

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/303471309>

Suivi et calcul de la trajectoire d'une personne dans un environnement fermé

Thesis · October 2015

DOI: 10.13140/RG.2.1.1252.6966

CITATIONS

0

READS

5,963

1 author:



[Zakaria El Kasri](#)

Mohammed V University of Rabat

1 PUBLICATION 0 CITATIONS

SEE PROFILE



PFE

POUR L'OBTENTION DU TITRE

MASTER EN INFORMATIQUE & TÉLÉCOMMUNICATIONS

FACULTÉ DES SCIENCES RABAT

UNIVERSITÉ MOHAMMED V

LABORATOIRE DE RECHERCHE EN INFORMATIQUE &
TÉLÉCOMMUNICATIONS

DOMAINE DE RECHERCHE : LA VISION PAR ORDINATEUR,
L'INTELLIGENCE ARTIFICIELLE

PRÉSENTÉ PAR

Zakaria EL KASRI

Suivi et calcul de la trajectoire d'une personne

dans un environnement fermé

Encadrants de PFE : Pr. Rachid OULAD HAJ THAMI

Soutenue le 21 octobre 2015
Devant la Commission d'Examen

JURY

M. Rachid OULAD HAJ THAMI	PES, École Nationale Supérieure d'Informatique et d'Analyse des Systèmes.
M. Lahoucine BALIHI	PA, Faculté des Sciences de Rabat.

REMERCIEMENTS

C'est avec un grand plaisir que je réserve cette page, en signe de gratitude et de reconnaissance à tous ceux qui m'ont aidé à la réalisation de ce travail.

Ce travail a abouti grâce à mon encadrant M. Rachid OULAD HAJ THAMI, Professeur d'enseignement supérieur à l'Ecole Nationale Supérieure d'Informatique et d'Analyse des Systèmes. Je tiens également à le remercier chaleureusement pour ses conseils, pour son soutien sans failles durant cette période de PFE, et la bonne pédagogie pour diriger un travail de recherche jusqu'à la fin. Ses nombreuses qualités, telles que sa motivation inaltérable et son esprit de communication ouvert, m'ont permis de mettre en valeur jour après jour l'ensemble des travaux effectués.

J'aimerais remercier les responsables du master Informatique et Télécommunications à la FSR, en particulier M. Driss Aboutajdine, Directeur du Centre National pour la Recherche Scientifique et Technique (CNRST), Professeur d'enseignement supérieur à la FSR et directeur du Laboratoire de Recherche en Informatique et Télécommunications (LRIT). Je tiens à remercier aussi Mme. S. Mouline, Professeur à la FSR, qui ont approuvé leur dévouement total à faire réussir ce Master.

Je remercie également, les membres du jury qui ont accepté de nous honorer par leur assistance, et quoi qu'il en soit, leur jugement sera accueilli avec enthousiasme.

Enfin, je ne voudrais pas terminer sans exprimer ma gratitude à mes collègues, mes amis, ma famille, et tout particulièrement ma mère, pour m'avoir encouragé moralement et matériellement tout au long de cette expérience.

L'un des sujets majeurs des prochaines années dans la planète entière sera le vieillissement de la population. Une des conséquences de l'augmentation du vieillissement de la population est l'augmentation du nombre de personnes âgées dépendantes, ce qui rend difficile la prise en charge de ces personnes. En effet, le nombre d'instituts spécialisés ne devrait pas permettre de toutes les accueillir. Actuellement, les personnes ont tendance à rester au domicile, soit pour des raisons liées au coût qu'implique un placement, ou soit parce que la personne elle-même préfère rester chez elle. Les personnes âgées vivant seules à domicile sont sujettes à des risques liés au vieillissement à long terme (mauvaise alimentation, hygiène insuffisante, perte d'autonomie...). Pour assurer un minimum de sécurité pour ces personnes on peut installer un système de vidéo-surveillance dans leur habitat pour en faire un Habitat Intelligent pour la Santé (HIS).

