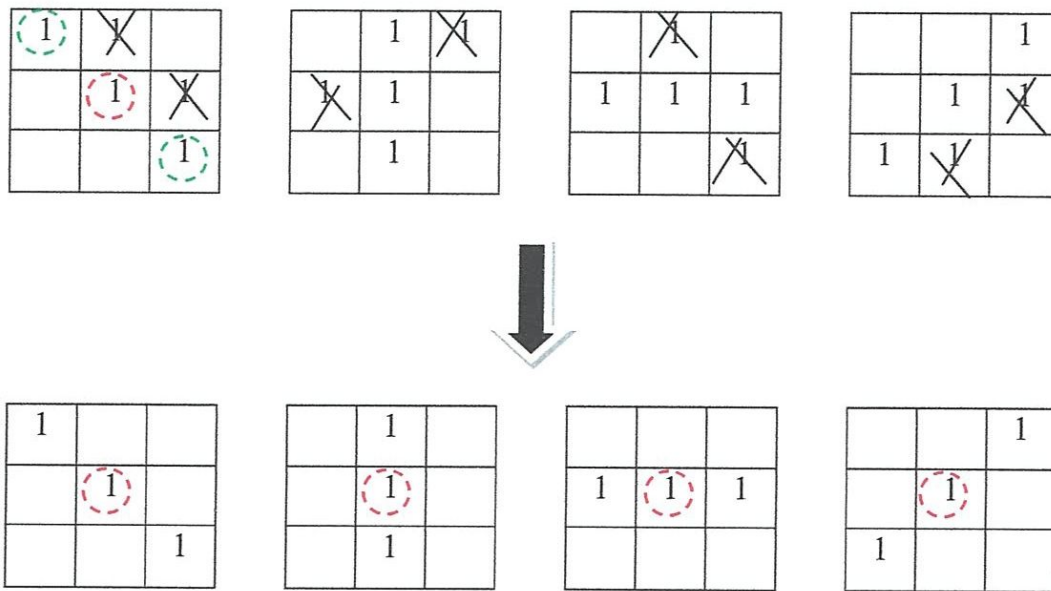


Figure 3.4 : Masques utilisés dans l'érosion.

## 6. AGMM :

Stauffer et Grimson (Stauffer 1999) ont présenté une nouvelle modélisation de l'arrière-plan basée sur une mixture de plusieurs gaussiennes. Les MoG (Mixture Of Gaussians) ont prouvé leurs efficacités face aux changements de luminosité, les mouvements répétitifs, l'encombrement, l'introduction et le retrait des objets de la scène et les mouvements lents des objets. La motivation sur laquelle est basée la proposition de ce modèle est qu'un modèle unimodal de l'arrière-plan ne pouvait gérer le bruit d'acquisition d'image, le changement de lumière et l'existence de multitude de surfaces pour un pixel particulier dans le même temps.

Une MoG est un modèle statistique qui suppose que les données proviennent d'une somme pondérée de plusieurs distributions de gaussiennes. Stauffer et Grimson (Stauffer 1999) ont présenté une méthode GMM adaptative pour modéliser un fond dynamique dans des séquences d'images. Si les K distributions de gaussiennes sont utilisés pour décrire l'historique d'un pixel, l'observation d'un pixel donné sera forcément dans l'un des K états à la fois.

Où K détermine la multimodalité de l'arrière-plan et la sélection de K est généralement basé sur la mémoire disponible et de la puissance de calcul. Stauffer et Grimson ont proposé de fixer K entre 3 et 5.

Dans cette méthode nous allons voir deux parties la partie initialisation et la partie mis a jour, tout d'abord nous allons commencer avec la première partie :

- **Initialisation :**

Chaque pixel est caractérisé par son intensité dans l'espace de couleurs HSV. Ensuite, la probabilité d'observer la valeur du pixel courant  $X_t$  est donnée par l'équation suivante dans le cas multidimensionnel.

$$P(X_t) = \sum_{i=1}^k w_{i,t} \eta(X_t, \mu_{i,t}, \Sigma_{i,t}) \quad (3.5)$$

$X_t$  la valeur d'un pixel au temps t, k est le nombre de gaussienne, et les paramètres d'une gaussienne  $w_{i,t}$ ,  $\mu_{i,t}$  et  $\Sigma_{i,t}$  sont respectivement le poids, la moyenne et la matrice de covariance de la  $i^{\text{ème}}$  gaussienne dans le temps t.

