

M/004. 437

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

13/ 825

Option : Ingénierie des Medias

Thème :

Détection de comportements suspects dans une vidéo de surveillance

Encadré par :

Mr Farou Brahim

Présenté par :

Chiheb Amira

Oubraham Sabrina

Juin 2013





Remerciements

Nous tenons tous d'abord à remercier dieu qui nous a donné le pouvoir et la volonté pendant 17ans consécutifs jusqu'à ce jour de fin d'études, tout en lui demandant de nous apporter d'avantage pour en continuer.

Ainsi, on tient à remercier Mr Farou Braslim pour son encadrement, son orientation et ces efforts afin qu'on puisse mener à bien ce projet.

Nos sincères remerciements aux membres de jury de nous faire honneur de juger cette humble mémoire.


Nous adressons notre profonde gratitude à tous l'ensemble du corps professoral et administratif du département d'informatique ayant contribué à notre formation.

En fin, nos remerciements vont à nos familles et à nos amis et à tous ce qui ont contribué de près ou de loin à la réalisation de ce mémoire.



Dédicaces

À tous ceux que j'aime, à tous ceux qui m'aiment, et surtout à tous ceux qui le méritent.



À mes chers parents qui ont toujours eu confiance en moi et à qui je dois chaque pas vers le succès, derrière toutes mes persévérances se cachent l'envie de vous rendre le minimum de ce que vous avez fait pour moi, j'espère pouvoir vous rendre heureux et ne jamais vous décevoir.

À mon cher et unique frère Chemssou que j'aime beaucoup.

À tout les membres de la famille Chiheb à Guelma et Litoufi à Sétif (tantes, Oncles, cousins et cousines : Imene, Karima et Wardiche).

À tout mes copains et copines (Zouzou, Selma, Sarah l'algéroise, Amine, Yousef, Mouh, Ramzi et Kadour) avec qui j'ai passée des moments inoubliables et qui m'ont toujours apporté soutien et consolation dans le besoin.

À ma chère et adorable binôme Sabrina et toute sa famille.

À mon encadreur Mr Farou Brahim et à tous les professeurs qui m'ont enseigné durant ces 5 années.

Sans oublier mes camarades de classe Master2 IM et à tout les étudiants du département d'informatique E8 que je connais.

Amira...





Dédicaces

ALHAMDOU LI ALLAH pour sa clémence et sa miséricorde.

A mes chers parents qui ont toujours été là pour moi, je vous suis reconnaissante pour les sacrifices consentis pour mon bien être. Votre soutien et votre éducation ont fait de moi ce que je suis aujourd'hui, j'espère pouvoir vous rendre heureux et ne jamais vous décevoir.

A mes frères Salim et Nadjib pour leurs soutiens morale, à tous les membres de la famille Oubraham et Haddad ou qu'ils sont (tantes, Oncles, cousins et cousines) et à qui j'adresse mon profond respect.

A ma merveilleuse amie, sœur et binôme Amira ainsi que sa famille.

A tous mes amis : Mousfida, Ferhat Oubraham, Doudou, djazira, Lamia Krim, Halima Belkadi, Mouna, Hanane Bouaziz, cherif et Zakj, Pour leur amitié, soutien morale, leurs conseils, et pour tous les instants inoubliables que j'ai passé avec vous je vous remercie.

A mon encadreur Mr Farou Brahim.

A tous mes camarades de classe master2 IM et à tous les étudiants du département d'informatique.

A tous ce qui m'aiment, et tous ceux que j'aime et surtout a tous ce qui méritent

Sabrina...



Résumé

La vidéosurveillance est une technologie en vogue, elle est très utilisée dans les applications de contrôles d'accès et reste l'un des domaines de vision par ordinateur qui attire l'attention de beaucoup de chercheurs.

L'intégration des systèmes de vidéosurveillance dans des endroits publics a rendu possible la surveillance des activités des humains pour garantir une meilleure rentabilité des services de sécurité. Cependant, comment un agent de sécurité peut-il analyser en temps réel des dizaines d'écrans de contrôle avec un taux d'erreur minimal? D'autre part, l'observation des écrans dans une longue durée est un travail ennuyeux. La réponse de ces deux problèmes tient en deux mots : vidéosurveillance.

C'est dans ce cadre que nous allons essayer de réaliser un système automatique qui permet de détecter les comportements jugés suspects dans une vidéo de surveillance.

Dans ce mémoire nous nous intéressons à la détection de l'individu par la GMM, l'extraction du squelette et le suivi de ce dernier dans des séquences vidéo filmées dans des conditions différentes, et enfin la reconnaissance des comportements suspects par l'intégration d'un simple réseau de neurones (PMC).

Mot clés: Vidéo surveillance, vision par machine, traitement d'images, reconnaissance, comportement humain

Sommaire

Liste des Figures..... 04

Liste des Tableaux..... 06

Liste d'Abbreviations et Acronymes..... 06

Introduction Générale..... 07

Chapitre I : La vidéosurveillance

Introduction 09

I. Généralités 09

 1. Qu'est-ce qu'une vidéo ?..... 09

 2. Vidéo numérique et analogique 09

 3. Les différents formats de fichiers vidéo 10

II. La vidéosurveillance 11

 1. Présentation 11

 2. Apparition 11

 3. Objectifs majeurs 11

 4. Domaines d'application 12

 4.1.Sécurité de sites publics et commerciaux 12

 4.2.Extraction d'informations 12

 4.3.Applications militaires 13

 4.4.Applications civiles 13

 5. Composants d'un système de vidéosurveillance 13

 6. Les différents types de système de vidéosurveillance 14

 6.1.Système sur réseaux IP 14

 6.2.Kit de vidéosurveillance 15

 6.3.Réseau « classique » de vidéosurveillance 15

 6.4.Système « hybride » de vidéosurveillance 15

 7. Processus d'un système de vidéosurveillance 16

 7.1.La détection 17

 7.1.1. Méthodes de soustraction de fond 17

7.1.1.1. Gaussienne simple	17
7.1.1.2. Mélange de gaussiennes	18
7.1.1.3. Différenciation temporelle	18
7.1.2. Calcul du flux optique	18
7.1.3. Suppression d'ombres	19
7.1.4. Détection de points caractéristiques dans une image	19
7.1.5. Détection d'objet basée sur la segmentation de l'image	19
7.1.6. Détection d'objet basée sur une reconnaissance de forme	21
7.2. La reconnaissance d'objets/personnes	21
7.3. Le Suivi	21
7.4. La reconnaissance d'actions/ comportements	24
7.4.1. Définition de la science du comportement	24
7.4.2. Le comportement humain	26
7.4.3. Quelques comportements suspects	26
7.5. Evaluation et déclenchement d'alertes	27
7.5.1. Alertes d'événements définis – modélisation explicite	28
7.5.1.1. Alertes basées sur le mouvement d'objets	28
7.5.1.2. Alertes basées sur la classe d'objets	28
7.5.1.3. Alertes basées sur le comportement	28
7.5.2. Alertes d'événements non-définis – modélisation implicite	29
8. Les différents systèmes de vidéosurveillance existants	29
8.1. Le projet ADVISOR-INRIA	29
8.2. Le projet CAVIARE-INRIA	29
8.3. Le projet CASSIOPEE-INRIA	29
Conclusion	30

Chapitre II : Conception du système

Introduction	31
1. Objectifs du système	31
2. Conception architecturale	32
2.1. Acquisition du flux vidéo	33
2.2. Fragmentation	33
2.3. Détection par GMM	33

2.4.Post Traitement	39
2.5.Extraction des points extrêmes (Suivi)	41
2.6.Segmentation LPE	42
2.7.Squelettisation	44
2.8.Extraction des paramètres	45
2.9.Reconnaissance des comportements suspects	45
Conclusion	49

Chapitre III : Réalisation du système

Introduction	50
1. Aspect matériel	50
2. Outils de développement	50
2.1.Eclipse	50
2.2.JDK « Java Development Kit »	51
2.3.JMF « Java Media Framework »	51
2.4.Neuroph studio	52
3. Description du système	52
4. Les types d'environnements	52
5. Tests et résultats	53
5.1.La détection par GMM	53
5.2.POST Traitement (Dilatation + érosion)	55
5.3.Suivi d'individu	56
5.4.Squelettisation et détection de contours	56
5.5.Résultats de la segmentation LPE	57
5.6.Apprentissage et reconnaissance des comportements suspects	61
Conclusion	63
Conclusion Générale et Perspectives	64
Bibliographie et web Graphie	66

Liste des Figures

Figure I.1 : <i>Exemple de détection d'objet abandonné</i>	12
Figure I.2 : Exemple de détection de personne dans une zone interdite	12
Figure I.3 : Exemple de surveillance du trafic routier	13
Figure I.4 : <i>composants d'un système de vidéosurveillance</i>	14
Figure I.5 : Vidéosurveillance sur réseau IP	14
Figure I.6 : Kit de vidéosurveillance	15
Figure I.7 : Système de vidéosurveillance analogique	15
Figure I.8 : <i>processus d'un système de vidéosurveillance</i>	16
Figure I.9 : la différenciation temporelle. (a) une scène avec deux objets en mouvement. (b) la différenciation échoue à détecter tous les pixels de l'objet à gauche puisqu'il est uniforme de couleur.les régions en mouvement sont marquées en rouge.....	18
Figure I.10: une image avec ses vecteurs de flux optique	19
Figure I.11: Illustration de détecteur de points d'intérêt: (a) Harris, (b) KLT, (c) SIFT	20
Figure I.12: Illustration de détection de zone par segmentation de l'image (a): (b) segmentation basée sur la méthode du Mean-Shift, (c) segmentation utilisant les Graph-Cuts	20
Figure I.13 : <i>Représentation d'objet (a) centre, (b) ensemble de points</i>	22
Figure I.14 : <i>Représentation d'objets (c) boite englobante rectangulaire, (d) ellipse</i>	22
Figure I.15 : <i>Représentation d'objet. (e) point de control sur contour, (f) contour complet, (g) silhouette d'objet</i>	23
Figure I.16: Représentation d'objet (h) Squelette	23
Figure I.17 : Quelques comportements suspects	27

Liste des Figures, Tableaux, Abréviations et Acronymes

Figure I.18 : <i>Exemples de surveillance d'un aéroport</i>	28
Figure II.1 : Architecture globale du système	32
Figure II.2 : Perceptron multi couches utilisé	46
Figure II.3 : <i>Détails des paramètres d'un PMC, mode de calcul de l'activation d'un neurone et fonction sigmoïde σ</i>	47
Figure III.1 : <i>Version d'Eclipse utilisé</i>	51
Figure III.2 : <i>Suivi d'individu. (A) Dans un environnement avec éclairage variant, (B) Dans un environnement simple et (C) dans un environnement a faible éclairage</i>	56
Figure III.3 : Résultat de la squelettisation après le prétraitement	56
Figure III.4 : Résultat de la détection de contour après le prétraitement	57
Figure III.5 : Résultat de la détection de contour avec filtre Laplacien	57
Figure III.6 : Résultat de la détection de contour avec filtre de Sobel	58
Figure III.7 : <i>Silhouette d'un individu en marchant</i>	58
Figure III.8 : <i>Silhouette d'un individu claquant les mains</i>	59
Figure III.9 : <i>Silhouette d'un individu en courant</i>	59
Figure III.10 : <i>Silhouette d'un individu en sautant</i>	59
Figure III.11 : <i>Squelette d'un individu claquant les mains</i>	60
Figure III.12 : <i>Squelette d'un individu en courant</i>	60
Figure III.13 : <i>Squelette d'un individu en sautant</i>	60
Figure III.14 : <i>Graphe d'apprentissage</i>	61
Figure III.15 : exemple des valeurs du test	62
Figure III.16 : <i>Résultat d'un comportement non-suspect</i>	62
Figure III.17 : <i>Résultat d'un comportement suspect</i>	63

Liste des Tableaux

Tableau I.1 : Les principaux formats vidéo	10
Tableau III.1 : Résultats du test de la GMM avec changement des paramètres	53
Tableau III.2 : Résultats du test de la <i>GMM selon le type d'environnement</i>	54
Tableau III.3 : Résultat du prétraitement (dilatation + érosion).....	55

Liste des Abréviations et Acronymes

CCTV : Closed Circuit TeleVision.

CCD : Charge Coupled Device.

IP : Internet Protocol.

INRIA : Institut Nationale de Recherche en Informatique et en Automatique.

GMM : Gaussien Mixture Model.

LPE : Ligne de Partage des Eaux.

RGB : Red Green Blue.

NDV : Niveau de Gris.

PMC : Perceptron multi couches.

IDE : Environnement de Développement Intégré.

API : Application Programming Interface.

JMF : Java Media Framework.

JDK : Java Development Kit.

Introduction Générale

Au cours de la dernière décennie, la sécurité des personnes, des biens et de l'information s'est imposée comme un enjeu majeur à l'échelle planétaire. Confrontées à des problèmes tels que la lutte au terrorisme, le renforcement de la sécurité intérieure et l'essor de la cybercriminalité, nos sociétés investissent de plus en plus pour assurer leur protection. Les technologies de l'information et des communications apportent des solutions nouvelles et sophistiquées pour la sécurité physique et informatique.

Parmi les solutions proposées, la vidéosurveillance constitue l'une des technologies de sécurité les plus anciennes et les plus répandues. Ces nouveaux systèmes génèrent une quantité colossale d'informations vidéo qui dépassent les capacités de surveillance des agents de sécurité.

La vidéosurveillance permet de traiter, de façon logicielle, la vidéo captée pour n'en retenir que les données pertinentes pour la sécurité. Celle-ci repose sur des systèmes qui incorporent de manière automatique les technologies de pointe de plusieurs domaines, notamment en vision par ordinateur, en traitement du signal, en intelligence artificielle et en fouilles de données.

Avec la généralisation de l'utilisation d'images numériques, l'analyse du mouvement dans les vidéos s'est révélée être un outil indispensable en particulier pour la vidéosurveillance.

Dans ce contexte, les systèmes d'analyse du mouvement tentent de localiser la (ou les) personne(s) présente(s) dans une séquence vidéo, d'identifier et de suivre les différentes parties du corps, et enfin d'interpréter les comportements et les situations selon l'application visée. Dans un cadre général d'un système d'analyse du mouvement humain, l'extraction de la silhouette est l'une des étapes initiales. La qualité du résultat a une forte incidence sur l'efficacité de l'ensemble du système, d'où l'importance de disposer d'approches fiables pour l'extraction de premier plan. Malgré les travaux réalisés, ce domaine est encore confronté à de nombreux défis induits par le non-contrôle de l'environnement tels que les changements de lumière, la présence d'ombres et le camouflage, qui rendent l'extraction de la silhouette assez ardue. Par ailleurs, l'information contenue dans la silhouette extraite est utilisée pour détecter

et suivre les mouvements de la personne dans la séquence et de déterminer son comportement.

L'objectif de notre travail est de réaliser un système qui permet de détecter des comportements jugés suspects dans une vidéo de surveillance.

Le présent mémoire est organisé comme suit :

Chapitre I : La vidéosurveillance : est composé de deux parties : la première partie contient quelques généralités sur la vidéo, des définitions relatives à celle-ci sont présentées brièvement.

Nous avons présenté dans la deuxième partie un état de l'art sur la vidéosurveillance, contenant son apparition, ses objectifs ainsi que ses domaines d'application, nous avons aussi défini les différentes approches de détection, de suivi et de reconnaissance avec une petite échappée à la psychologie du comportement afin de déterminer le minimum des comportements suspects qu'un être humain peut avoir. Et nous avons achevé ce chapitre avec les différents types de systèmes existants jusqu'à aujourd'hui.

Chapitre II : Conception : nous présentons l'objectif de ce système, les différentes étapes de conception, ainsi que les différents algorithmes et les méthodes implémentées avec l'amélioration apportée.