Cette installation permet un suivi à distance de l'occupant. Notre but est de suivre les mouvements des personnes surveillées et détecter les situations de danger ou d'anomalie pour lancer une alerte auprès des services compétents et accélérer les secours. Plus le système est apte à effectuer cette tâche, moins l'intervention de l'homme est nécessaire.

Un système de vidéosurveillance est en général composé de trois modules : la détection, le suivi et la reconnaissance. Nous présentons dans cette première étape du travail un état d'art ainsi qu'une étude comparative des différentes techniques et méthodes pour la détection et le suivi de personnes dans une séquence vidéo.

Mots-clés : Analyse de vidéo, suivi d'objet, calcul de trajectoire, analyse de trajectoire.

TABLE DES MATIÈRES

Introduction générale	9
1 La vidéosurveillance intelligente	13
Introduction	13
1.1 Évolution des systèmes de vidéosurveillance intelligents	14
1.2 Domaines d'application de la vidéosurveillance intelligente . .	15
1.3 Structure d'un système de vidéosurveillance intelligente	16
Conclusion	17
2 État de l'art	18
Introduction	18
2.1 Détection d'objets en mouvement	18
2.1.1 Points d'intérêt	19
2.1.2 Soustraction d'arrière-plan	20
2.1.3 Différence temporelle	22
2.1.4 Apprentissage supervisé	23
2.2 Suivi d'objets	23
2.2.1 Représentation des objets	23
2.2.2 Filtre de Kalman	25
2.2.3 Filtre à particules	26
2.2.4 Méthode de recouvrement	27
2.2.5 Mean-Shift	28
2.2.6 CamShift	28
2.2.7 Flot optique	28
Conclusion	29

TABLE DES MATIÈRES

3 Implémentation et comparaison	31
Introduction	31
3.1 Détection de personne	34
3.1.1 Résultats	34
3.1.2 Discussion	39
3.2 Suivi de personne	41
3.2.1 Résultats	41
3.2.2 Discussion	47
Conclusion	48
Conclusion générale et perspectives	49
Bibliographie	52

TABLE DES FIGURES

1	Prototype d'une chaine de traitement pour l'analyse de trajectoires d'individus.	11
1.1	Les trois modules d'un système de videosurveillance intelligente	17
2.1	Exemple de points d'intérêt extraits avec le détecteur de Harris [18]	19
2.2	Résultat de la segmentation par soustraction de fond.(a) Image courante. (b) Image du fond. (c) Segmentation d'objet	21
2.3	Illustration du problème d'ouverture des algorithmes basés sur la différence entre images [20].	22
2.4	Exemples de représentation d'objets [24] . (a)-(b) Représentation à l'aide d'un ensemble de points, (c)-(d)-(e) Représentation à l'aide d'un ensemble de formes géométriques, (f) Représentation à l'aide du squelette, (g)-(h)-(i) Représentation à l'aide du contour et de la silhouette d'un objet.	24
2.5	Le cycle du filtre de Kalman.	25
2.6	Principe d'un filtre à particules [30] où l'on retrouve les trois grandes étapes : sélection, prédiction et mesure.	27
2.7	Le flux optique [33]. (a et b) Région de la bouche sur deux images consécutives. (c) Champ de vecteur de vitesses.	29
3.1	Résultat de la détection des points d'intérêt en utilisant le détecteur de Harris.	34
3.2	Résultat de la détection des points d'intérêt en utilisant le détecteur de Harris. (a) masque utilisé, (b) les points détectés.	35
3.3	Résultat de la segmentation par soustraction de fond.(a) Image du fond. (b) Image courante. (c) Segmentation d'objet.	36