$$\eta(X_t, \mu, \Sigma) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(X_t - \mu)^T \Sigma^{-1} (X_t - \mu)} \quad (3.6)$$

$\eta$  est la fonction de densité de probabilité gaussienne:

Au début du processus d'extraction de l'arrière-plan, pour chaque pixel, les distributions des K gaussiennes sont initialisées avec une moyenne prédéfinie, une variance élevée et un faible poids. Pour obtenir un arrière-plan stable, nous avons utilisé les 50 premiers frames pour l'apprentissage des mixtures de gaussiennes.

- **Mise à jour des paramètres :**

Il existe plusieurs techniques disponibles pour estimer les paramètres d'une GMM, la méthode la plus populaire et la plus utilisée pour l'estimation est le maximum de vraisemblance (ML). Le but de l'estimation ML est de trouver les paramètres du modèle qui maximisent la vraisemblance du GMM sur un ensemble de données d'apprentissage.

Pour une séquence de T vecteurs d'apprentissage de  $X = \{x_1, \dots, x_T\}$ , si on suppose qu'il existe une indépendance entre les vecteurs, alors la vraisemblance du GMM peut être écrite selon l'équation suivante.

$$P(X | \lambda) = \prod_{t=1}^T p(X_t | \lambda) \quad (3.7)$$

Malheureusement, cette expression est une fonction non linéaire des paramètres  $\lambda$  et la maximisation directe est impossible. Cependant, les estimations des paramètres ML peuvent être obtenues en utilisant de manière itérative un cas particulier de l'algorithme espérance-maximisation (EM). L'idée de base de l'algorithme EM est, en commençant par un modèle

initial  $\lambda$ , pour estimer un nouveau modèle  $\lambda'$ , tel que  $P(X|\lambda') \geq p(X|\lambda)$ . Le nouveau modèle devient alors le modèle initial pour la prochaine itération et le processus est répété jusqu'à ce qu'un seuil de convergence soit atteint. Le modèle initial est généralement obtenu en utilisant une forme binaire de l'histogramme. Pour chaque itération d'EM, la ré estimation des formules suivantes est utilisée garantissant une augmentation monotone de la valeur de vraisemblance du modèle

$$\bar{w}_i = \frac{1}{T} \sum_{t=1}^T P'(i|X_t, \lambda) \quad (3.8)$$

$$\bar{\mu}_i = \frac{\sum_{t=1}^T P'(i|X_t, \lambda) X_t}{\sum_{t=1}^T P'(i|X_t, \lambda)} \quad (3.9)$$

$$\bar{\sigma}_i^2 = \frac{\sum_{t=1}^T P'(i|X_t, \lambda) X_t^2}{\sum_{t=1}^T P'(i|X_t, \lambda)} - \bar{\mu}_i^2 \quad (3.10)$$

Où  $\bar{w}_i$ ,  $\bar{\mu}_i$ ,  $\bar{\sigma}_i^2$  représente respectivement les nouvelles valeurs du poids, de la moyenne et de la variance. La probabilité a posteriori pour le composant  $i$  est donnée par l'équation suivante :

$$P'(i|X_t, \lambda) = \frac{w_i \eta(X_t, \mu_{i,t}, \Sigma_{i,t})}{\sum_{k=1}^M w_k \eta(X_t, \mu_{k,t}, \Sigma_{k,t})} \quad (3.11)$$

Malgré la réputation de l'algorithme ML ce dernier occupe énormément d'espace mémoire et nécessite une puissance de calcul non négligeable ce qui rend son utilisation inappropriée dans un environnement en temps réel ou d'un système embarqué. En plus, EM converge vers un maximum local, mais nous n'avons aucune garantie que ce maximum soit « bon », c.-à-d. soit proche en valeur du maximum global. Pour se rapprocher de ce but, on doit encore alourdir le calcul.