Chapitre III : Réalisation : est consacré à la partie réalisation de notre système, ou nous présentons l'environnement du travail (matériel et logiciel), les testes et les résultats obtenus avec l'évaluation des performances des algorithmes utilisés.

Et enfin, nous terminerons par une conclusion générale englobant l'essentiel de notre travail, notre point de vue et les différentes perspectives résultantes.

Chapitre I : La vidéosurveillance

Introduction

I. Généralité

1. Qu'est-ce qu'une vidéo ?
2. Vidéo numérique et analogique
3. Les différents formats de fichiers vidéo

II. Vidéosurveillance :

1. Présentation
2. Apparition
3. Objectifs majeurs
4. Domaines d'application
 - 4.1 Sécurité de sites publics et commerciaux
 - 4.2 Extraction d'informations
 - 4.3 Applications militaires
 - 4.4 Applications civiles
5. Composants d'un système de vidéosurveillance
6. Les différents types de système de vidéosurveillance
 - 6.1 Système sur réseaux IP
 - 6.2 Kit de vidéosurveillance
 - 6.3 Réseau « classique » de vidéosurveillance
 - 6.4 Système « hybride » de vidéosurveillance
7. Processus d'un système de vidéosurveillance
 - 7.1 La détection
 - 7.2 La reconnaissance d'objets/personnes
 - 7.3 Le suivi
 - 7.4 La reconnaissance d'actions/comportements
 - 7.5 Evaluation et déclenchement d'alertes
8. Les différents systèmes de vidéosurveillance existants
 - 8.1 Le projet ADVISOR-INRIA
 - 8.2 Le projet CAVIARE-INRIA
 - 8.3 Le projet CASSIOPEE-INRIA

Conclusion

Chapitre I

La vidéosurveillance

Introduction :

Le traitement d'images est une science récente qui permet d'analyser et de traiter des images numériques par un ensemble de techniques et de méthodes afin d'exploiter les informations contenues dans ces dernières. De nos jours, l'utilisation du traitement d'images touche pratiquement tous les domaines en particulier la vidéosurveillance. Cette dernière nous aide à observer plusieurs positions différentes en même temps et à reconnaître rapidement les événements anormaux qui se passent dans la scène.

I. Généralités :

1. Qu'est-ce qu'une vidéo ?

Avant de définir une vidéo il faut d'abord connaître qu'est-ce qu'une image numérique ?

L'image numérique est la représentation d'une entité physique sous une forme électrique ou informatique, elle est composée d'un ensemble de pixel ou le pixel est la plus petite entité représentative de l'image. [29]

Une vidéo est une succession d'images à une certaine cadence [29]. On caractérise la fluidité d'une vidéo par le nombre d'images par secondes (en anglais frame rate), exprimée en FPS (Frames per second, en français frames par seconde) [29].

D'autre part la vidéo au sens multimédia du terme est généralement accompagnée de son, c'est-à-dire de données audio. [29]

2. Vidéo numérique et analogique :

On distingue généralement deux grandes familles d'images animées :

- **La vidéo analogique** : représente l'information comme un flux continu de données analogiques, destinée à être affichée sur un écran de télévision (basé sur le principe du balayage). Il existe plusieurs normes pour la vidéo analogique (PAL, NTSC, SECAM, DIGITAL 8). [29]

- La vidéo numérique : consiste à coder la vidéo en une succession d'images numériques. [29]

3. Les différents formats de fichiers vidéo :

Il existe plusieurs formats de fichiers vidéo, les plus courants sont décrits dans le tableau ci-dessus : [20]

Formats / extension	utilisation	Logiciel/support
AVI .avi	Prévoit une définition d'image maximale de 320 x 240 pixels et une cadence maximale de 30 fps.	Normalement tous les logiciels de lecture vidéo
WMV .wmv	Un format conteneur de Microsoft et sa famille de codecs vidéo. On le trouve souvent chez des vidéos en streaming ou en téléchargement.	Windows Media Player
MOV .mov	Format conteneur d'Apple, pouvant contenir un très large choix de codecs, y compris haute définition.	Quicktime
DivX et Xvid .xvid/.avi/.mpg	Les codecs vidéo les plus utilisés. Xvid est le concurrent libre de DivX.	Nécessite l'installation des codecs DivX / Xvid à jour, qui sont inclus dans de nombreux lecteurs vidéo
MKV .mkv .mka .mks	Format conteneur russe, pouvant contenir de très nombreux codecs.	Se lit avec VLC, Media Player Classic et quelques autres .
H264 / MPEG 4 AVC	C'est le codec utilisé pour la haute définition.	Se lit avec la majorité des lecteurs vidéo .
FLV / .flv	utilisé en quantité sur Internet	S'ouvre parfaitement avec VLC
Real Video/ .rmvb	Codec vidéo propriétaire de Real Networks	RealPlayer ou RealAlternative
MPEG1 .mpeg/.mpe, .mpg/.mpv, .dat	appelé aussi format VCD 360 X 240 (pixels)	La plupart des logiciels de lecture vidéo
MPEG2 .mpg	appelé aussi format SVCD 480 X 576 (pixels)	La plupart des logiciels de lecture vidéo
MPEG2 .mpeg/.mpe, .mpg/.mpv, .mp2/.m2p.	appelé aussi format DVD 720 X 576	La plupart des logiciels de lecture vidéo
Theora .ogv ou .ogg	Theora est un format de compression vidéo. il a pour but de créer un ensemble de standards ouverts concernant le traitement de signaux multimédia (son, vidéo).	Lecture avec VLC, entre autres, ou Firefox 3.5 et ultérieurs

Tableau I.1 : Les principaux formats vidéo [20]

II. La vidéosurveillance

1. Présentation :

La vidéosurveillance consiste à placer des caméras de surveillance dans un lieu public ou privé en gardant les traces des personnes observées et en détectant des schémas suspects de mouvement ou comportement afin de prendre une décision fondée sur cette analyse de scène. L'opérateur est alerté quand il y a un écart par rapport à un comportement normal.

L'évolution technologique des systèmes de vidéosurveillance a débuté avec les systèmes CCTV (Closed-Circuit TeleVision) analogiques. [19]

Actuellement, la majorité des systèmes utilisent des caméras à capteurs CCD (Charge-Coupled Device, ou dispositif à transfert de charge) pour capturer des images discrétisées et des techniques analogiques pour la distribution et le stockage de ces images. [19]

2. Apparition :

Depuis les années 1970, la vidéosurveillance a fait son apparition en ayant pour but principal de prévenir le terrorisme dans le monde. L'Angleterre fut le premier pays à instaurer le système de vidéosurveillance.

Actuellement, dans le monde, le nombre de caméras s'élève à environ 25 millions, dont 4.2 au Royaume-Uni. La répartition de ces dispositifs est très inégale dans le monde. Au niveau des individus, la vidéosurveillance a des effets néfastes sur la vie privée des individus, mais aussi des avantages pour leur sécurité. La vidéosurveillance a entraîné l'élaboration de certaines législations qui ont suscité de nombreux débats. [19]

3. Objectifs majeurs:

Les raisons de l'installation de systèmes de vidéosurveillance sont diverses :

- Fournir une interprétation automatique des scènes filmées et de comprendre et prédire les actions et interactions des objets observés.
- Prédire les incidents en détectant des comportements suspects et déclencher des alarmes. comme exemples de prédictions :
Un sac abandonné, une-pièce ou objet disparu d'un endroit, un véhicule dépassant une vitesse maximale...etc.

4. Domaines d'application :

On dénombre quatre grandes catégories dans lesquelles on retrouve l'application de la vidéosurveillance :

4.1 Sécurité de sites publics et commerciaux :

- Surveillance des banques, centres commerciaux, aéroports, ports, gares, métros, musées, stades, propriétés privées et parkings pour la prévention et la détection d'intrusions, de crimes ou d'actes de dégradation de biens. [16] [21]
- Surveillance des autoroutes et des chemins de fer pour la détection d'accidents. [5]
- Surveillance des forêts pour la détection d'incendies [30]
- Contrôle d'accès à des sites privés.



Figure I.1 : Exemple de détection d'objet abandonné [3]

4.2 Extraction d'informations :

- Mesure du débit de la circulation routière et d'encombrement de piétons. [28]
- Comptage de personnes dans des espaces publics.
- Extraction de statistiques dans des activités sportives. [11]



Figure I.2 : Exemple de détection de personne dans une zone interdite [13]

4.3 Applications militaires :

- Surveillance des frontières.
- Surveillance de sites sécurisés.
- Assistance du contrôle et du commandement sur le terrain.

4.4 Applications civiles :

- Mesure de la vitesse des véhicules [17]
- Détection de franchissement de feu rouge et de franchissement de ligne continue



Figure I.3 : Exemple de surveillance du trafic routier [17]

5. Composants d'un système de vidéosurveillance :

Un système de vidéosurveillance est essentiellement composé de : [19]

- a. Des caméras pour filmer les zones à surveiller.
- b. Des câbles ou des liaisons sans fil pour relier les caméras au réseau ou à l'enregistreur.
- c. Des alimentations au niveau de chaque caméra ou déportées grâce aux câbles qui véhiculent le courant.
- d. Un enregistreur pour sauvegarder les images
- e. Un logiciel pour accéder aux images.
- f. Un écran pour visualiser les images de l'enregistreur
- g. Une connexion internet pour accéder aux caméras à distance.

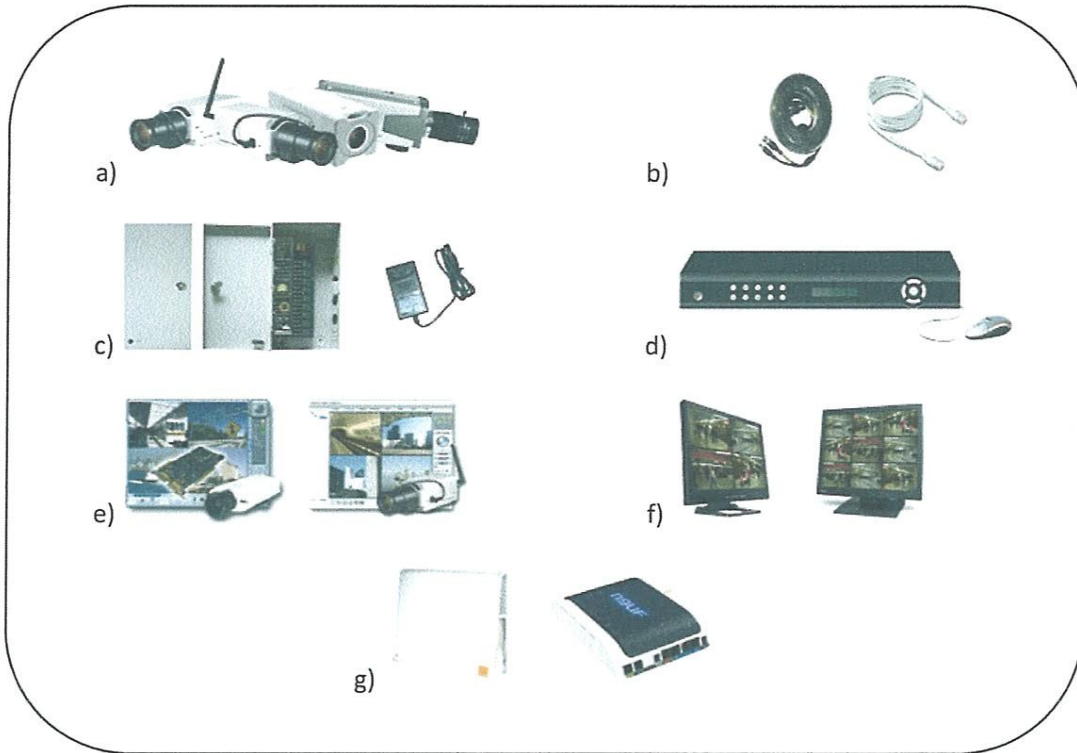


Figure I.4 : composants d'un système de vidéosurveillance [19]

6. Les différents types de système de vidéosurveillance :

6.1 Système sur réseaux IP :

Ce système relie un réseau de caméras IP à un système d'enregistrement numérique. D'une part, cela permet de pouvoir stocker une quantité importante d'images, sans perte de qualité. D'autre part, le fait d'informatiser un système de surveillance permet de profiter des technologies de communication comme Internet. Cette technologie permet également d'économiser et de mutualiser les câbles réseaux. [19]

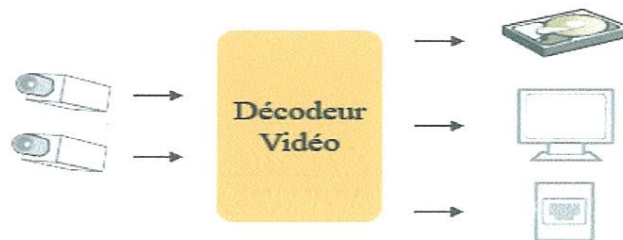


Figure I.5 : Vidéosurveillance sur réseau IP [19]

6.2 Kit de vidéosurveillance :

Un « kit » c'est le genre de caméras utilisées dans les petits magasins. Il regroupe en général une ou deux caméras et un moniteur. Ces systèmes sont plutôt utilisés à titre de prévention et n'enregistrent pas ce qu'ils voient. C'est en quelque sorte de la vidéosurveillance bon marché qui est proposée comme une solution de sécurité peu coûteuse. [19]

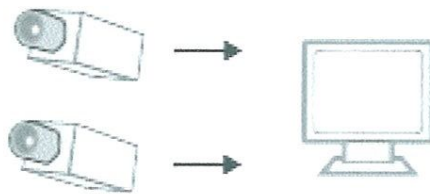


Figure I.6 : Kit de vidéosurveillance [19]

6.3 Réseau « classique » de vidéosurveillance :

Le réseau est basé sur un système analogique, dans la plupart des cas un enregistrement est limité. Il s'agit d'une des méthodes les plus anciennes, donc des plus répandues dans un grand nombre d'établissements. [19]

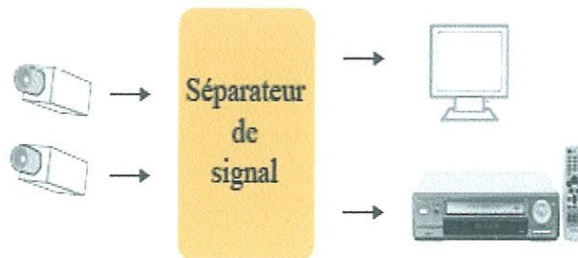


Figure I.7 : Système de vidéosurveillance analogique [19]

6.4 Système « hybride » de vidéosurveillance :

Les systèmes hybrides intègrent les systèmes classiques de vidéosurveillance basés sur les caméras analogiques et les caméras en réseau. Ils permettent d'intégrer aisément les deux types de systèmes en place sur un seul serveur ou de faciliter l'évolution d'un système de

vidéosurveillance analogique vers le numérique, sans remettre en cause l'existant et d'introduire de nouvelles fonctions comme la détection de disparition / apparition d'objet et le comptage d'objets ou de personnes. [19]

7. Processus d'un système de vidéosurveillance :

Il y'a 4 taches importantes dans un système de vidéosurveillance automatisé :

- La détection et la segmentation d'objets ou de personnes cible afin de les isolés du reste de l'image.
- La reconnaissance d'objets ou de personnes.
- Le suivi d'objets ou de personnes (position, vitesse, directions de ses mouvements).
- La reconnaissance d'actions (comportements).