TABLE DES FIGURES

3.4	Résultat de la segmentation par soustraction de fond. (a) Résultat sans lissage. (b) Résultat après lissage.	36
3.5	Résultats de la méthode Différence Temporelle. (a) 100 ^{ème} trame de la sequence. (b) 103 ^{ème} trame de la sequence. (c) Résultat de la différence.	37
3.6	Résultats de la méthode Double Différence Temporelle. (a) 72 ^{ème} trame de la sequence. (b) 75 ^{ème} trame de la sequence. (c) 78 ^{ème} trame de la sequence. (d) la différence entre (a) et (b). (e) la différence entre (c) et (d). (f) Résultat final de la détection de la personne en mouvement.	38
3.7	Résultats du suivi avec le Mean-shift. (a) Définition de la région d'intérêt. (b) 88 ^{ème} trame de la sequence. (c) 145 ^{ème} trame de la sequence. (d) 217 ^{ème} trame de la sequence.	42
3.8	Erreur du suivi en utilisant Mean-Shift. (a) Définition de l'objet d'intérêt en incluant des pixels du fond. (b) 46 ^{ème} trame de la sequence. (c) 71 ^{ème} trame de la sequence. (d) 119 ^{ème} trame de la sequence.	43
3.9	Résultats du suivi avec le CamShift. (a) Définition de la région d'intérêt. (b) 59 ^{ème} trame de la sequence. (c) 177 ^{ème} trame de la sequence. (d) 219 ^{ème} trame de la sequence.	44
3.10	Résultats du suivi avec le filtre de Kalman. (a) 25 ^{ème} trame de la sequence. (b) 74 ^{ème} trame de la sequence. (c) 99 ^{ème} trame de la sequence. (d) 216 ^{ème} trame de la sequence.	46
3.11	Trajectoire mesurée/estimée de la personne en mouvement.	46

LISTE DES TABLEAUX

3.1	Séquences vidéo d'expérimentation.	33
3.2	Résumé des résultats présentés, où le nombre d'étoiles est proportionnelle à l'efficacité de la méthode.	39
3.3	Tableau comparatif, où le nombre d'étoiles est proportionnelle à l'efficacité de la méthode.	47

INTRODUCTION GÉNÉRALE

La vidéosurveillance est un ensemble de caméras disposés dans une zone publique ou privée pour mieux la surveiller. Elle a pour objectif principal, la sécurité et la protection des personnes et des biens. Ceci est réalisable en plaçant des caméras vidéo aux endroits à surveiller et qui sont reliés à un terminal qui permet de visualiser les flots vidéos fournis par toutes les caméras. En outre, une installation de vidéosurveillance permet d'enregistrer les vidéos filmées par les caméras de surveillance, ces vidéos ont la particularité d'être volumineuses, il devient donc impossible pour un surveillant humain d'analyser toutes ces vidéos, afin d'y déceler les comportements ou événements suspects. Or, observer plusieurs écrans pendant de longues heures pousse la concentration à ses limites, après seulement 20 minutes, un opérateur peut manquer jusqu'à 90% de l'activité d'une scène [1]. C'est pourquoi, depuis que la puissance de calcul des ordinateurs le permet, on cherche à automatiser les tâches réalisées jusqu'à présent par des opérateurs humains.

Il s'agit de la vidéosurveillance intelligente, celle-ci repose sur l'utilisation des outils informatiques et technologiques qui ont pour but de détecter automatiquement des événements, distinguer un comportement normal d'un comportement anormal et par suite générer une alerte afin d'alléger la charge de travail de l'opérateur qui surveille plusieurs moniteurs.

La vidéosurveillance est appliquée dans différents domaines. Nous nous intéressons dans ce mémoire à un domaine particulier qu'est la surveillance des personnes fragiles chez elles ou dans les établissements spécialisés, c'est un enjeu social et économique important devant le vieillissement de population, la longévité et les maladies qui l'accompagnent (Alzheimer, Parkinson, etc.)