Enfin, EM présente le risque de trouver des solutions « dégénérées » présentant au moins une composante ajustée sur 1 seul individu, alors que la vraisemblance d'une telle gaussienne peut être infinie, mettant ainsi à mal notre algorithme.

Pour une éventuelle application en temps réel, nous avons utilisé l'algorithme d'approximation K-Means, pour la mise à jour du modèle. Chaque valeur d'un nouveau pixel est vérifiée, à travers les  $k$  distributions de gaussiennes, pour déterminer si la distance entre le pixel et chacune des gaussiennes sont inférieure à l'écart type de distribution selon l'équation suivante :



$$\frac{|P_t - \mu_i|}{\sigma_i} < 2.5 \quad (3.12)$$

**Phase 1 :**

Si aucune distribution ne vérifie l'équation (3.12), alors le pixel est associé à l'avant-plan. Dans ce cas, nous allons ajouter une nouvelle distribution en attribuant à la gaussienne une moyenne, une variance et un poids selon les équations (3.13), (3.14), (3.15) :

$$\sigma_k^2 = \text{une Grande variance initiale} \quad (3.13)$$

$$w_k = \text{faible poids} \quad (3.14)$$

$$\mu_k = X_t \quad (3.15)$$

Si le nombre de gaussienne est maximum alors

Remplacer la gaussienne qui a le poids minimum par la nouvelle gaussienne.

**Phase 2 :**

Si non s'il existe au moins une distribution qui vérifie l'équation (3.12), les poids de toutes les gaussiennes seront mis à jour selon l'équation suivante :

$$w_{k,t} = (1 - \alpha) \cdot w_{k,t-1} + \alpha M_{k,t} \quad (3.16)$$

Où

$\alpha$  : est le coefficient d'apprentissage qui détermine la vitesse d'adaptation de modèle.

$M_{k,t}$  est égal à 1 pour la gaussienne qui correspond à l'arrière-plan et égal à 0 pour les autres. Après la mise à jour des poids, une étape de normalisation est effectuée pour garantir que la somme des poids est toujours égale à 1. Les autres paramètres des distributions ayant vérifié l'équation sont mis à jour selon les équations (3.17), (3.18), (3.19) :

$$\mu_{k,t} = (1 - \varphi_k) \cdot \mu_{k,t-1} + \varphi_k \cdot P_t \quad (3.17)$$

$$\sigma_{k,t}^2 = (1 - \varphi_k) \cdot \sigma_{k,t-1}^2 + \varphi_k (P_t - \mu_{k,t})^T (P_t - \mu_{k,t}) \quad (3.18)$$

Où

$$\varphi_t = \alpha\eta(P_t/\mu_k\sigma_k) \quad (3.19)$$

La suppression d'une gaussien se faite selon deux paramètres le poids et l'âge telle que on cherche la gaussienne qui a le poids minimum et l'âge maximum puis on la supprime.

### 7. Extraction des objets en mouvement : *(Supprimer les candidats)*

Pour décider si  $X_t$  est inclus dans l'arrière-plan, les distributions sont ordonnées selon la valeur  $w_{k,t}/\sigma_{k,t}$ . Cet ordre utilise l'hypothèse selon laquelle un pixel d'arrière-plan correspond à un poids élevé avec une variance faible pour la raison que l'arrière-plan est pratiquement constant et il est plus présent que les objets en mouvement. Les premières B distributions qui vérifient l'équation (3.16) sont sélectionnés pour représenter l'arrière-plan.

$$B = \arg \min(\sum_{k=1}^b w_{k,t} > T) \quad (3.16)$$

Le seuil T représente la portion minimum du poids total attribué au modèle d'arrière-plan. Si une petite valeur T est choisie, l'arrière-plan devient unimodal. Si T est plus élevée, une distribution multimodale causée par un mouvement répétitif de l'arrière-plan peut résulter de plusieurs objets de l'arrière-plan ce qui permet aux pixels de l'arrière-plan d'accepter plus d'une distribution gaussienne. L'utilisation d'un seuil T unique pour le modèle engendre une mauvaise classification surtout lorsque la scène contient à la fois des zones dynamique et statique. Un seuil plus élevé permet d'avoir une meilleure classification dans un environnement dynamique, mais génère des détections incorrectes pour les objets en mouvement dans un arrière-plan statique. La figure suivante montre un simple processus de classification où les pixels accumulés définissent l'arrière-plan représenté par les distributions de gaussiennes tandis que les pixels dispersés sont classés comme premier plan.