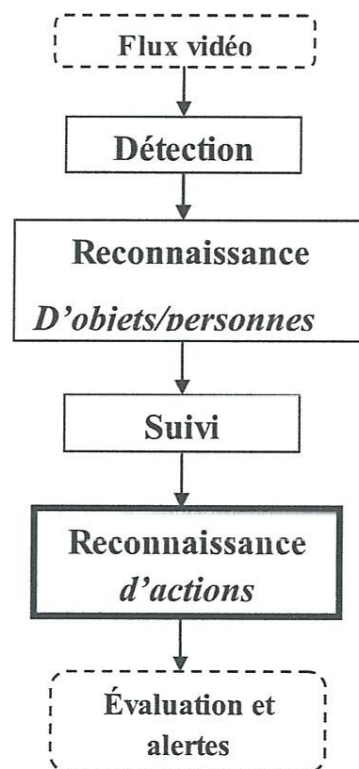


Figure I.8 : processus d'un système de vidéosurveillance

⎓ : Représente l'entrée et la sortie. ⎓ : Représente les taches du système.

⎓ : La tâche principale de notre système (implémentation).

7.1 La détection :

L'identification des objets est une tâche critique pour de nombreuses applications de traitement d'image et de vision par ordinateur. Elle consiste à étiqueter chaque pixel d'une image et l'associer ou non à une région en mouvement dans la scène. En effet elle fournit une classification des pixels, soit en avant plan (mobile) ou arrière plan (statique). La détection joue un rôle très important dans le système de vidéosurveillance, car le résultat de cette étape va influencer sur toutes les étapes suivantes.

Les techniques les plus utilisées dans la détection sont :

7.1.1 Méthodes de soustraction de fond :

La méthode d'extraction consiste à estimer une représentation appropriée de la scène appelée « modèle de fond », puis à chercher tout changement ou déviation par rapport au modèle dans chaque image d'entrée traitée. Les régions de l'image d'entrée où il y aura des changements significatifs par rapport au modèle de fond correspondront aux éléments d'intérêt « forme ». Donc la soustraction de fond sert à déterminer les régions de l'image où une activité potentiellement intéressante est en train de se produire. Les pixels constituant ces régions sont marqués pour des traitements ultérieurs afin d'éliminer le bruit dû à une soustraction de fond imparfaite. [18]

La soustraction de fond reste l'approche la plus utilisée pour la détection et ceci est dû notamment au fait que c'est une méthode rapide et simple, mais généralement sensible à des changements dynamiques telle que l'éclairage ou lié au mouvement ou changement de fond ainsi que le bruit de caméra.

7.1.1.1 Gaussienne simple :

C'est une technique statique de soustraction du fond qui consiste à calculer l'arrière plan à partir d'une image moyenne avec écart-type de la scène pour chaque pixel (avant l'arrivée d'objets mobiles). En soustrayant chaque nouvelle image de cette image moyenne et en utilisant un seuillage basé sur l'écart-type, nous obtenons les objets mobiles. Cette méthode peut s'adapter aux longs changements de la scène en mettant à jour de façon récursive le modèle en utilisant un filtre adaptatif simple. [8]

7.1.1.2 Mélange de gaussiennes :

Le fond de la scène contient souvent des objets non statiques, le mouvement de fond de ces objets provoque la variation d'intensité des pixels, donc la représentation de cette dernière par une gaussienne simple ne tiendra pas compte de cette situation.

Cette méthode compare chaque pixel à l'ensemble des modèles existants à cet endroit pour trouver une correspondance (matching). Les paramètres du modèle correspondant sont mis à jour grâce à un facteur d'apprentissage. Si aucune correspondance n'est trouvée, le modèle le moins probable est éliminé et remplacé par une nouvelle gaussienne avec les statistiques initialisées par les valeurs du pixel actuel. [7]

7.1.1.3 Différence temporelle :

La différence temporelle détecte les régions en mouvement en faisant la différence des trames consécutives (deux ou trois), pixel par pixel, dans une séquence vidéo. Cette méthode est très adaptative pour les changements dynamiques de scène, en revanche elle échoue généralement à détecter les pixels de certains types d'objets en mouvement. [12]



Figure I.9 : la différenciation temporelle. (a) une scène avec deux objets en mouvement. (b) la différenciation échoue à détecter *tous les pixels de l'objet à gauche puisqu'il est uniforme de couleur*. les régions en mouvement sont marquées en rouge. [12]

7.1.2 Calcul du flux optique :

La technique du flux optique peut segmenter les objets en mouvement à partir d'une caméra mobile ou statique. Les vecteurs de flux sont utilisés pour diviser l'image en segments de mouvements égaux. Le mouvement de fond sera différent de celui des objets en mouvement et donc les objets en mouvement pourront être segmentés de l'arrière plan.

Cette méthode est connue pour sa complexité calculatoire et donc la mise en œuvre en temps réel est difficile et coûteuse. Dans les applications de surveillance, le flux n'est pas toujours correct. [25]

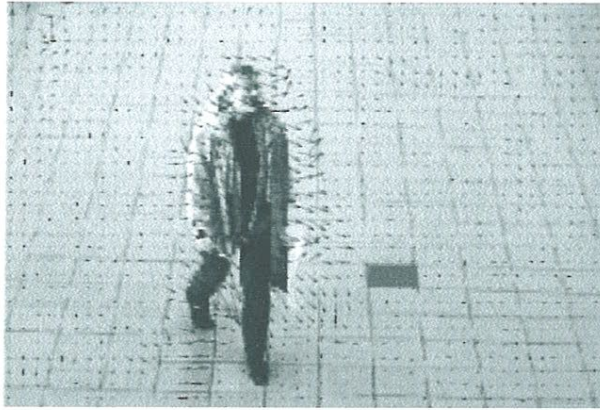


Figure I.10: une image avec ses vecteurs de flux optique [25]

7.1.3 suppression d'ombres :

L'ombre se produit quand un objet occulte partiellement ou totalement la lumière directe à partir d'une source d'éclairage. La présence d'ombres dans une image peut modifier le perçu de la forme et la couleur de l'objet et donc nuire à la segmentation. La plupart des méthodes existantes pour la détection d'ombre exploitent l'information provenant de changement d'apparence de pixel en présence d'ombre. Elles peuvent être classées en méthodes fondées sur un modèle ou méthodes sans modèle.

Les techniques fondées sur un modèle comptent sur des modèles représentant la connaissance a priori la géométrie de la scène, d'objets, d'élimination. D'autre part, les méthodes sans modèles détectent l'ombre en utilisant des propriétés spectrales, spatiales et temporelles.

Malgré la diversité des approches qui ont été développées il n'y a pas de méthodes généralement acceptées pour la détection d'ombre dans les séquences d'images, cela est dû au fait que la plupart des méthodes utilisent les caractéristiques spectrales telles que, l'intensité, la teinte ou la saturation pour identifier les pixels de l'ombre et ces derniers peuvent changer dans diverses circonstances. [12]

7.1.4 Détection de points caractéristiques dans une image :

Les méthodes de détection de points dans l'image listent des points caractéristiques de l'image ayant une texture riche dans le voisinage de ces points. La qualité d'un point d'intérêt dépend notamment de son invariance aux changements d'illumination.

Les méthodes les plus usuelles de détection de points d'intérêt sont le détecteur de point de Harris, le détecteur KLT et le détecteur SIFT. Une fois la détection effectuée, une étape de mise en correspondance est effectuée afin d'apparier les points d'une image à l'autre dans un but de suivi d'objet. [10]

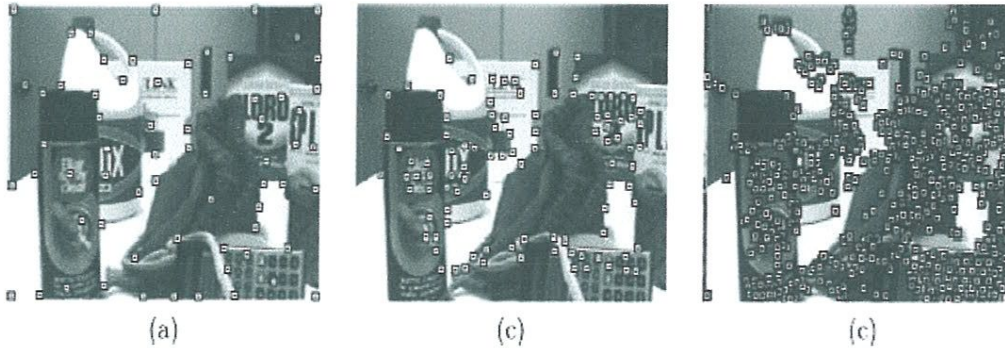


Figure I.11: Illustration de détecteur de points d'intérêt: (a) Harris, (b) KLT, (c) SIFT. [10]

7.1.5 Détection d'objet basée sur la segmentation de l'image :

Le but des algorithmes de segmentation d'image est de partitionner l'image en régions similaires. Les techniques récentes de détections d'objets basées sur la segmentation d'image utilisée pour le suivi d'objets sont les suivantes : segmentation basée sur la méthode du Mean-Shift de Comaniciu et Meer, segmentation utilisant les Graph-Cuts comme Shi et al. Et segmentation fondée sur les contours actifs. [10]

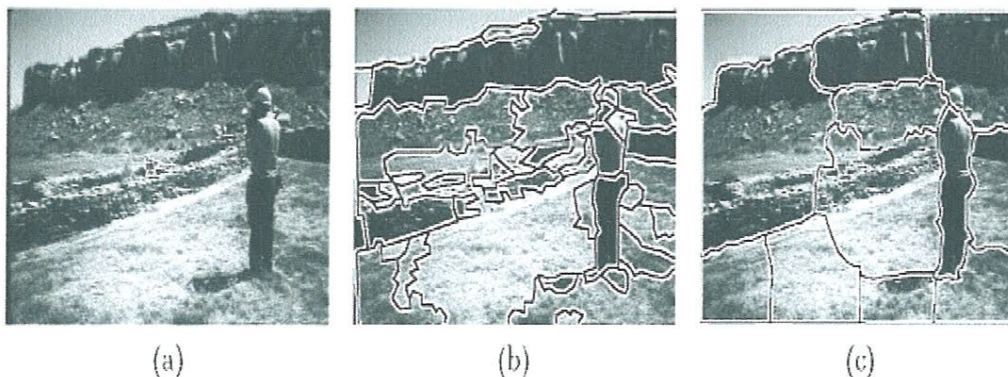


Figure I.12: Illustration de détection de zone par segmentation de l'image (a): (b) segmentation basée sur la méthode du Mean-Shift, (c) segmentation utilisant les Graph-Cuts.

[10]

7.1.6 Détection d'objet basée sur une reconnaissance de forme :

La caractéristique principale de ces techniques est qu'elles sont efficaces avec des caméras mobiles et des objets fixes. De plus, elles permettent de séparer les objets d'intérêt et de différencier différentes classes d'objets (personne, véhicule, bagage, etc.). Ces méthodes de détection reposent sur l'apprentissage automatique de différentes vues de l'objet à partir d'exemples positifs (images d'une personne) et négatifs (images ne représentant pas de personne). Cela revient à classer les détections selon un certain nombre de labels : par exemple, un label personne ou autre objet, personne assise, debout, couchée, etc. [10]

7.2 La reconnaissance d'objets/personnes:

La reconnaissance de personnes se fait avant le suivi, et consiste à identifier un individu cible parmi un ensemble d'individus. Dans les systèmes actuels, les principales approches de reconnaissance de personnes utilisées sont la reconnaissance de visage, la reconnaissance de la démarche et la reconnaissance basée sur l'apparence. [27]

Les deux premières catégories d'approches font partie des méthodes de reconnaissance biométrique. Ces dernières ont l'avantage d'être non invasives et d'utiliser des caractéristiques uniques à chaque personne. [27]

Les méthodes basées sur l'apparence font référence à l'apparence extérieure des personnes, et utilisent généralement des caractéristiques décrivant leur forme et leur taille ou leurs vêtements telles que la couleur et la texture. Ces méthodes ont également l'avantage d'être non invasives et sont particulièrement adaptées au contexte de la vidéosurveillance du point de vue des modalités de prise de vue des séquences d'images, ce qui n'est pas toujours le cas des méthodes biométriques qui imposent généralement des contraintes en termes de résolution et de conditions de prises de vue. [27]

7.3 Le suivi :

L'objectif du suivi tout au long du flux vidéo consiste à associer ces objets détectés dans l'image courante avec ceux détectés dans les images précédentes et de déterminer leur position de manière continue et fiable. Dans les systèmes de vidéosurveillance, le suivi en temps réel représente une étape primordiale pour des applications d'analyse d'activités et de compréhension d'événements. En effet, le processus de suivi d'un individu cible permettra de

générer son chemin parcouru qui représente un des moyens les plus utilisés pour décrire l'activité d'une personne.

Diverses stratégies de suivi de cibles ont été développées, on peut les regrouper en trois grandes familles : [18]

1. **Les méthodes de suivi basées point** : Les objets détectés dans des images consécutives sont représentés par un point typiquement centré, ou par un ensemble de points ou l'association de ces points est basée sur l'état précédent de l'objet en termes de position et de mouvement. [18]

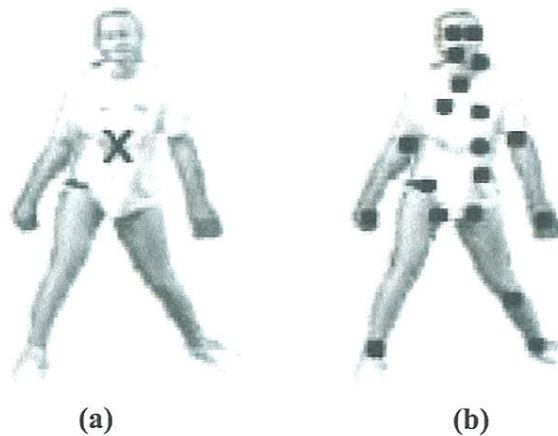


Figure I.13 : Représentation d'objet (a) centre, (b) ensemble de points [4]

2. **Les méthodes basées noyau** : Le terme « noyau » se réfère à la forme et à l'apparence de l'objet, ce dernier peut être représenté par une forme rectangulaire ou elliptique. Les objets sont suivis en calculant le mouvement du noyau dans des images consécutives. [18]

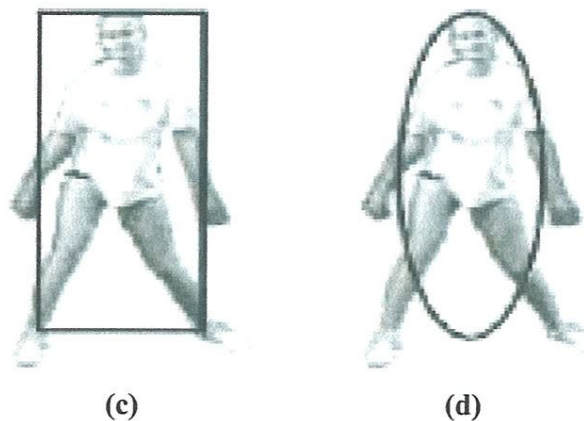


Figure I.14 : Représentation d'objets (c) boîte englobante rectangulaire, (d) ellipse [4]

3. **Les méthodes basées silhouette** : Elles effectuent le suivi d'un objet en estimant sa région dans chaque image puis utilisent les informations extraites de cette région pour le suivi. Pour un modèle d'objet, la silhouette est suivie soit par l'évolution de contours ou le squelette d'objet. [18]

- **La silhouette et le contour d'objet :**

La représentation de contour définit les bords d'un objet. La région intérieure du contour est appelée la silhouette d'objet. La représentation de silhouette et contour est appropriée pour le suivi de forme non rigide. [4]

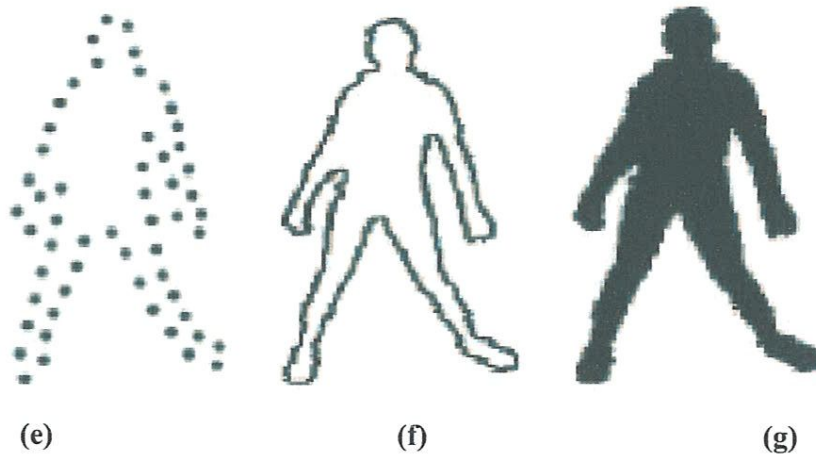


Figure I.15 : Représentation d'objet. (e) point de contrôle sur contour, (f) contour complet, (g) silhouette d'objet [4]

- **Modèle de Squelette :**

Le squelette d'un objet peut être extrait par l'application de transformation d'axe médian de la silhouette d'objet. Ce modèle est couramment utilisé comme représentation de forme pour la reconnaissance des objets. La représentation par squelette peut être utilisée pour modéliser des objets à la fois articulés et rigides. [1]

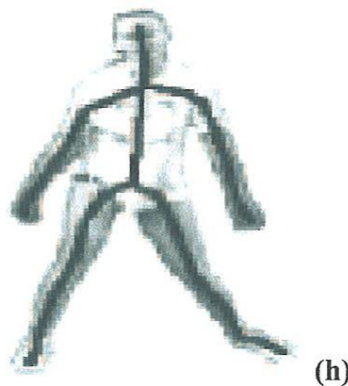


Figure I.16: Représentation d'objet (h) Squelette. [4]

7.4 La reconnaissance d'actions/ comportements :

La reconnaissance d'actions se fait après le suivi et consiste à reconnaître les actions des individus cibles dans des séquences vidéo afin de permettre une description sémantique ultérieure, de sorte que le système puisse comprendre le comportement et l'activité de chaque personne. [18]

Le comportement de l'être humain est le reflet de ses désirs et de ses préférences, de ses motivations et de ses affects, de ses pensées et de ses croyances. La connaissance de ce dernier, dans toutes ses modalités, est un élément essentiel pour comprendre l'être humain, et la place qu'il occupe parmi les autres êtres vivants.