Nous nous intéressons plus particulièrement aux personnes âgées vivant seules à domicile et atteintes d'Alzheimer. Cette maladie est dite "neuro-dégénérative", elle se caractérise par la dégénérescence progressive et irréversible des cellules nerveuses (les neurones). La disparition de ces cellules entraîne des troubles cognitifs (altération de la mémoire, perturbation du langage, etc.), des troubles du comportement (agressivité verbale, déambulation, etc.) et, dans certains cas, des symptômes psychotiques.

D'autres symptômes peuvent évoquer la maladie d'Alzheimer chez une personne âgée, sans pour autant être caractéristiques de cette maladie : par exemple, la fatigue chronique, la perte d'autonomie, un état dépressif, des chutes répétées, un amaigrissement, etc.

Afin de sécuriser la personne malade d'une part et de rassurer les aidants d'autre part, la vidéosurveillance est une des technologies qui prennent en compte les spécificités des malades d'Alzheimer et détectent automatiquement certaines situations anormales.

Un avantage de cette technologie est qu'elle n'entrave pas la liberté de la personne dans le sens où la personne n'a pas de capteur à porter et qu'il n'y a pas de bouton à actionner.

Stratégie envisagée

La Figure 1 illustre les différentes étapes nécessaires afin d'atteindre l'objectif de notre projet.

Tout d'abord, le flux vidéo est obtenu à l'aide des caméras placées dans chaque pièces de l'appartement. Ensuite, on applique au flux vidéo un ensemble d'algorithmes qui permettent d'en extraire des informations compréhensibles par les humains.

Le premier bloc de traitement consiste en la détection et la segmentation des personnes au premier plan, afin de les isoler du reste de l'image pour des analyses ultérieures. Une fois qu'un individu d'intérêt a été détecté dans la scène, la procédure de suivi est déclenchée pour déterminer sa position de manière continue et fiable tout au long du flux vidéo.

En effet, le processus de suivi d'un individu cible permettra de générer son chemin parcouru appelé « trajectoire ».

L'analyse de trajectoire est un des moyens les plus utilisés dans les systèmes de surveillance pour décrire l'activité d'une personne.

Le système pourra alors générer des alertes en cas de détection des situations de danger.

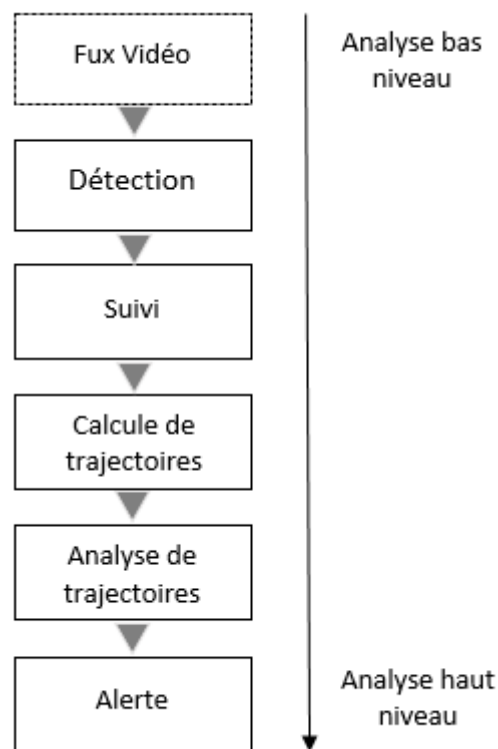


FIGURE 1 – Prototype d'une chaîne de traitement pour l'analyse de trajectoires d'individus.

Dans cette première étape du travail, nous nous intéressons au module de la détection et celui du suivi, les autres étapes de ce prototype sont considérées comme perspectives de ce travail.

Organisation du document

Ce mémoire est organisé de la manière suivante :

- Le chapitre 1 introduit la vidéosurveillance intelligente , nous y citons des application dans divers domaines de la vidéo surveillance, le prototype, etc.