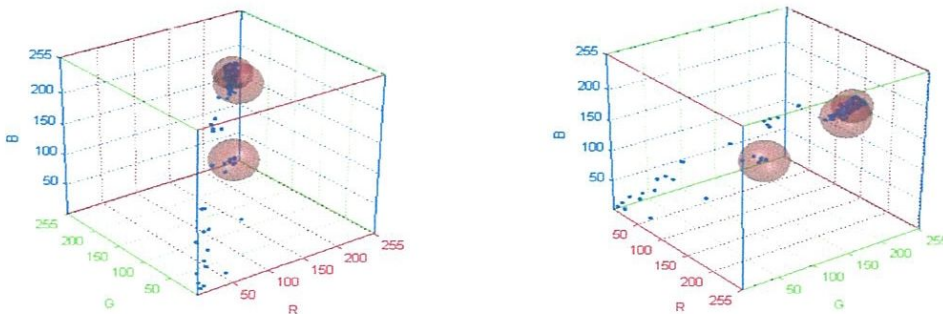


Figure 3.5: Processus de classification.

## 8. Conclusion :

Dans ce chapitre nous avons présenté l'architecture générale de notre système ainsi que les différents modules qui le composent. Nous avons pris le soin de détailler chaque processus afin de montrer notre apport à travers les idées implanté dans presque tous les niveaux du système proposé.

# **CHAPITRE IV**

## **Testes et résultats**

## 1. Introduction :

Pour valider les approches proposées, une étape de tests est indispensable. L'analyse de cette dernière va nous permettre de peser les pour et les contre et de détecter les points forts et les faiblesses des algorithmes utilisés. Afin d'expérimenter le système, nous allons d'abord effectuer une implémentation des algorithmes et des modules du système proposé. Cette implémentation est réalisée à l'aide d'un langage de programmation et sur une plateforme bien spécifique. L'expérimentation nécessite également la présence d'une base de tests à travers lequel nous allons mesurer la performance du système.

## 2. Outils utilisés:

### 2.1 JAVA :

Java est un langage de programmation et une plate-forme informatique qui ont été créés par *Sun Microsystems* en 1995. Beaucoup d'applications et de sites Web ne fonctionnent pas si Java n'est pas installé et leur nombre ne cesse de croître chaque jour. Java est rapide, sécurisé et fiable [S10].

On peut faire de nombreuses sortes de programmes avec Java des applications sous forme de fenêtre ou de console, des applets qui sont des programmes Java incorporés à des pages web, des applications pour appareils mobiles avec J2ME etc[CYS 13].

Les applications Java peuvent être exécutées sur tous les systèmes d'exploitation pour lesquels a été développée une plateforme Java, dont le nom technique est JRE (Java Runtime Environment -Environnement d'exécution Java). Cette dernière est constituée d'une JVM (Java Virtual Machine- Machine Virtuelle Java), le programme qui interprète le code Java et le convertit en code natif.

Mais le JRE est surtout constitué d'une bibliothèque standard à partir de laquelle doivent être développés tous les programmes en Java.

### 2.2 Eclipse :

Eclipse IDE est un environnement de développement intégré libre (le terme Eclipse désigne également le projet correspondant, lancé par IBM) extensible, universel et polyvalent, permettant potentiellement de créer des projets de développement mettant en œuvre n'importe quel langage de programmation. Eclipse IDE est principalement écrit en Java (à l'aide de la bibliothèque graphique SWT, d'IBM), et ce langage, grâce à des bibliothèques spécifiques, est également utilisé pour écrire des extensions [HUT 08].

### 2.3 JMF (Java Media Framework) :

JMF est une API (Application Programming Interface) Java permettant de manipuler aisément toutes sortes de contenus multimédia avec Java tels que du son ou de la vidéo. Elle offre les outils nécessaires pour faire de l'acquisition, du traitement et du transport de médias basés sur le temps. L'avantage est de pouvoir concevoir des applications utilisant des éléments multimédias (Webcam, micro, vidéos...) et pouvant s'exécuter sur différentes plates-formes logicielles (principalement Windows et Linux).