Au fil du temps, dans une volonté de compréhension scientifique, de nombreuses approches se sont développées, visant à construire cette connaissance sur des analyses et des bases solides, s'appuyant sur la recherche.

7.4.1 Définition de la science du comportement :

Le discours sur le comportement change selon les époques et les écoles. Ce vaste domaine de recherche aborde l'être humain dans sa dimension cognitive, sociale, morale et culturelle.

Chacune des approches utilisées dans les sciences du comportement a proposé sa propre définition. Nous envisageons ici, dans un survol rapide, les définitions correspondant aux principales d'entre elles : l'éthologie, le béhaviorisme, la psychologie cognitive et la psychanalyse. [15]

a. L'approche éthologique :

L'éthologie place le comportement sous la dépendance de l'instinct. Le comportement instinctif est une séquence motrice (mouvements, émissions sonores, attitudes posturales) qui se déroule de façon invariante et est exécuté de la même façon par tous les individus de la même espèce. [15]

Ce type de comportement peut être le résultat d'une chaîne de réflexes déclenchés par un stimulus approprié, ou se déclencher spontanément sous l'influence de facteurs endogènes et se dérouler à vide, en l'absence de stimulus. La théorie éthologique insiste donc sur la transmission héréditaire de comportements sélectionnés par les nécessités de l'adaptation au milieu. [15]

b. L'approche behavioriste :

Ce que l'approche éthologique doit à Charles Darwin, l'approche behavioriste le doit à Ivan Pavlov.

Le behaviorisme fait de l'étude du comportement l'objectif central d'une école de psychologie qui se veut une psychologie objective. Le paradigme de la psychologie objective est donc celui de la mise en relation d'une réponse comportementale et d'un stimulus extérieur à l'organisme. Les termes de « succès », de « plaisir » ou de « satisfaction », ainsi que, à l'inverse, de « punition » ou de « déplaisir », sont couramment employés par la psychologie behavioriste pour formaliser cette notion de réponse favorable ou défavorable à l'organisme. Le comportement a donc pour fonction de favoriser l'interaction entre l'organisme et l'environnement. [15]

c. L'approche cognitiviste :

L'approche cognitiviste du comportement a une double origine. Elle est née en réaction contre le behaviorisme, en réhabilitant la notion d'états mentaux. Mais elle est aussi tributaire de la « révolution cognitive » des années 1950 qui a donné naissance aux sciences cognitives au sens large, dont la psychologie fait partie depuis le début. La psychologie cognitive postule que l'appareil mental est constitué de représentations, états mentaux possédant une structure et un contenu accessibles à la description objective. [15]

d. L'approche psychanalytique :

À l'inverse de l'idéal d'une psychologie objective mis en avant par l'approche behavioriste (cognitiviste), l'approche psychanalytique se veut une psychologie subjective. Elle a la particularité d'insister sur l'histoire individuelle de chaque sujet pour comprendre la dynamique intrapsychique qui conditionne son comportement et peut être à l'origine de ses conflits internes. La théorie psychanalytique se présente avant tout comme une théorie clinique du comportement, à l'opposé d'autres théories qui en font un objet d'études de laboratoire. [15]

7.4.2 Le comportement humain :

Le terme « comportement » désigne les actions d'un être vivant. Il a été introduit en psychologie française en 1908 par Henri Piéron comme équivalent français de l'anglais-américain behavior. On l'utilise notamment en éthologie (humaine et animale) ou en psychologie scientifique. Il peut aussi être pris comme équivalent de conduite dans l'approche psychanalytique. [31]

Il existe une très forte interaction entre le comportement et la pensée. S'il est évident que la pensée influence le comportement, l'on sait aujourd'hui que le comportement influence de façon certaine la pensée. Les techniques de manipulation les plus efficaces sont basées sur ce principe, baptisé congruence. L'homme a besoin, pour son équilibre, de justifier son propre comportement et de s'auto persuader que ses actes sont légitimes et cohérents. L'expérience conditionne donc la pensée. [31]

Il existe deux types de comportement :

a. Comportement Normal :

Ensemble des activités et des réactions normales observables chez un individu placé dans son milieu de vie et dans des circonstances données. [31]

b. Comportement Anormal (suspect) :

Un comportement suspect est tout comportement anormal ou l'anormalité est tout schème de comportement inadapté, perturbateur ou nuisible pour l'individu ou la société. Utilisation de trois normes :

A. Norme culturelles : déviant des normes sociales (homosexualité).

B. Norme du malaise subjectif: évaluation de l'individu.

C. Norme de comportement inadapté. Pensées sentiments et actes nuisent au fonctionnement de leur vie intime ou sociale. [31]

7.4.3 Quelques comportements suspects :

On peut définir quatre mouvements dans nos vidéos qui présentent des comportements suspects :



Figure 1.17 : Quelques comportements suspects [14]

7.5 Évaluation et déclenchement d'alertes :

L'analyse de scène permet l'interprétation d'événements et le déclenchement d'alertes, ces deux derniers sont des objectifs majeurs d'un système de vidéosurveillance. Un événement dit « anormal » peut être automatiquement détecté comme étant un écart par rapport à des modèles communs d'activité. Par conséquent, savoir comment définir et représenter un comportement « normal » est un point crucial.

Nous pouvons définir deux types d'alertes qui peuvent être générés par un système de vidéosurveillance: [24]

- Des alertes d'événements définis par l'utilisateur (nécessitant une modélisation « Explicite » des événements).

- Des alertes d'événements non-définis par l'utilisateur (avec une modélisation « implicite » des événements).

7.5.1 Alertes d'événements définis – modélisation explicite :

Dans cette approche, on trouve tous les systèmes qui nécessitent la définition explicite de ce qui constitue des événements normaux ou anormaux.

Trois types sont définis : [24]

7.5.1.1 Alertes basées sur le mouvement d'objets :

Ces alarmes dépendent des propriétés des mouvements des objets dans l'espace surveillée, par exemple détection d'objets abandonnés, un bagage sans surveillance dans un aéroport, ou une voiture garée dans une zone interdite. [24]

7.5.1.2 Alertes basées sur la classe d'objets :

Ces alarmes prennent en compte le type d'objets en plus des propriétés de leurs mouvements, par exemple la surveillance dans aéroport, le système déclencherait une alerte en cas de présence ou de mouvement spécifique d'individus sur le tarmac, mais pas celles des avions. [24]

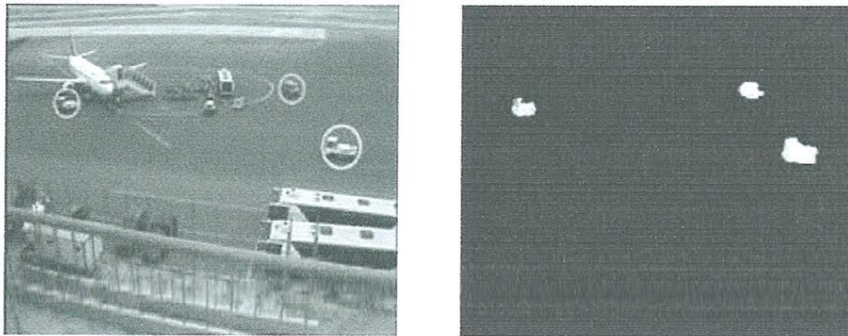


Figure 1.18 : Exemples de surveillance d'un aéroport [24]

7.5.1.3 Alertes basées sur le comportement :

Ces alarmes sont généralement basées sur l'apprentissage et l'analyse des habitudes de déplacement d'objets sur de longues périodes de temps. Ces alarmes sont utilisées dans des applications spécifiques et utilisent une quantité importante d'informations de contexte, par exemple, Détection de comportements suspects dans les parkings de stationnement, par exemple, une personne essayant d'ouvrir plusieurs voitures. [24]

7.5.2 Alertes d'événements non-définis – modélisation implicite :

Contrairement aux alertes d'événements définis, ici aucune connaissance a priori n'est fournie, le système génère des alertes quand il détecte une activité qui s'écarte de la norme ou dite « activité habituelle ». Par exemple, lors de la surveillance d'une rue, le système apprend que les véhicules se déplacent sur la route et les personnes se déplacent sur le trottoir. [24]

8. Les différents systèmes de vidéosurveillance existants :

Dans cette section, nous présentons quelques systèmes de vidéosurveillance en développement ou finalisés, chez les chercheurs, mais aussi les industriels.

8.1 Le projet ADVISOR-INRIA :

ADVISOR « Annotated Digital Video for Surveillance and Optimised Retrieval » : est un projet européen de l'équipe ORION de l'INRIA Sophia-Antipolis, en vidéo surveillance multi-caméras, impliquant trois partenaires académiques (Univ. Kingston, Univ. Reading, KCL Londres) et trois partenaires industriels (THALES, BULL, VIGITEC), entre janvier 2000 et mars 2003. Ce système a pour objectif de sécuriser les transports publics par la détection automatique de situations anormales temps réel pouvant conduire à des accidents, de la violence ou des actes de vandalisme. [22]

8.2. Le projet CAVIARE-INRIA :

CAVIARE « Context Aware Vision Picture-based Active Recognition » : est un projet européen de l'INRIA Grenoble, qui étudie les techniques d'analyse d'images pour améliorer les performances des systèmes de surveillance dans les environnements urbains et les centres commerciaux. [23]

8.3 Le projet CASSIOPEE-INRIA :

Le projet **CASSIOPEE** du groupe Crédit Agricole avec Eurotelis et Ciel pour la conception de sites de vidéosurveillance bancaire, a pour objectif de détecter des comportements définis comme « à risque » à partir d'une acquisition vidéo continue et d'une connaissance a priori de l'agence. L'objectif n'est pas de détecter à coup sûr les comportements, mais plutôt de détecter des comportements potentiellement intéressants afin que leur prise en charge soit réalisée par la station centrale de vidéo surveillance pour lever le doute.

Le projet regroupe les compétences d'une banque, d'un intégrateur de systèmes d'acquisition vidéo, d'un opérateur de télésurveillance et de l'INRIA. [2]

Conclusion :

Dans ce chapitre, nous avons présenté une littérature sur la vidéosurveillance et nous avons pu découvrir l'étendue des usages de ces systèmes ainsi que les différentes manières dont ils pouvaient être exploités pour l'analyse de scène et le déclenchement automatique d'alertes. Nous avons par ailleurs pu mettre en évidence l'avancement des méthodes de détection d'objets mobiles, des approches de reconnaissance ainsi que les méthodes de suivi d'objets mobiles et à la fin nous avons présenté quelques systèmes de vidéosurveillance existants.

Chapitre II : conception du système

Introduction

1. Objectif du système
2. Conception architecturale
 - 2.1 Acquisition du flux vidéo
 - 2.2 Fragmentation
 - 2.3 Détection par GMM
 - 2.4 Post traitement.
 - 2.5 Extraction des points extrêmes (Suivi)
 - 2.6 Segmentation LPE
 - 2.7 Squelettisation
 - 2.8 Extraction des paramètres
 - 2.9 Reconnaissance des comportements suspects

Conclusion

Chapitre II

Conception du système

Introduction

La surveillance, qui était autrefois accomplie uniquement par des agents de sécurité, a vu arriver la technologie comme une révolution. Alarmes, vidéosurveillance permettent aux agents d'être plus efficaces, car ils visualisent directement de leur poste de sécurité tous les points clé des locaux à surveiller. Pourtant l'homme reste imparfait. En effet, qui pourrait rester devant des écrans de surveillance sans aucune faille ? C'est impossible. C'est là qu'apparaît la vidéosurveillance, elle permet d'aider les agents dans leur travail. Le programme peut détecter une intrusion à tout moment et en avertir l'agent qui prendra les mesures nécessaires. La surveillance en est rendue bien plus efficace et moins contraignante pour les agents.

Ce chapitre est une représentation des détails et des approches théoriques utilisées dans le cadre de la conception de notre système de détection de comportements suspects dans une séquence vidéo.

1. Objectif du système :

L'objectif de notre travail est de mettre en place un outil logiciel qui permettra de détecter, suivre et reconnaître quelques comportements suspects à partir d'une base de vidéos que nous avons conçus.

Les principaux objectifs à atteindre sont :

- Acquisition d'un flux vidéo à partir d'une base créée dans des conditions de nature différente.
- Fragmentation du flux vidéo et la détection de l'objet en mouvement (l'individu) par la méthode GMM.
- Suivi de l'objet à travers l'extraction de ces points extrêmes.
- Extraction de la silhouette (approche contours, approche squelette).
- Extraction des paramètres du squelette.

- Apprentissage pour la reconnaissance de quelques comportements suspects.
- Déclenchement d'alerte.