- Le chapitre 2 présente un état de l'art des techniques utilisées en vidéosurveillance pour détecter et suivre d'un objet. Nous citons dans la section 2.1 les approches les plus utilisées pour définir les zones d'intérêt. Ce sont les endroits de la scène où le mouvement sera estimé. Une fois ces zones définies, un algorithme d'estimation du mouvement est utilisé. Nous présentons les algorithmes les plus connus dans la Section 2.2.
- Le chapitre 3 présente les résultats d'une étude comparative des différentes méthodes de détection et de suivi de personne.
- Enfin nous terminons ce mémoire par une conclusion générale.

CHAPITRE 1

LA VIDÉOSURVEILLANCE INTELLIGENTE

Introduction

La vidéosurveillance intelligente est un ensemble de caméras disposées dans un espace à surveiller, ces caméras sont reliées à un système informatique qui permet le traitement et l'analyse des données reçues afin de fournir une interprétation automatique des scènes filmées et de comprendre et prédire les actions et interactions des objets observés.

La vidéosurveillance de base se limite bien souvent à la visualisation d'images (capture), la transmission des données sous forme numérique à un écran de contrôle et à l'enregistrement des images dans une banque de données. Dans ce type de vidéosurveillance, l'humain joue un rôle primordial. Pourtant, plusieurs études ont montré qu'après 20 minutes d'observation seulement, un opérateur peut ne pas voir jusqu'à 90% de l'activité d'une scène. Pour pallier à cela, la vidéosurveillance intelligente combine des logiciels et un matériel d'analyse des images intelligents qui viennent assister l'opérateur humain et le rendent plus réactif, et alors les erreurs d'inattention sont moins fréquentes. Les systèmes de vidéosurveillance intelligente utilisent des algorithmes mathématiques pour détecter des objets en mouvements dans l'image et filtrer les mouvements non pertinents. Ils créent une base de données consignnant les attributs de tous les objets détectés et leurs propriétés de mouvements. La prise de décision par le système ou la recherche d'événements d'intérêt dans des séquences archivées se fait à partir de règles (par ex. : si une personne traverse une limite, envoyer une alerte).

1.1 Évolution des systèmes de vidéosurveillance intelligents

L'évolution technologique des systèmes de vidéosurveillance a débuté durant les années 60 avec les systèmes CCTV (Closed-Circuit TeleVision) analogiques. Les ordinateurs ne permettaient pas encore d'assister les opérateurs à cause des faibles avancées matérielles et logicielles. Puis, sont arrivées les caméras numériques permettant de filmer avec une meilleure résolution d'images et réduction du bruit. Dans les années 80, des ordinateurs encore plus puissants ont vu le jour. Cette évolution technologique permet de classer les systèmes de vidéo-surveillance en trois générations [2].

- **1^{ère} génération (1960-1980)** : ces systèmes sont basés sur des sous-systèmes analogiques pour l'acquisition, la transmission et le traitement des images. Ils étendent la portée visuelle de l'homme en transmettant les sorties de plusieurs caméras surveillant un ensemble de sites à des écrans dans une salle de contrôle centrale. Ces systèmes sont très fiables et simple à utiliser, ils ne requièrent pas de compétences informatiques. Ils ont cependant plusieurs inconvénients et limites tels que la nécessité de disposer d'une large bande passante pour la transmission des données, la difficulté d'archivage et de récupération d'événements en raison du besoin d'une grande quantité de flux vidéo, la qualité de la vidéo enregistrée se détériore avec le temps et la difficulté de détection d'événements en temps réel qui dépend uniquement des opérateurs humains, qui ont une capacité et une durée d'attention limitées. Les générations suivantes viennent résoudre ces inconvénients en tirant parti des progrès matériels et logiciels.
- **2^{ème} génération (1980-2000)** : les systèmes de deuxième génération sont « hybrides » dans le sens où ils utilisent des sous-systèmes analogiques et numériques pour pallier certains inconvénients de leurs prédécesseurs. Ils ont exploité les premières avancées des méthodes de traitement de la vidéo numérique en permettant d'avertir les opérateurs de certains événements. La plupart des travaux durant la deuxième génération a été axée sur la détection d'événements en temps réel.
- **3^{ème} génération (2000-)** : Contrairement aux générations précédentes, la troisième génération permet le traitement et la distribution d'images au niveau du capteur en utilisant des caméras intelligentes capables de numériser et compresser le signal acquis ainsi que l'application d'algorithmes de traitement d'images. Elle gère également le stockage