### 3. Base de vidéo :

Nom de vidéo	Nombre de frame	Résolution	Description
Bootstrap	3054	160×120	Des personnes en train de faire des commandes dans une cafeteria.
Camouflage	352	160×120	Une personne marche devant un écran ordinateur avec des barres d'interférence.
ForegroundAperture	2112	160×120	Une personne sommeil après il réveil en bureau.
LightSwitch	2714	160×120	Changement de lumière dans chambre contienne un ordinateur allumé.
MovedObject	1744	160×120	Un objet déplacé en arrière-plan dans une salle de conférence.
TimeOfDay	5889	160×120	Changement de lumière progressive dans Easy Living lab
WavingTrees	286	160×120	Déplacement d'un arbre en arrière-plan.

*Tableau 4.1 : Base des vidéos.*

### 4. Testes et résultats :

#### 4.1 Résultat de l'algorithme de fragmentation :

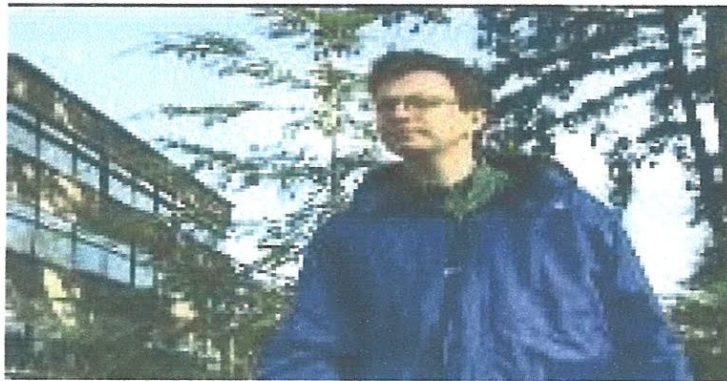
Après l'exécution de l'algorithme de fragmentation qui sert a découper une vidéo en séquence d'image on obtient le résultat suivant :



*Figure 4.1 : Vidéo fragmentée.*

## 4.2 Résultat de l'algorithme de canny :

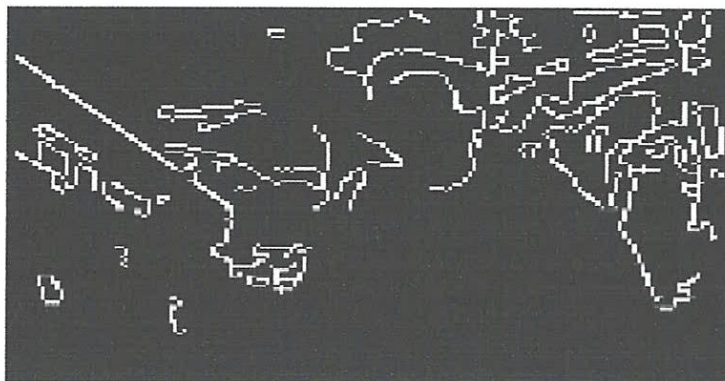
Comme nous avons vu dans le chapitre trois l'algorithme de canny consiste à détecter le contour d'une image en niveaux de gris voila le résultat de cet algorithme :



*Figure 4.2* : Image originale.



*Figure 4.3* : Image en niveaux de gris.



*Figure 4.4* : Image avec l'algorithme de canny.

## 4.3 Résultat des opérations morphologiques :

Parmi les opérations morphologiques nous avons travaillé avec la dilation et l'érosion le résultat chaque opération est montrée par les figures suivantes :



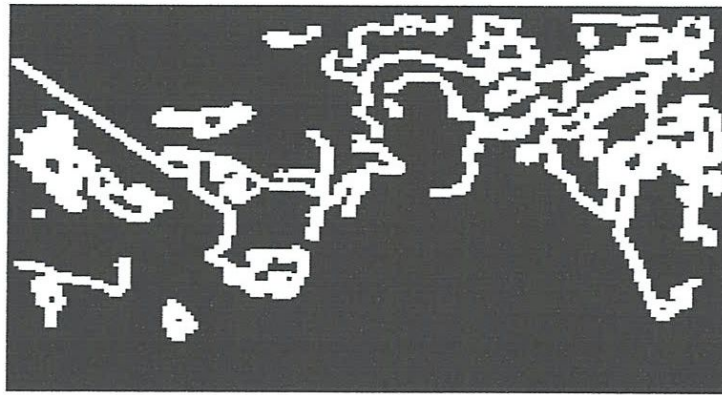


Figure 4.5 : Image après l'application des opérations morphologiques.