2. Conception architecturale :

Le système réalisé est constitué des étapes illustrées par le diagramme suivant :

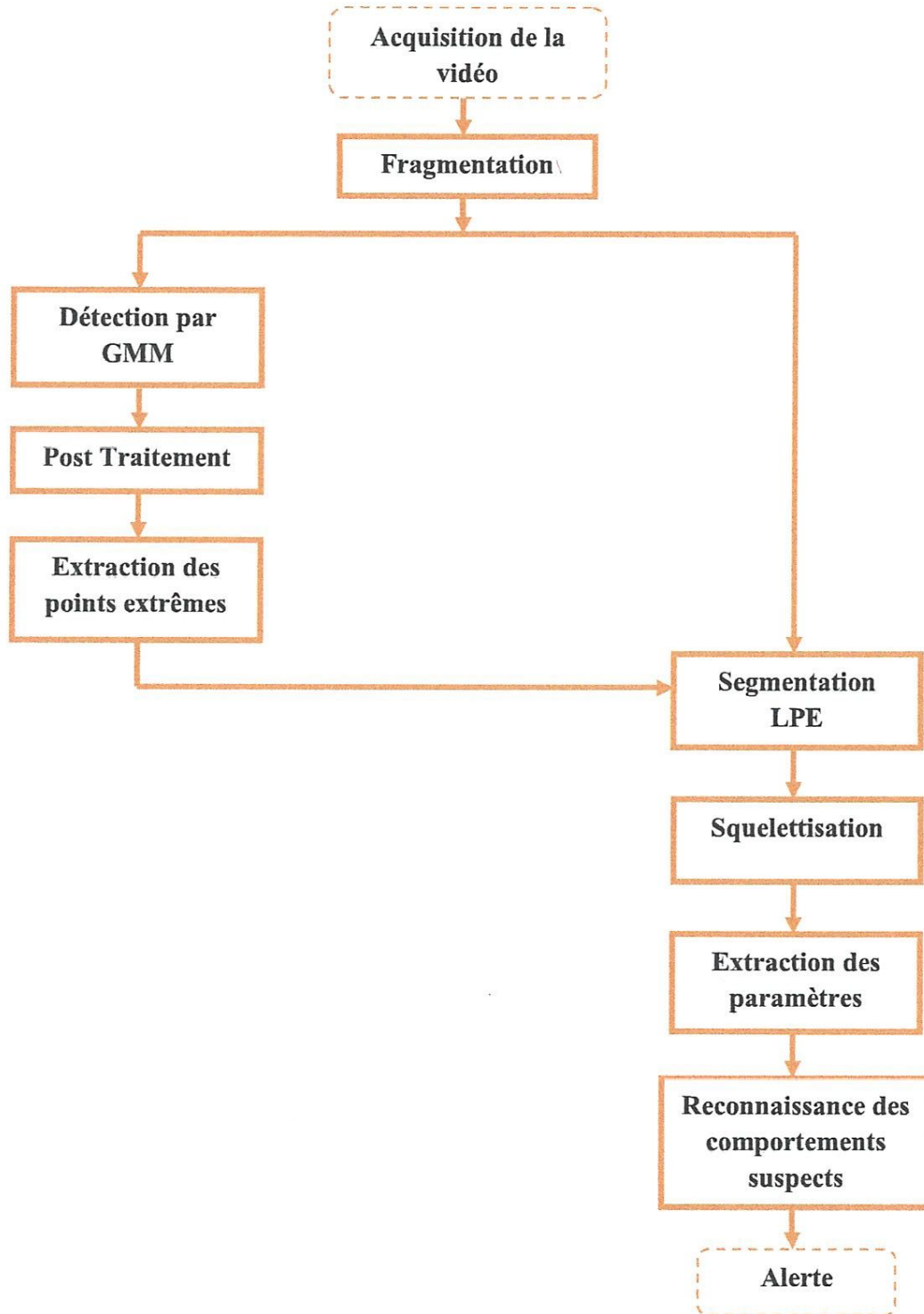


Figure II.1 : Architecture globale du système.

2.1 Acquisition du flux vidéo :

L'entrée de notre système est un flux vidéo lue à partir d'une base créée dans des conditions de nature différente par l'intermédiaire du web Cam du PC utilisé.

Ce flux représente une séquence d'images à une cadence de 24 images par seconde, sachant que cette dernière influence sur la détection, le suivi et la reconnaissance du mouvement.

2.2 Fragmentation :

La fragmentation est la décomposition du flux vidéo en un ensemble d'images, pour cela on a utilisé ces instructions :

➤ **Récupérer un contrôle sur le flux :**

```
FrameGrabbingControl fgc = (FrameGrabbingControl)Afficher.getControl  
(‘javax.media.control.FrameGrabbingControl’);
```

➤ **Mettre le résultat dans un buffer :**

```
Buffer = fgc.grabFrame();
```

➤ **Convertir le buffer en un buffer d'image :**

```
Buffer_image = new BufferToImage((VideoFormat)Buffer.getFormat());
```

➤ **Créer l'image :**

```
Buffer_image.createImage (buffer);
```

L'ensemble d'images de sorties sera traité par la suite.

2.3. Détection par GMM :

Une grande partie de la littérature concernant la soustraction de fond est inspirée de Stauffer et Grimson. Leur approche permet notamment de gérer les variations graduelles de luminosité ainsi que les fonds multimodaux. Le principe est le suivant : Dans l'espace de couleur RVB chaque pixel est représenté par un mélange de k gaussiennes pondérées (en pratique k est souvent fixé entre 3 et 5). Ces gaussiennes vont servir à représenter les distributions d'un historique récent des modes du fond et des objets (formes). Ainsi, pour un pixel donné à un temps t la probabilité d'observer X_t est : [6]

$$P(X_t) = \sum_{i=1}^k \omega_{i,t} \eta(X_t; \mu_{i,t}, \Sigma_{i,t}) \quad (\text{II.1})$$

Où chaque distribution gaussienne est représentée par sa moyenne $\mu_{i,t}$, sa matrice de covariance $\Sigma_{i,t}$, est pondérée par $\omega_{i,t}$. $\omega_{i,t} \in [0, 1]$ est significatif du nombre de fois que le mode i a été observé. η est la densité de probabilité de la loi normale :

$$\eta(X; \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{3}{2}} |\Sigma|^{\frac{3}{2}}} e^{-1/2(X-\mu)^T \Sigma^{-1} (X-\mu)} \quad (\text{II.2})$$

Durant la phase d'initialisation les paramètres $\omega_{i,t}$, $\mu_{i,t}$ et $\Sigma_{i,t}$ sont estimés à l'aide d'une approximation en ligne de l'algorithme K-means. Durant la phase de classification, les modes sont classés en modes représentant le fond et modes représentant l'objet. Pour cela ils sont ordonnés par $r_{i,t} = \omega_{i,t} / \sigma_{i,t}$ décroissant.

Les modes que l'on considère comme représentant le fond sont les premiers modes (pour cette nouvelle relation d'ordre) dont la somme des poids excède un certain seuil T . Ainsi si les $\omega_{i,t}$ sont ré-indices de sorte que $r_{i,t} \geq r_{i+1,t}$ alors les modes représentant le fond sont les B premiers :

$$B = \text{arg } \min_b \left(\sum_{k=1}^b \omega_{k,t} > T \right) \quad (\text{II.3})$$

Où T représente la proportion des observations représentant le fond. Il est donc possible que le fond soit représenté par plusieurs modes. Quant aux modes restants, ils représentent les objets, dont la présence n'est que temporaire. Si X_t est la nouvelle valeur du pixel, un test d'appartenance est effectué pour lui attribuer un mode et donc le classer comme fond ou objet :

$$\sqrt{(X_t - \mu_{i,t})^T \Sigma_{i,t} (X_t - \mu_{i,t})} < k \sigma_{i,t} \quad (\text{II.4})$$

Où k est une constante fixée à $k = 2, 5$

Si le pixel appartient effectivement à l'un des modes existants, les paramètres de ce mode sont mis à jour de la manière suivante :

$$\omega_{i,t+1} = (1 - \alpha) \omega_{i,t} + \alpha \quad (\text{II.5})$$

$$\mu_{i,t+1} = (1 - \rho) \mu_{i,t} + \rho X_t \quad (\text{II.6})$$

$$\sigma_{i,t+1}^2 = (1 - \rho) \sigma_{i,t+1}^2 + \rho (X_t - \mu_{i,t+1})^2 \quad (\text{II.7})$$

$$\text{Avec: } \rho = \alpha \eta (X_t, \mu_{i,t}, \sigma_{i,t}) \quad (\text{II.8})$$

Où $\alpha \in [0, 1]$ est le coefficient d'apprentissage de l'algorithme, il contrôle la vitesse d'adaptation du modèle aux observations. Pour les autres modes, seul le poids est mis à jour :

$$\omega_{i,t+1} = (1 - \alpha) \omega_{i,t} \quad (\text{II.9})$$

Si le pixel n'a pu être affecté à aucun des modes existants alors le mode de poids le plus faible est supprimé puis remplacé par un nouveau avec des paramètres initialisés arbitrairement :

$$\omega_{i,t+1} = \omega_0 \quad (\text{II.10})$$

$$\mu_{i,t+1} = X_t \quad (\text{II.11})$$

$$\sigma_{i,t+1}^2 = \sigma_0 \quad (\text{II.12})$$

Avec ω_0 un poids faible et σ_0 une grande variance.

L'algorithme est fondé sur le principe suivant :

➤ **Conversion de l'image en niveau de gris :**

Chaque pixel d'une image couleur possède 3 composantes couleurs (Rouge, Vert, Bleu), la conversion de cette image en niveau de gris est calculée par la moyenne de ces 3 composantes RGB et utiliser cette valeur moyenne pour chacune des composantes.

L'algorithme suivant permet un gain de temps et d'espace mémoire

Pour chaque pixel :

```
{  
    Rouge = récupérer la valeur de la couleur rouge du  
    pixel ;  
    Vert = récupérer la valeur de la couleur verte du pixel ;  
    Bleu = récupérer la valeur de la couleur bleue du pixel ;  
    NDG = ((Rouge+Vert+Bleu)/3) ;  
    NvPixel = (0xff000000|Rouge<<16|Vert<<8|Bleu) ;  
}
```

➤ **Apprentissage :**

Les valeurs prises par les pixels du fond sont modélisées par plusieurs gaussiennes qui sont comprises en 2 et 7.

L'avantage de cette étape est de garder le fond est sauvegarder ces caractéristiques pour cela, on laisse passer pendant quelques secondes, que des images du fond (environs 30).

➤ **Extraction du fond :**

Après la mémorisation des caractéristiques du fond dans la phase précédente, on laisse passer l'objet en mouvement (l'individu) afin de le séparer du fond.

Pour cela, on calcule la distance entre chaque pixel de l'image et chacune des gaussiennes qu'ils l'associent, puis on passe à la mise à jour des paramètres selon le résultat précédent.

L'algorithme est défini comme suit :

Pour chaque pixel :

```

{
  Pour i allant de 1 a k faire
    {
      Si ( $\frac{|X_t - \mu_t|}{\sigma_t} < 2,5$ ) {
        Existe = 1 ;
        Tab[i] = 1 ;
      }
      Sinon Tab[i] = 0 ;
    }
  Si (Existe==1)
    {
      Pour i allant de 1 a k
        {
          Si (Tab[i] == 1)
            {
               $\omega_{i,t} = (1 - \alpha)\omega_{i,t-1} + \alpha$  ;
               $\varphi_i = \alpha \eta(X_t / \mu_i \sigma_t)$  ;
               $\mu_{i,t} = (1 - \varphi_i)\mu_{i,t-1} + \varphi_i X_t$ 
               $\sigma_{i,t}^2 = (1 - \varphi_i)\sigma_{i,t}^2 + \varphi_i (X_t - \mu_{i,t})^T (X_t - \mu_{i,t})$  ;
            }
          Sinon  $\omega_{i,t} = (1 - \alpha) \omega_{i,t-1}$  ;
        }
      }
    }
  Sinon {  $\sigma_i^2 = \sigma_i^2$  ;  $\omega_t = \varphi_t$  ;  $\mu_t = X_t$  ; }

```

Normalisation ;

Les indices :

K : Le nombre de gaussienne associé a chaque pixel.

X_t : La valeur du pixel courant.

μ_{i,t} : La moyenne de la *i*^{ème} gaussienne dans le temps **t**.

σ_i : Ecart-type de la *i*^{ème} gaussienne dans le temps **t**.

ω_{i,t} : Le poids calculé pour la *i*^{ème} gaussienne dans le temps **t**.

α : Le taux d'apprentissage.

Les variables de contrôles : Existe, Tab.

Les Formules :

$\frac{|X_t - \mu_i|}{\sigma_i} < 2,5$: La distance entre le pixel **X** et chacune des gaussiennes.

$\omega_{i,t} = (1 - \alpha)\omega_{i,t-1} + \alpha$: La mise à jour des poids des gaussiennes.

$\mu_{i,t} = (1 - \alpha)\mu_{i,t-1} + \alpha X_t$: La mise à jour des moyennes de gaussiennes.

$\sigma_{i,t}^2 = (1 - \alpha)\sigma_{i,t}^2 + \alpha(X_t - \mu_{i,t})^T(X_t - \mu_{i,t})$: La mise a jour des écart-type des gaussiennes.

Soustraction fond/forme :

Cette étape a pour but de déterminer si le pixel appartient au fond, en ordonnant les **k** gaussiennes selon la valeur : $\omega_{i,t}/\sigma_{i,t}$, ensuite calculer la valeur de **B** selon la formule suivante :

$$\mathbf{B} = \text{argmin} (\sum_{i=1}^k \omega_{i,t} > T)$$

Donc les **B** premières sont les gaussiennes correspondantes au fond.

A la fin l'image sera convertie en une image binaire ou le pixel appartenant au fond prend la valeur 0 sinon il prend la valeur 1.

2.4 Post traitement :

La méthode utilisée pour l'extraction du fond produit parfois de fausses détections qui se manifestent par des pixels blancs isolés, ainsi que des pixels non détectés qui se manifestent par des trous dans l'objet.

Pour cela nous avons appliqué un filtrage morphologique, ce dernier permet de modifier la valeur d'un pixel de l'image en fonction de la valeur de ces voisins en utilisant des masques.

- **Dilatation** : permet de parcourir tous les pixels de l'image excepté les bords, et consiste à boucher les trous sur le corps de l'individu selon l'algorithme suivant :

```
Test = vrai ;
Nb : nombre de voisinage du pixel portant la couleur blanche ;
Tant que (Test) faire {
    Test= faux ;
    Pour chaque pixel
        {
            Si
            ((masque1)or(masque2)or(masque3)or(masque4)or(masque5)or(masque6))
            or ((nb=3) and ((masque7)or(masque8)))
            or ((nb=4) and ((masque7)or(masque8)or(masque9)))
            or (nb>=5)    {
                Pixel=couleur blanche ;
                Test=vrai ;
            }
        }
    }
}
```

Chapitre II : Conception du système

Les masques utilisés sont définis ci-dessous :

1	1	1
1	X	0
1	0	0

Masque1

0	0	1
0	X	1
1	1	1

Masque2

0	0	1
0	X	0
1	0	0

Masque3

1	0	0
0	X	0
0	0	1

Masque4

?	1	?
?	X	?
?	1	?

Masque5

?	?	?
1	X	1
?	?	?

Masque6

1	?	?
?	X	?
?	?	1

Masque7

?	?	1
?	X	?
1	?	?

Masque8

?	1	?
1	X	1
?	1	?

Masque9

1 : Signifie la couleur blanche.

0 : Signifie la couleur Noire.

? : Le pixel peut prendre la valeur 0 ou 1.

- **Erosion (éliminer des points isolés)** : consiste à balayer l'image de façon à supprimer les points rebelles selon l'algorithme suivant :

```
Test= vrai ;  
Nb : nombre de voisinage du pixel portant la couleur noire ;  
Tant que (Test) faire  
  {  
    Test=faux ;  
    Pour chaque pixel  
      {  
        Si (nb=8)  
          {  
            Pixel= couleur noir ;  
            Test=vrai ;  
          }  
        }  
      }  
  }
```

2.5 Extraction des points extrêmes (Suivi):

C'est la localisation spatiale et temporelle de l'objet en mouvement (l'individu) dans la vidéo. Le principe de cette étape consiste à mettre l'objet dans une boîte englobante. Lorsque plusieurs objets sont présents dans la scène, on applique les traitements restants sur le plus grand cadre qui représente l'individu.

Afin de dessiner les cadres, on suit les étapes suivantes :

Calculer l'histogramme horizontal de l'image.

Compter le nombre des objets Horizontaux.