### 5. Résultat d'AGMM :

L'exécution de cette partie permet de créer deux fichiers base de séparation et base de dilatation et nous donne le résultat montré dans la figure suivante :


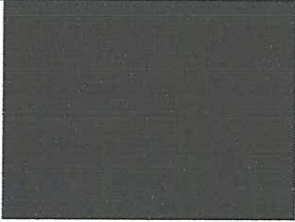
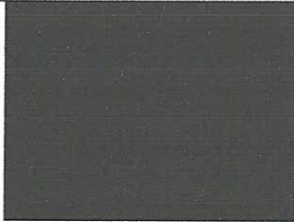

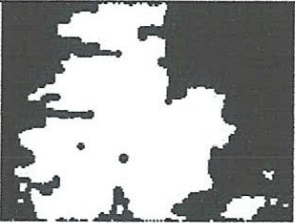


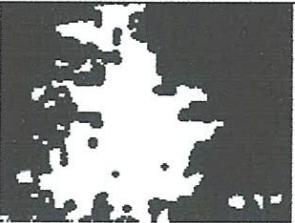
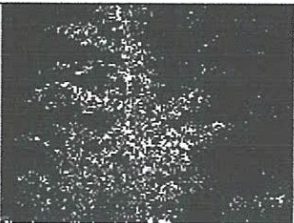
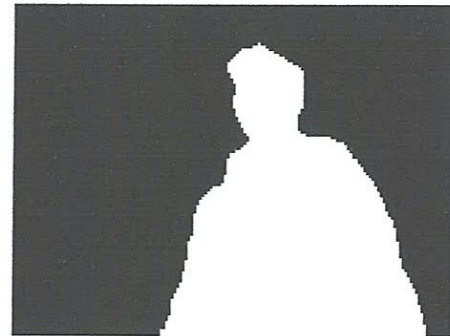
		
		
		
Images originales	Dilatation	Séparation

Figure 4.6 : Echantillons d'image de dilatation et de séparation.

Dans cette figure on peut voir que les objets en mouvements sont bien détectés.



(a)



(b)

Figure 4.7 : WavingTrees(a) image originale, (b) image segmenté a la main.

Nbr_Gaus_Min = 3, Nbr_Gaus_Max = 9	Nbr_Gaus_Min = 4, Nbr_Gaus_Max = 9	Nbr_Gaus_Min = 5, Nbr_Gaus_Max = 9
Nbr_Gaus_Min = 3, Nbr_Gaus_Max = 10	Nbr_Gaus_Min = 4, Nbr_Gaus_Max = 10	Nbr_Gaus_Min = 5, Nbr_Gaus_Max = 10
Nbr_Gaus_Min = 3, Nbr_Gaus_Max = 11	Nbr_Gaus_Min = 4, Nbr_Gaus_Max = 11	Nbr_Gaus_Min = 5, Nbr_Gaus_Max = 11
Résultats : le meilleur résultat est obtenu lorsque le nombre de gaussiennes varie entre 4 et 10.		

Figure 4.8 : Résultats des testes avec changement de paramètre (Waving Trees).



Figure 4.9 : TimeOfDay(a) image originale, (b) image segmenté a la main.

Nbr_Gaus_Min = 3, Nbr_Gaus_Max = 9	Nbr_Gaus_Min = 4, Nbr_Gaus_Max = 9	Nbr_Gaus_Min = 5, Nbr_Gaus_Max = 9
Nbr_Gaus_Min = 3, Nbr_Gaus_Max = 10	Nbr_Gaus_Min = 4, Nbr_Gaus_Max = 10	Nbr_Gaus_Min = 5, Nbr_Gaus_Max = 10
Nbr_Gaus_Min = 3, Nbr_Gaus_Max = 11	Nbr_Gaus_Min = 4, Nbr_Gaus_Max = 11	Nbr_Gaus_Min = 5, Nbr_Gaus_Max = 11
Résultat : le meilleur résultat est obtenu lorsque le nombre de gaussiennes varie entre 5 et 11.		

Figure 4.10 : Résultats des testes avec changement de paramètre (TimeOfDay).

## **6. Conclusion :**

Dans ce chapitre nous avons présenté les différents résultats que nous avons obtenus en intégrant les étapes de détection développées dans le chapitre trois. Nous pouvons dire que les idées proposées ont fortement contribué dans l'amélioration des résultats par rapport au modèle de base et cela dans tous les vidéos testés.

## Conclusion générale

Les idées soutenues dans ce mémoire ont une finalité pratique et un objectif précis, qui sont la conception et l'implémentation d'un système de détection des objets en mouvement à travers une caméra fixe. Pour atteindre cet objectif, nous avons proposé une amélioration d'un algorithme largement rependu dans la littérature. En effet, nous avons proposé de transférer les mixtures de gaussienne de leurs structures d'origine statique vers une structure flexible et dynamique permettant ainsi de refléter la nature réelle des informations traitées.

Le principe de base de notre système repose sur la visualisation du flux vidéo qui arrive sur notre ordinateur et la capture des séquences d'images de la source pour pouvoir les traiter par la suite. Le résultat est un ensemble d'objets en mouvement encapsulé dans des boîtes de suivi.

Les résultats obtenus sont largement acceptables et encourageants, mais nous suggérons de continuer la recherche dans ce sens pour améliorer davantage les résultats.

Comme perspective, le système pourrait être amélioré en augmentant sa vitesse d'exécution, en améliorant son résultat par le biais de nouvelles techniques, ou encore en facilitant son utilisation, car un des problèmes majeurs est le nombre important de variables qui peut jouer un rôle très important dans la reconnaissance de l'objet.

Pour élargir le champ d'action de notre système, travailler avec une caméra amovible pourrait être intéressant. Le problème est que certains algorithmes étudiés seraient inutilisables dans ce cas, en particulier celles qui utilisent un background.

Le suivi, d'objets visuels, reste un sujet de recherche ouvert en raison du manque de système de suivi holistique satisfaisant, prenant en compte les distorsions intrinsèques et extrinsèques, variations d'éclairément, occlusions, bruits et erreurs dans la correspondance et la classification d'objets ne sont qu'une partie des problèmes actuellement rencontrés dans la détection et le suivi des objets en mouvement.

# Bibliographie

[VAL 09] Valérie Gouaillier « La vidéosurveillance intelligente » Rapport de veille technologique et commerciale Mars 2009 Mise à jour 8 avril 2009.

[VER 07] Nicolas verbeke « Suivi d'objets en mouvement dans une séquence vidéo » Thèse présentée à l'université PARIS DESCARTES pour l'obtention du grade de Docteur 2007.

[BRU 12] Mathieu Brulin « Analyse sémantique d'un trafic routier dans un contexte de vidéosurveillance » Thèse présentée à L'UNIVERSITÉ BORDEAUX I ÉCOLE DOCTORALE DE MATHÉMATIQUES ET INFORMATIQUE pour l'obtention du grade de Docteur 2012.

[LME 03] Alexandre Lemieux « SYSTEME D'IDENTIFICATION DE PERSONNES PAR VISION NUMERIQUE » mémoire présenté la Faculté des études supérieures de l'Université Laval QUEBEC pour l'obtention du grade de maître `es sciences (M.Sc.) 2003.

[BOU 12] Hajar BOUIROUGA « Reconnaissance des scènes vidéo pour adulte » Thèse présentée à Université Mohammed V-Agdal Faculté des Sciences Rabat Maroc 2012.

[MAN 12] Antoine Manzanera « Traitement Et Reconnaissance d'Images » Cours TERI – Master 2 UPMC Paris 6 2012.

[GRE 05] Thomas GRENIER « Apport de l'espace des caractéristiques et des paramètres d'échelle adaptatifs pour le filtrage et la segmentation d'image » thèse présenté à L'institut National des Sciences Appliquées de Lyon pour obtenir Le grade de docteur 2005.