Deviser l'image en sous image selon le nombre d'objets trouvées.

Pour chaque sous image

{

Calculer l'histogramme vertical.

Compter le nombre des objets verticaux.

Extraire les extrémités des objets verticaux.

Dessiner les rectangles englobant les objets.

Extraire le plus grand cadre.

}

2.6 Segmentation LPE :

La segmentation des images numériques est une méthode permettant de regrouper les différents pixels qui composent une image en région, c'est-à-dire en ensemble de points connexes partageant un certain critère d'homogénéité.

De nombreuses techniques existent ; celle que nous avons développée c'est la ligne de partage des eaux (LPE).

La LPE possède, en traitement d'images, la même signification que celle qui nous permet d'appréhender les bassins versants des différents fleuves d'un pays. Ce sont les lignes de crête qui délimitent ces bassins et qui forment donc la ligne de partage des eaux.

Pour segmenter une image, il est possible de réaliser dans un premier temps une détection de contours avec un filtre de Sobel, voici les masques de ce dernier :

➤ **Masque verticale :**

-1	0	1
-2	0	2
-1	0	1

➤ Masque horizontale :

-1	-2	-1
0	0	0
1	2	1

Après la détection de contours, nous considérons une image comme un tableau à deux dimensions dont chaque case porte une valeur de niveau de gris. Nous commençons par ordonner l'ensemble des pixels (les cases du tableau) par niveau de gris croissant, puis l'inondation par les pixels dont le niveau de gris est le plus bas.

Nous allons donc utiliser un algorithme à pile FIFO. Les pixels à examiner seront placés dans une pile, laquelle sera vidée à un autre moment du processus d'inondation pour découvrir la LPE. Le processus se décompose en plusieurs étapes :

- Elaboration d'un tableau trié des pixels par ordre croissant de NDG.
- Parcours des pixels d'iso-niveau, pour les marquer et les placer dans la pile en fonction de leur voisinage.
- Parcours de la pile de pixels pour décider s'ils appartiennent à la LPE et leur attribuer une valeur ou attribuer une valeur à leurs voisins.
- Deuxième parcours des pixels pour les placer dans la pile ou placer leurs voisins.

A l'issue de ces étapes les pixels auront obtenu une valeur qui correspondra à :

- ✓ Une appartenance à la LPE caractérisée par un niveau = 0.
- ✓ Une appartenance à un bassin versant caractérisée par un niveau >0

Tous les pixels appartenant à un même bassin versant posséderont le même niveau.

2.7 Squelettisation :

L'intérêt de la squelettisation est la reconnaissance de forme, elle permet d'affiner les contours. Il existe plusieurs techniques pour obtenir le squelette d'un objet et celle qu'on a utilisée est la technique d'amincissement successive (thinning) et s'arrêter lorsque l'amincissement n'apporte plus de changement.

Pour l'algorithme de squelettisation, on considère les définitions suivantes :

1. P_0 est un point frontière Nord si $P_2 = 0$.
2. P_0 est un point frontière Est si $P_4 = 0$.
3. P_0 est un point frontière Sud si $P_6 = 0$.
4. P_0 est un point frontière Ouest si $P_8 = 0$.
5. P_0 est un point 8-terminal si un seul de ses voisins est noir, il s'agit en fait d'une minutie de type terminaison.
6. P_0 est un point 8-isolé si aucun de ses voisins n'est noir.
7. P_0 est un point 8-simple si la connexité de ses 8 voisins n'est pas altérée quand on le transforme en pixel blanc.

La squelettisation consiste à répéter les opérations d'érosion suivantes jusqu'à ce que plus aucun pixel ne soit changé:

- Tous les pixels noirs vérifiant (1) et (7) et ne vérifiant pas (5) et (6) sont changés en pixels blancs (érosion des points frontières Nord).

- Tous les pixels noirs vérifiant (2) et (7) et ne vérifiant pas (5) et (6) sont changés en pixels blancs (érosion des points frontières Est).

- Tous les pixels noirs vérifiant (3) et (7) et ne vérifiant pas (5) et (6) sont changés en pixels blancs (érosion des points frontières Sud).

- Tous les pixels noirs vérifiant (4) et (7) et ne vérifiant pas (5) et (6) sont changés en pixels blancs (érosion des points frontières Ouest).

2.8 Extraction des paramètres :

En statistiques, l'histogramme est un graphique que l'on utilise pour observer la répartition des valeurs que prend une variable aléatoire. En ce qui nous concerne, nous allons calculer l'histogramme dit « d'allure » qui détermine l'aspect ou l'apparence de l'individu présent dans l'image, cela va nous permettre de définir son attitude.

L'algorithme de calcul d'histogramme d'allure est défini comme suite :

```
Début ;  
Initialiser l'histogramme ;  
Pour chaque pixel ;  
    Extraire les composantes Rouge, Verte et Bleue ;  
    Calculer la moyenne des composantes couleurs RGB ;  
Parcourir l'image résultante;  
    Histogramme [largeur]=Histogramme [largeur] + image [hauteur][largeur] ;  
Fin ;
```

2.9 Reconnaissance de comportements suspects :

Pour la reconnaissance de comportements suspects, nous avons eu recours aux PMC.

Le perceptron multi couches (PMC) est un type particulier de réseaux de neurones artificiels, il est organisé en couches ou :

- Chaque couche est un groupe de neurones uniformes sans connexions les uns avec les autres.
- Une couche reçoit un vecteur d'entrée et le transforme en vecteur de sortie.
- Les dimensions d'entrée et de sortie peuvent être différentes.
- Sa spécificité est son fonctionnement dit en passe avant, parce que l'information qui y circule ne le fait que dans un sens ; il n'y a donc pas de boucles au niveau des connexions neuronales.

- Possède au moins deux couches, une dite cachée et une couche de sortie.

Le PMC que nous avons conçu a comme paramètre d'entrée 640 neurones, qui représentent la largeur des images utilisées, deux couches cachées qui peuvent contenir autant de neurones que l'on veut, et une couche de sortie contenant un seul neurone pour dire que le comportement est suspect ou non suspect.

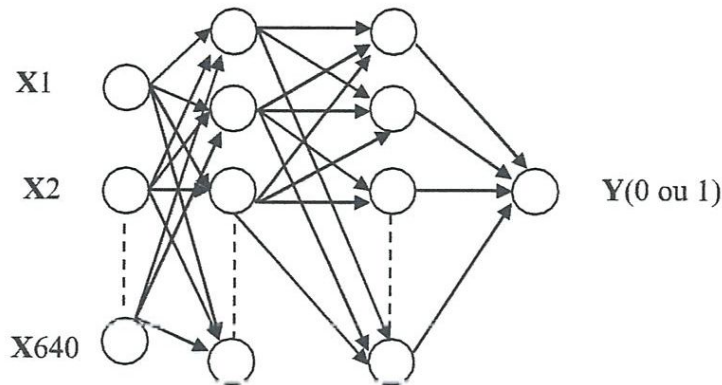


Figure II.2 : Perceptron multi couches utilisées

Un PMC, comme tout autre réseau de neurones, est paramétré non seulement par le nombre de neurones qui le constituent (sur les couches d'entrée, cachée et de sortie), le graphe de connectivité, mais également par les poids des connexions qui lient les neurones entre eux. Suivant l'inspiration biologique des PMC, deux neurones qui sont connectés avec un arc muni d'un poids positif s'activent simultanément et inversement. Des neurones réalisant des tâches similaires sont donc censés s'activer ensemble. [30]

Un autre paramètre qui entre en jeu dans la définition d'un PMC est la fonction d'activation ou fonction de transfert associée à chaque neurone : cette fonction constitue précisément le calcul élémentaire réalisé par chaque neurone, elle est de type sigmoïde (car prenant la forme de S). [9]

La fonction est la suivante :

$$\sigma = \frac{1}{1 + \exp(-x)}$$

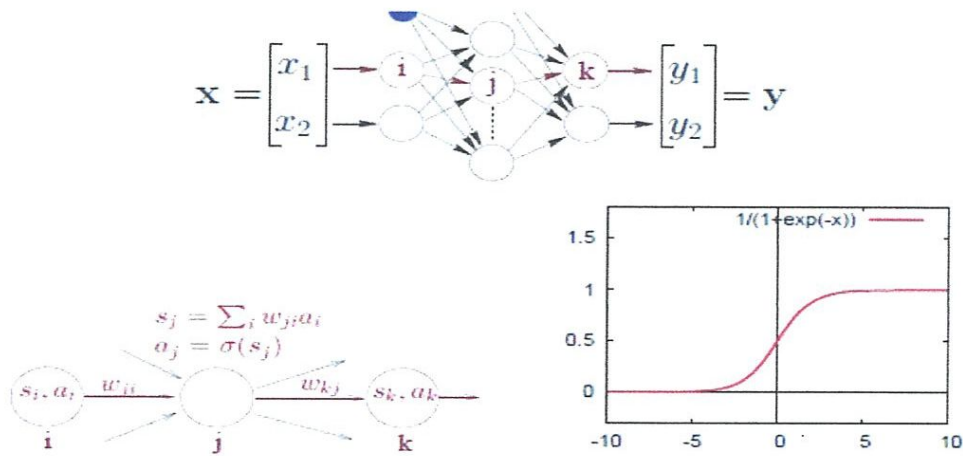


Figure II.3 : Détails des paramètres d'un PMC, mode de calcul de l'activation d'un neurone et fonction sigmoïde σ [9]

La figure représente en détail un PMC avec les poids w_{ji} (w_{ij} désigne le poids de la connexion du neurone i vers le neurone j), le mode de calcul de l'activation d'un neurone, et fournit une représentation de la fonction σ .

De manière plus détaillée : [9]

- les neurones de la couche d'entrée transmettent sans les modifier les signaux qu'elles reçoivent ; en d'autres termes, la fonction d'activation ou de transfert utilisée par ces neurones est la fonction identité $h(x) = x$.
- les neurones de la couche cachée et de la couche de sortie calculent leur activation de la même manière ; par exemple, pour le neurone j :
 1. l'intensité du signal reçu par j est calculée par :

$$S_j = \sum_{i \in \text{entrée}} w_{ji} a_i$$

2. son activation est alors calculée par :

$$a_i = \sigma(S_j)$$

Comme tout autre algorithme de reconnaissance, le PMC à deux modes de fonctionnement :

- **Phase d'apprentissage :** L'objectif premier de l'apprentissage est celui de déterminer automatiquement les poids des connexions entre les neurones, ou d'une autre manière comment déterminer w pour obtenir le comportement désiré ? Pour cela on utilise la méthode de descente du gradient (algorithme de rétro-propagation de gradient).

- **Phase de reconnaissance** : Comment évaluer la sortie Y pour une entrée X donnée ?

Voici les étapes de la phase d'apprentissage :

1. Présentation d'un exemple parmi l'ensemble d'apprentissages en fonction d'un critère donné.
2. Calcul de l'état du réseau.
3. Calcul de l'erreur = fonction (sortie - sortie désirée).
4. Calcul des gradients par l'algorithme de rétro-propagation du gradient.
5. Modification des poids.
6. Critère d'arrêt sur l'erreur, nombre de présentation d'exemples.
7. Retour en 1.

La classification d'un objet en fonction de l'activation d'un PMC se fait à la lecture de l'activation des neurones de la couche de sortie. Donc, attribuer à un objet la classe k si c'est le neurone k de la couche de sortie qui a l'activation la plus élevée. Ainsi, on souhaite qu'à chaque objet X à classer de classe k , le vecteur de sortie Y calculé par le PMC vérifie cette propriété. C'est ce critère qui permet de définir l'algorithme d'apprentissage par rétro-propagation du gradient.

L'algorithme est défini comme suit :

- Répéter jusqu'à convergence

Pour chaque pour chaque exemple (X_p ; Y_p) faire

1. propager l'information (calculer l'activation du PMC)

2. pour chaque neurone de sortie k calculer

$$\sigma_k \leftarrow a_k(1 - a_k)(y_{pk} - a_k)$$

3. pour chaque neurone caché j calculer

$$\sigma_j \leftarrow a_j(1 - a_j) \sum_{i \in \text{entrée}} \sigma_k w_{kj}$$

4. calculer les pas Δw_{ji}^t et Δw_{kj}^t par ($\eta > 0$)

$$\Delta w_{ji}^t = + \eta \sigma_j a_i \quad (\text{c.entrée à c.cachée})$$

$$\Delta w_{kj}^t = + \eta \sigma_k a_j \quad (\text{c.cachée à c.sortie})$$

5. mettre à jour les w_{ji} et les w_{kj} par

$$w_{ji} \leftarrow w_{ji} + \Delta w_{ji}^t \quad (\text{c.entrée à c.cachée})$$

$$w_{kj} \leftarrow w_{kj} + \Delta w_{kj}^t \quad (\text{c.cachée à c.sortie})$$

Conclusion :

Dans ce chapitre nous avons présenté les différents outils de développement utilisés, on a défini les objectifs de notre système, les différentes étapes de conception, ainsi que les différents algorithmes et les méthodes implémentées pour la réalisation de notre travail.

Chapitre III : Réalisation du système

Introduction

1. Aspect matériel
2. Outils de développement
 - 2.1. Eclipse
 - 2.2. JDK « Java Development Kit »
 - 2.3. JMF « Java Media Framework »
 - 2.4. Neuroph Studio
3. Description du système
4. Les types d'environnements
5. Tests et résultats
 - 5.1. La détection par GMM
 - 5.2. Post Traitement (Dilatation+Erosion)
 - 5.3. Suivi d'individu
 - 5.4. Squelettisation et détection de contours
 - 5.5. Résultats de la segmentation LPE
 - 5.6. Apprentissage et reconnaissance des comportements suspects

Conclusion

Chapitre III

Réalisation du système

Introduction

La réalisation d'un logiciel désigne la création d'un produit fini à partir d'un document nommé cahier des charges dont lequel on spécifie les besoins fonctionnels et non fonctionnels du système.

Ce chapitre est dédié à l'implémentation de notre application et à la présentation des différents composants matériels, qui présentent les caractéristiques matérielles de l'ordinateur utilisé lors de la validation, et logiciels qui présentent les langages de programmation utilisés pour son fonctionnement ainsi que les tests et les résultats obtenus après chaque phase d'exécution.

1. Aspect matériel :

Notre projet a été développé sur un micro portable :

- Processeur : Intel (R) Core (TM) i3
- Capacité Mémoire (RAM) : 2 Go.
- Vitesse d'horloge : 2.53 Ghz.
- Capacité disque dur : 500Go.
- Système d'exploitation : Windows 7.

2. Outils de développement :

Pour la réalisation de notre système, nous avons utilisé les logiciels suivants :

2.1. 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 (Java, PHP, C, C++, Ruby...etc).

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. [32]

Il existe plusieurs versions d'Eclipse et celle qu'on a utilisée est bien Eclipse IDIGO :

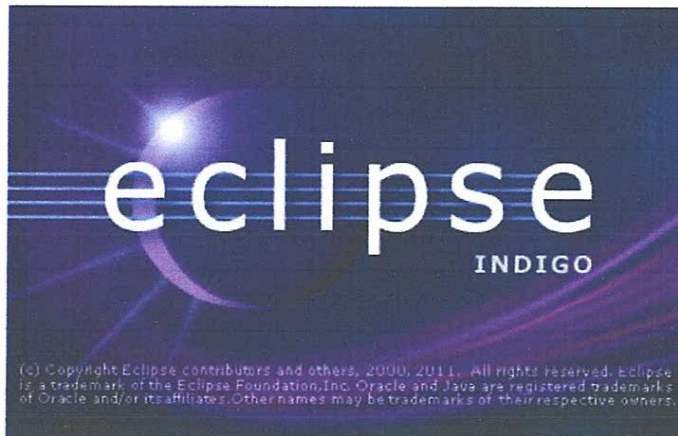


Figure III.1 : Version d'Eclipse utilisée

2.2. JDK « Java Development Kit » :

C'est un logiciel édité par Sun pour le développement d'application en Java, ou le code Java est compilé pour être transformé en bytecode afin que la machine virtuelle Java (JVM) puisse l'interpréter. [31]

2.3. JMF « Java Media Framework » :

JMF est une large et transparente API (Application Programming Interface) utilisée pour les applications multimédias (Son/Vidéo).