[BEL 12] Bellal Salah eddine « reconnaissance d'objets dans une image application au tri/contrôle de qualité » mémoire de magister présenté à Université Hadj Lakhdar Batna Faculté de Technologie Département Génie Industriel pour obtenir diplôme de magister 2012.

[VAP 92] Vladimir Naumovich VAPNIK et Vlamimir VAPNIK. Statistical learning theory. T. 1. Wiley New York, 1992.

[LIN 02] Horng-Horng LIN, Tyng-Luh LIU et Jen-Hui CHUANG. « A probabilistic SVM approach for background scene initialization ». In : Image Processing. 2002. Proceedings. 2002 International Conference on. T. 3. IEEE. 2002, p. 893–896.

- [WAN 08] Huan WANG, Ming-wu REN et Jing-Yu YANG. « Background modeling method based on sequential kernel density approximation ». In : Pattern Recognition, 2008. CCPR'08. Chinese Conference on. IEEE. 2008, p. 1–6.
- [TAV 06] Alireza TAVAKKOLI, Mircea NICOLESCU et George BEBIS. « A novelty detection approach for foreground region detection in videos with quasistationary backgrounds ». In : Advances in Visual Computing. Springer, 2006, p. 40–49.
- [KIM 08] Hansung KIM et al. « Background subtraction using generalised Gaussian family model ». In : Electronics letters 44.3 (2008), p. 189–190.
- [ALL 07] Mohand Said ALLILI, Nizar BOUGUILA et Djemel ZIOU. « Finite generalized Gaussian mixture modeling and applications to image and video foreground segmentation ». In : Computer and Robot Vision, 2007. CRV'07. Fourth Canadian Conference on. IEEE. 2007, p. 183–190.
- [STA 99b] Chris STAUFFER et W Eric L GRIMSON. « Adaptive background mixture models for real-time tracking ». In : Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on. T. 2. IEEE. 1999.
- [CHA 10] Theekapun CHAROENPONG, Ajaree SUPASUTEEKUL et Chaiwat NUTHONG. « Adaptive background modeling from an image sequence by using KMeans clustering ». In : Electrical Engineering/Electronics Computer Telecommunications and Information Technology (ECTI-CON), 2010 International Conference on. IEEE. 2010, p. 880–883.
- [XU 10] Huazhong XU, Pei LV et Lei MENG. « A people counting system based on head-shoulder detection and tracking in surveillance video ». In : Computer Design and Applications (ICCD), 2010 International Conference on. T. 1. IEEE. 2010, p. V1–394.
- [CYS 13] cysboy « Apprenez à programmer en Java » Licence Creative Commons 6 2.0 Dernière mise à jour le 8/01/2013.
- [HUT 08] Stéphane Huot « Programmation Orientée Objet – Java Introduction à Eclipse » Formation initiale 2ème Année 2008 Département Informatique.

# Web graphie

[S1]: <http://www.micros-services.fr/professionnels/votre-video-surveillance>

[S2]: <http://www.ocean-securite.com/les-objectifs-et-le-fonctionnement-des-systemes-de-videosurveillance-c9-p22.html>

[S3]: <http://videosurveillance.e-monsite.com/pages/histoire-de-la-videosurveillance/>

[S4]: <http://docplayer.fr/1403720-Systeme-de-videosurveillance.html>

[S5]: <https://helpx.adobe.com/fr/premiere-pro/using/interlacing-field-order.html>

[S6]: <http://fr.hkvstar.com/technology-news/cctv-tips-tv-lines-analog-vs-pixels-digital-resolution.html>

[S7]: [https://fr.wikipedia.org/wiki/Image\\_de\\_r%C3%A9f%C3%A9rence](https://fr.wikipedia.org/wiki/Image_de_r%C3%A9f%C3%A9rence)

[S8]: <http://www.gralon.net/articles/photo-et-video/photo-et-video/article-les-pixels---definition-et-caracteristiques-2014.htm>

[S9]: <http://edgestracing.aurelie-fruitiere.fr/canny.html>

[S10]: [https://www.java.com/fr/download/faq/whatis\\_java.xml](https://www.java.com/fr/download/faq/whatis_java.xml)