Elle permet :

- ✓ La lecture de différents formats multimédias tels qu'AVI, MPEG, WAVE.
- ✓ Capturer le son et la vidéo depuis un microphone et/ou une caméra et enregistrer les données dans un format supporté.
- ✓ Modifier les flux audio/vidéo en temps réel.

La version actuelle de JMF est 2.1.1e, cette dernière est une initiative de Sun qui souhaite apporter une solution « processeur de média basé sur un time line » à Java. [26]

2.4. Neuroph Studio :

Neuroph Studio est un outil utilisé pour le développement des réseaux de neurones et peut être intégré dans un programme Java à l'aide d'une bibliothèque nommée Neuroph Java Library.[34]

3. Description du système :

Notre système est composé de plusieurs parties :

- A. La détection de l'individu avec la GMM.
- B. L'application des prétraitements nécessaires afin de corriger les défauts issus après la détection et diminuer le bruit existant.
- C. La mise d'individu cible dans une boîte englobante.
- D. La détection des contours de l'individu, remplissage de la silhouette et l'extraction du squelette.
- E. L'apprentissage du squelette par un réseau de neurones pour la reconnaissance des comportements suspects.

4. Les types d'environnements :

Le système réalisé a été testé dans trois types d'environnements qui ont donné des résultats différents :

- **Environnement avec éclairage variant** : caractérisé par un milieu ouvert où la lumière est changeante.
- **Environnement simple** : caractérisé par un fond simple, et un seul objet en mouvement avec un éclairage stable.
- **Environnement avec faible éclairage** : caractérisé par un espace fermé.

Pour chaque environnement nous avons créé des vidéos de comportement défini suspect ou non suspect.

Après les tests réalisés la meilleure estimation des paramètres est dans la deuxième case du tableau, ou la gaussienne $\sigma=7$, $T=0.3$, et $\alpha = 0.01$ puisque l'objet en mouvement est bien détecté et bien représenté. Selon ces paramètres nous allons terminer les tests restants.

➤ Selon le type d'environnement :


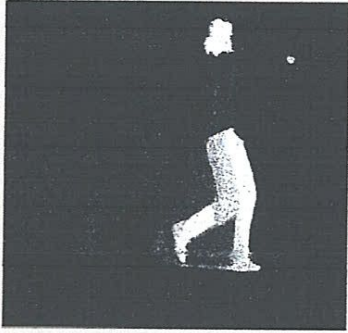

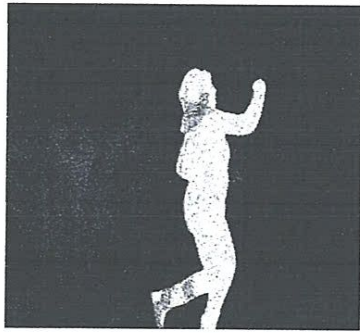


Type d'environnement	Image originale	Image résultante
Avec éclairage variant		
Simple		
Avec faible éclairage		

Tableau III.2 : Résultats du test de la GMM selon le type d'environnement

5. Tests et résultats :

5.1 La détection par GMM :

➤ Avec changement des paramètres :


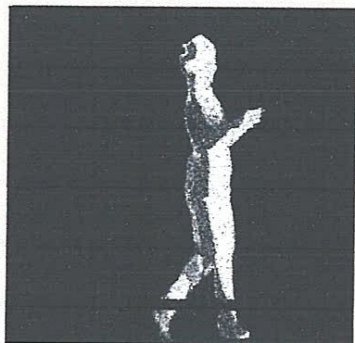
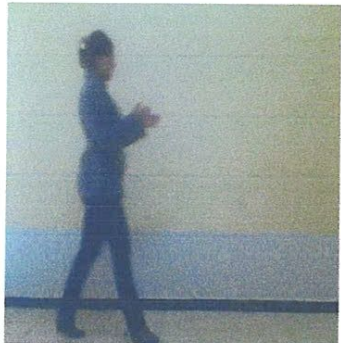
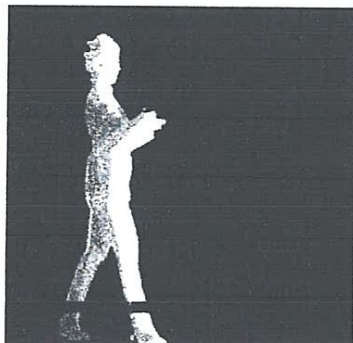
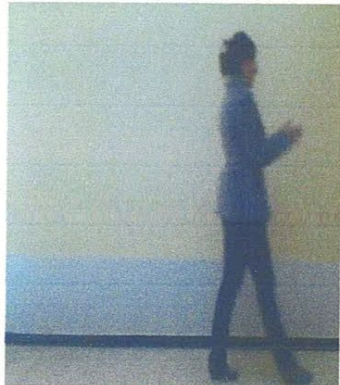
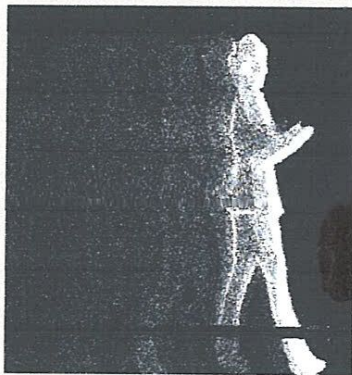
Paramètres	Image originale	Résultat GMM
<p>Gaussienne=7 $T=0.8$ $\alpha = 0.01$</p>		
<p>Gaussienne=7 $T=0.3$ $\alpha = 0.01$</p>		
<p>Gaussienne=3 $T=0.4$ $\alpha = 0.01$</p>		

Tableau III.1 : Résultats du test de la GMM avec changement des paramètres

Selon les tests effectués, la détection d'individu avec la GMM donne de bons résultats dans un environnement simple dont le fond est statique et la lumière est stable, ou plus au mieux dans un environnement fermé avec un éclairage faible. Tandis que dans un espace ouvert ou la lumière est changeante, elle détecte mal l'individu surtout quand ce dernier porte des vêtements dont la couleur est proche à la couleur du fond.

5.2 Post Traitement (Dilatation+Erosion):

La dilatation permet de bien définir l'objet, mais ce qui nous intéresse le plus est bien l'érosion, car elle sert à éliminer le bruit et ça va influencer sur les étapes suivantes.

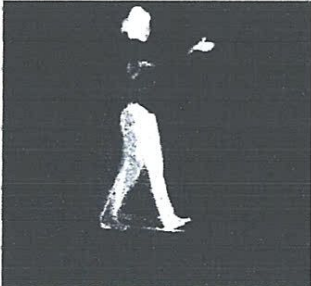



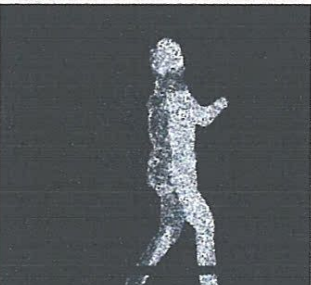

Type d'environnement	Image de GMM	Image traitée
Avec éclairage variant		
Simple		
Avec faible éclairage		

Tableau III.3 : Résultat du post traitement (dilatation + érosion)

5.3 Suivi d'individu:

Cette étape permet de mettre l'individu cible dans une boîte englobante afin de le suivre et d'appliquer le reste des traitements sur le cadre dessiné ci-dessous.

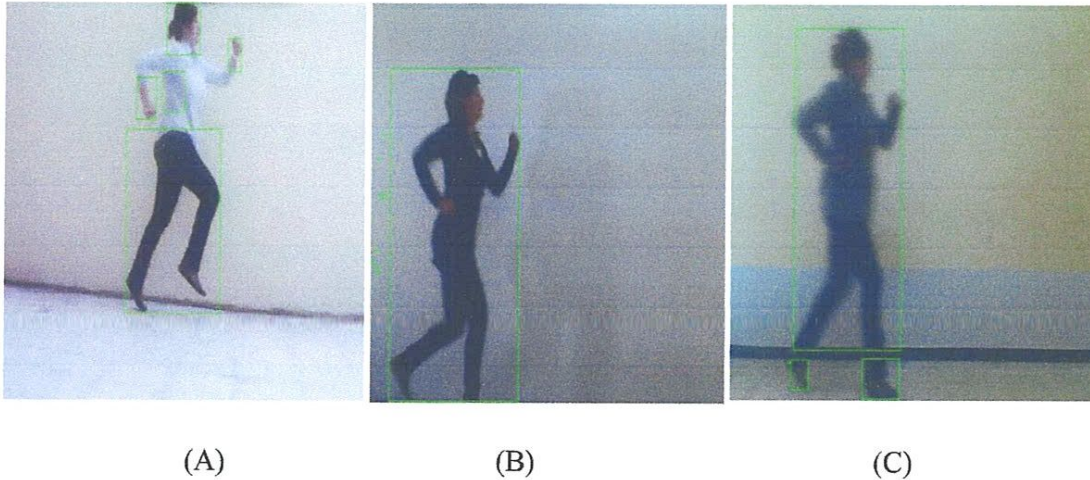


Figure III.2 : *Suivi d'individu.* (A) Dans un environnement avec éclairage variant, (B) Dans un environnement simple et (C) dans un environnement a faible éclairage.

5.4 Squelettisation et détection de contours :

Dans un premier temps nous avons appliqués la squelettisation directement après post traitement, mais le squelette extrait est déformé, car l'objet n'est pas totalement rempli.

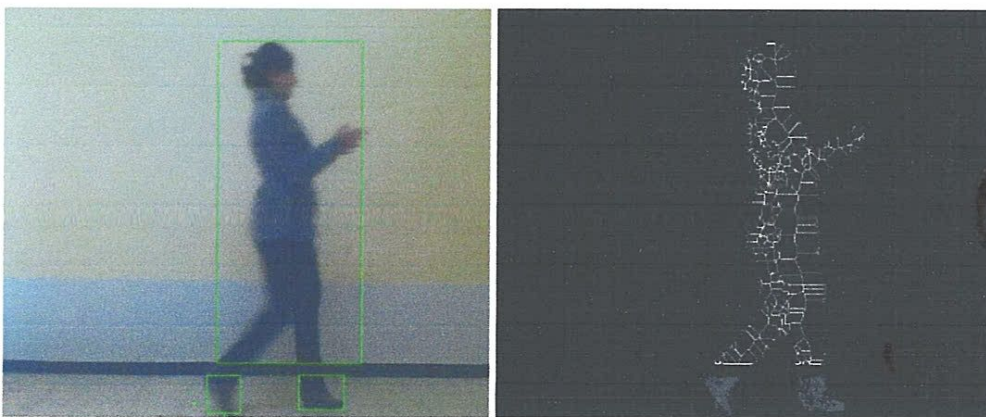


Figure III.3 : Résultat de la squelettisation après le post traitement

L'idée suivante était d'extraire le contour de l'objet englobé pour être examiné par la suite, comme l'objet est déformé les contours sont ouverts et ne servons à rien.

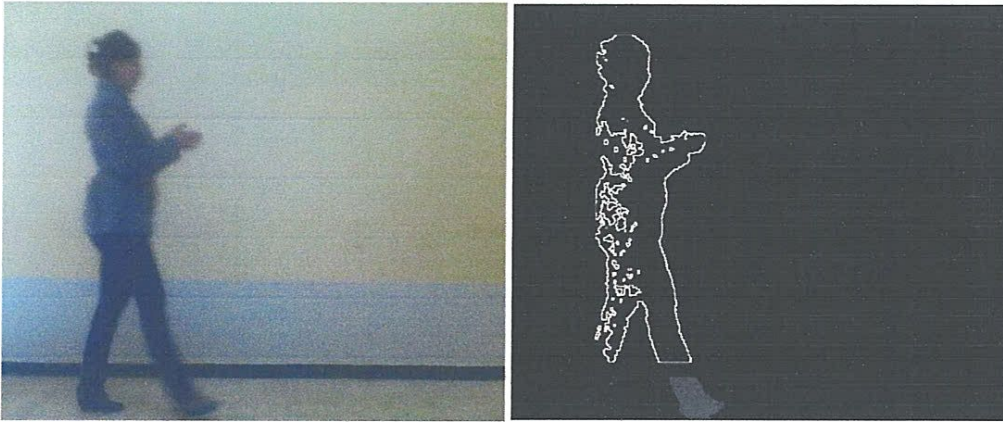


Figure III.4 : Résultat de la détection de contour après le post traitement

L'objet déformé ne nous a pas permis d'avancer même si on avait abouti à une dilatation successive, donc la seule solution pour dépasser ce problème est bien l'application d'une segmentation (LPE).

5.5 Résultats de la segmentation LPE :

➤ Détection de contours :



Figure III.5 : Résultat de la détection de contour avec filtre Laplacien



Figure III.6 : Résultat de la détection de contour avec filtre de Sobel

Selon les résultats obtenus, nous avons constaté que le filtre de Sobel est bien meilleur que celui de Laplace.

➤ **Remplissage de la silhouette :**



Figure III.7 : Silhouette d'un individu en marchant



Figure III.8 : Silhouette d'un individu claquant les mains

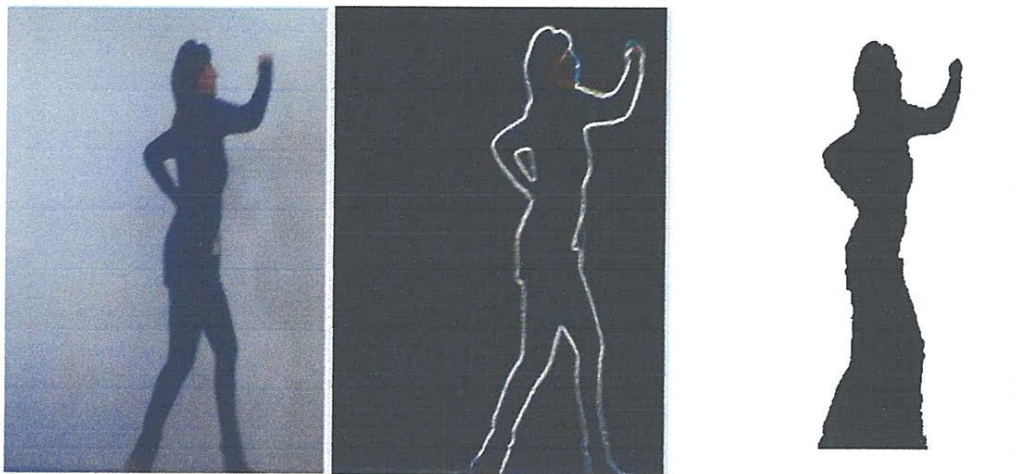


Figure III.9 : Silhouette d'un individu en courant

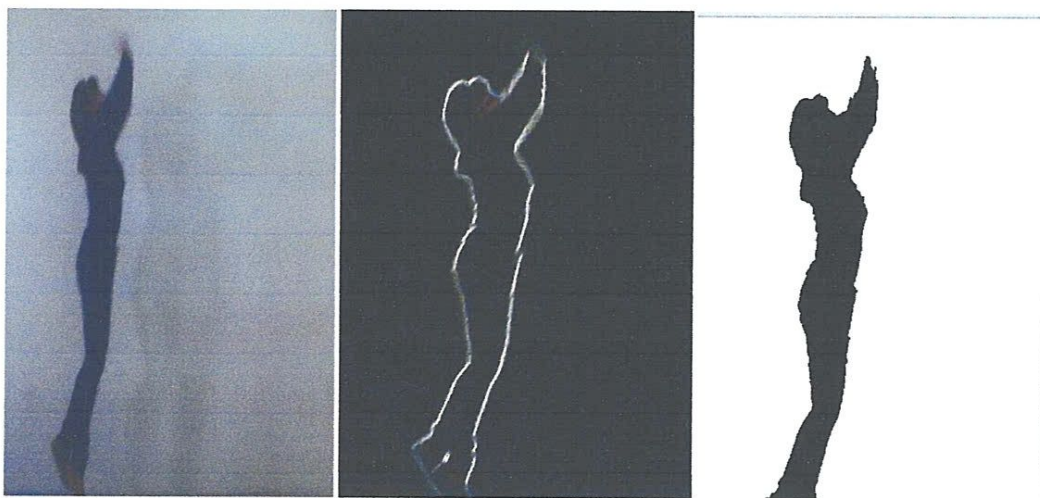


Figure III.10 : Silhouette d'un individu en sautant

➤ Squelettisation :

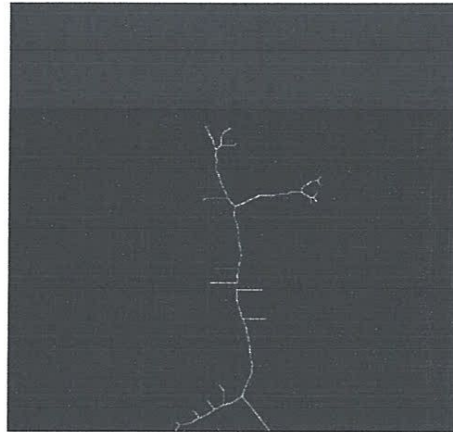


Figure III.11 : Squelette *d'*un individu claquant les mains

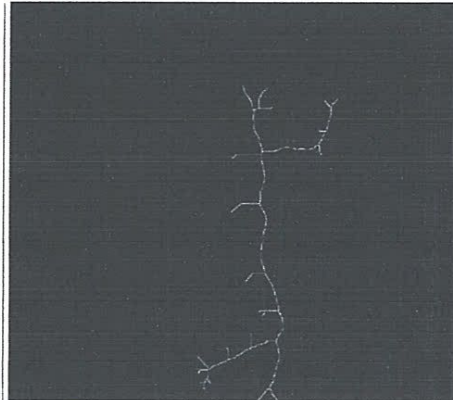


Figure III.12 : Squelette *d'*un individu en courant

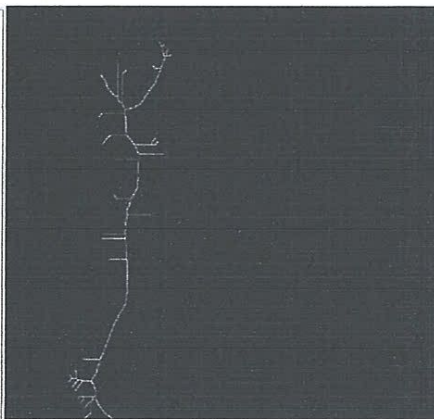


Figure III.13 : Squelette *d'*un individu en sautant

5.6 Apprentissage et reconnaissance des comportements suspects :

Le PMC utilisé à :

- Entrées = 640.
- 2 couches cachées (8,8).
- Sortie = 1. (0 : Non suspect, 1 : Suspect).
- Pas d'apprentissage = 0.2 ;
- Erreur maximale = 0.01.

Pour tester l'efficacité de notre système, nous avons créé deux bases :

- Une base d'Apprentissage qui comporte 45 images de squelettes dessinés.
- Une base de Test qui représente la vidéo d'entrée.

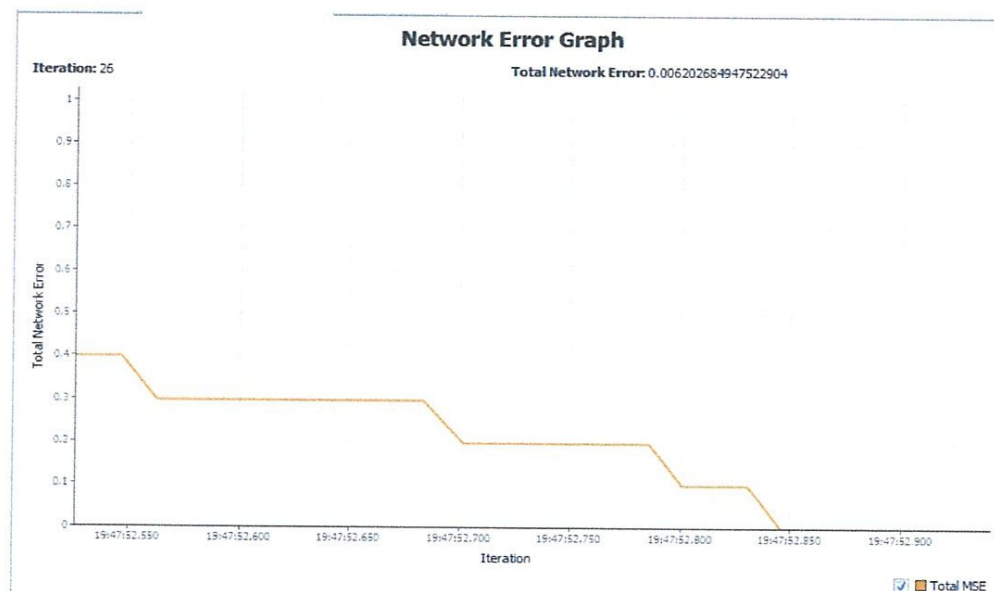


Figure III.14 : Graphe d'apprentissage.

Selon le graphe, l'apprentissage s'arrête après 26 itérations lorsque l'erreur totale est inférieure à 0.01.

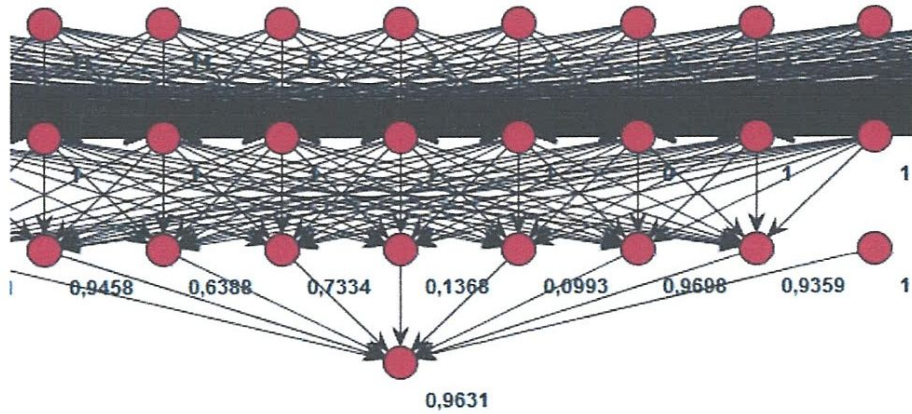


Figure III.17 : Résultat d'un comportement suspect.

Conclusion

Dans ce chapitre nous avons décrit notre système de détection de comportements suspects basé sur la méthode de détection GMM, de suivi par silhouette (contour, squelette) et de reconnaissance par PMC. Nous avons aussi présenté les tests et les résultats obtenus avec l'évaluation et la discussion de ces derniers.

Conclusion Générale et Perspectives

Vu la nécessité d'utiliser des applications de contrôle d'accès, la vidéosurveillance a émergé comme un secteur actif de recherches, enjambant des disciplines telles que le traitement d'images et la vision par ordinateur. Elle reste un sujet scientifiquement complexe dès lors que l'on souhaite intégrer un degré d'automatisation élevé tant dans la détection, le suivi et la reconnaissance d'un objet spécifique présent dans la scène.

Durant ces dernières années, de nombreuses méthodes de détection, de suivi et de reconnaissance ont été proposées, dont plusieurs ont été appliquées avec succès. Le choix d'une méthode doit être basé sur les conditions spécifiques de chaque application.

Dans un tel cadre, nous avons pu constater que parmi toutes les méthodes de détection, la mixture de gaussienne (GMM) reste l'une des approches les plus fiables, et le principal avantage de cet algorithme est de pouvoir gérer certains fonds dynamiques et de s'adapter aux variations graduelles de luminosité. Tandis que l'approche basée silhouette (squelette) est la méthode de suivi la plus utilisée et cela due au fait qu'elle permet une bonne localisation de l'individu cible afin de mieux décrire et reconnaître les actions ou comportements supposés suspects de ce dernier.

À travers les tests effectués, les résultats que nous avons obtenus sont en général satisfaisants et encourageants, mais nous suggérons de continuer à améliorer les algorithmes utilisés.

Actuellement, les systèmes de vidéosurveillance se tournent vers la combinaison de deux types de capteurs :

- Une caméra statique qui permet d'avoir à tout instant une vision globale de la scène.
- Une caméra dynamique qui peut à tout instant être pilotée afin d'obtenir une image de plus haute résolution de la cible.

Conclusion Générale et Perspectives

L'intérêt de combiner ces deux types de caméras dans un même système de vidéosurveillance est de pouvoir exécuter un suivi dynamique haute résolution d'une cible grâce à la caméra dynamique et plus robuste grâce à la vision globale de la caméra statique.

Enfin, nous espérons avoir pu communiquer aux lecteurs du présent mémoire la passion qui nous a animées tout au long de notre modeste incursion au sein de ce domaine.

Bibliographie et Web graphie

[1] : A. Ali et J.K. Aggarwal. « Segmentation and Recognition of Continuous Human Activity » . IEEE Workshop on Detection and Recognition of events in video. Vol.0 page 28, 2001.

[2] : A. Avanzi, F. Brémond, C. Tornieri, M. Thonnat « Design and Assessment of an Intelligent Activity Monitoring Platform ». In EURASIP Journal on Applied Signal Processing, Special Issue on « Advances in Intelligent Vision Systems: Methods and Application », Vol. 2005, N°14, pages 2359-2374, August 2005.

[3] : A. Singh, S. Sawan, M. Hanmandlu, V.K. Madasu, B.C. Lovell, « An abandoned object detection system based on dual background segmentation », Proceedings of the 6th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pages 352-357, Genova, Italy, 2009.

[4] : A. Yalmaz, X.Li et M.shah. « Contour Based Object Tracking with Occlusion Handling in Video Acquired Using Mobile Cameras ». IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol.26 n°11 page 1531, 2004.

[5]: A. Yoneyama, Y. Chia-Hung, and C. Jay Kuo, « Robust vehicle and traffic information extraction for highway surveillance », EURASIP Journal on Applied Signal Processing vol. (14), pages 2305-2321, 2004.

[6] : C. Guillot, « Détection d'objets stationnaires par une paire de caméras PTZ », Thèse de doctorat en informatique spécialité vision pour la robotique. Université Blaise Pascal-Clermont-Ferrand II, 2012.

[7] : C. Stauffer et W. Grimson. « Learning Patterns of Activity Using Real-time tracking ». IEEE Transaction on Pattern Analysis and Machine Intelligence, vol 22 n°8 pages 747-757, 2000.

[8]: C.R.Wren, A.Azarbayejani, T.Darell et A.P.Pentland. Pfunder : « Real-time tracking of the human body ». IEEE Transaction on Pattern Analysis and Machine Intelligence, vol 19 n°7 pages 780-785, 1997.

[9]: F. Rossi, « Réseaux de neurones : le perceptron multi couches », Cours Université Paris-IX Dauphine.

[10] : J. Badri, « Système de Vision Hybride » Thèse de doctorat en Automatique et Informatique Appliquée, Université Blaise Pascal- Clermont-Ferrand 2, 2008.

[11] : J. R. Wang and N. Parameswaran, « Survey of Sports video analysis: research issues and applications », Proceedings of the Pan-Sydney area workshop on Visual Information Processing (VIP), Darlinghurst, Australia, 2003.

[12] : H. Huynh, « Vidéosurveillance pour appartements intelligents », Thèse de doctorat en Informatique, Université de la Méditerranée, Faculté des Sciences de Luminy, 2010.

[13] : H.M. Dee, and S.A. Velastin, «How close are we to solving the problem of automated visual surveillance?» A review of real-world surveillance, scientific progress and evaluative mechanisms, Machine Vision and Applications, vol. 19 (5-6), pages 329-343, 2008.

[14] : H. Tbia, M. Gouiffes, L.Lacassagne, « La Reconnaissance des activités humaine à partir des vecteurs de mouvements quantifiés », Institut d'Electronique Fondamentale d'Orsay, France.

[15] : G. Mógio membre du Comité d'éthique (COMETS), « Réflexion sur Ethique et Sciences du comportement humain ». 23 Février 2007.

[16] : K. Dimitropoulos, T. Semertzidis and N. Grammalidis «Video and signal based surveillance for airport applications », Proceedings of the 6 th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pages 170-175, Genova, Italy, 2009.

- [17] : K. Robert, « Night-time traffic surveillance a robust framework for multi-vehicle detection, Classification and Tracking », Proceedings of the 6th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pages 1-6, Genova, Italy, 2009.
- [18] : L. Hamoudi, « Application de techniques d'apprentissage pour la détection et la reconnaissance d'individus », Thèse de doctorat en Automatique, Génie Informatique, Ecole des Mines de Douai, 2011.
- [19] : M. Ferdjani, A. Seriai, A. Seddik, « La vidéosurveillance a distance », Rapport du projet TER, Université Montpellier2 sciences et techniques.
- [20] : M. Griot, « Tableau des différents principaux formats vidéo », avril 2010
- [21] : N. Krahnstoever, P. Tu, T. Yu K. Patwardhan, D. Hamilton, B. Yu, C. Greco and G. Doretto, «Intelligent video for protecting crowdcd sports venues», Proceedings of the 6 th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), pages 116-121, Genova, Italy, 2009.
- [22] : N. T Siebel , S.J Maybank « The ADVISOR Visual Surveillance System ». In Proceedings of the ECCV 2004 Workshop, « Applications of Computer Vision » (ACV'04), Prague, Czech Republic, pages 103-111, ISBN 80-01-02977-8 , May 2004.
- [23] : P. Jorge, A.J. Abrantes, J .S. Marques « Estimation of the Bayesian network architecture for object tracking in video sequences ». In Proceedings of the 17th International Conference on Pattern Recognition ICPR, Vol. 2, pages 732-735, Cambridge, August 2004.
- [24] : P. Wang, G. Abowd, J. Rehg, « Quasi-periodic Event analysis for social game retrieval », Proceedings of the IEEE International Conference on Computer Vision (ICCV), pages 112-119, Kyoto, Japan, 2009.
- [25] : P. Withagen, « Object Detection and Segmentation for Visual surveillance ». PhD thesis, university of Amsterdam. 2006.

[26] : T. Badard, F. Hubert « Découverte d'une API multimédia –JMF », Cours Université Laval, Département des sciences géométriques Québec, Canada.

[27] : W. Zhao, R. Chellapa, P. J. Phillips, A. Rosenfeld, « Face recognition: a literature survey », ACM Computing Surveys, vol. 35 (4), pages 399-458, 2003.

[28] : Z. Li, E. Tan, J. Chen, T. Wassantachat, «On Traffic Density Estimation with a Boosted SVM classifier », Proceedings of the Digital Image Computing: Techniques and Applications (DICTA), pages 117-123, Washington, DC, USA, 2008.

[29] : www.europeschool.net. Janvier 2013

[30] : <http://www.paratronic.fr/feux.html>. Février 2013

[31] : <http://marchal.tripod.com/Sante%20mentale/anormal.html>. Mars 2013

[32] : <http://www.jmdoudoux.fr/java/dejac/indexavecframes.html>. Avril 2013

[33] : http://fr.wikipedia.org/wiki/Java_Development_Kit. Mai 2013

[34] : <http://neuroph.sourceforge.net> Mai 2013