distribué et la récupération de données vidéo basées sur le contenu. Cette génération profite du faible coût du matériel, de la disponibilité élevée des unités de calcul et de la possibilité de communiquer avec des appareils mobiles et des réseaux à faible débit. L'objectif principal de cette génération est d'émettre une alarme en temps réel pour aider les opérateurs à repérer instantanément des événements.

1.2 Domaines d'application de la vidéosurveillance intelligente

La multiplication des finalités assignées à la vidéosurveillance. Elle n'est plus utilisée uniquement à des fins sécuritaires, mais est aussi mobilisée pour d'autres tâches telles que la régulation du trafic routier, la gestion urbaine, l'assistance aux personnes, l'analyse de comportement de clients (dans les centres commerciaux), etc.

Généralement les domaines d'application de la vidéosurveillance sont variés et peuvent être publics ou privés. On peut citer :

Reconnaissance d'activités :

- Analyse de trajectoire [3].
- L'analyse de prise des repas .
- l'analyse de prise de médicaments .
- Reconnaissance de gestes [4], de visages et d'expressions faciales [5].
- La détection de chutes [6].

Surveillance des lieux publics :

- Surveillance d'aéroports [7], de métros [8], de stationnements [9] ou de magasins [10].
- Surveillance des autoroutes et des chemins de fer pour la détection d'accidents.
- Surveillance des forêts pour la détection d'incendies.
- Contrôle d'accès à des sites privés.

1.3. STRUCTURE D'UN SYSTÈME DE VIDÉOSURVEILLANCE INTELLIGENTE

Extraction d'informations :

- Mesure du débit de la circulation routière et d'encombrement de piétons [11].
- Établissement de profils de consommateurs dans des centres commerciaux.
- Comptage de personnes dans des espaces publics.
- Extraction de statistiques dans des activités sportives [12].

Applications militaires :

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

Applications civiles :

- Mesure de la vitesse des véhicules [13].
- Détection de franchissement de feu rouge et de franchissement de ligne continue [14].

1.3 Structure d'un système de vidéosurveillance intelligente

Un système de vidéosurveillance est en général composé de trois modules comme illustré dans la figure 1.1 :

- **Le module de détection**

A chaque nouvelle image, l'objet en mouvement doit être détecté et segmenté en régions. La qualité de cette détection aura une influence considérable sur les performances finales du système, puisque la non détection d'un acteur engendrera obligatoirement la non détection ou reconnaissance de l'action.

- **Le module de suivi**

Le suivi d'objet consiste à localiser et maintenir l'identité des objets détectés au cours du temps. Cette étape est une estimation temporelle des trajectoires des différents objets mobiles.

1.3. STRUCTURE D'UN SYSTÈME DE VIDÉOSURVEILLANCE INTELLIGENTE

- **Le module de reconnaissance**

Il permet de détecter un événement suspect et génère une prise de décision à savoir si le comportement est normal ou anormal.

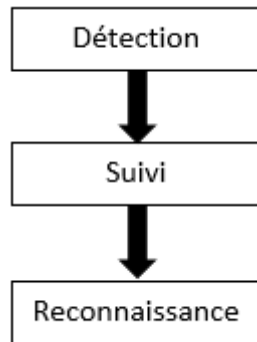


FIGURE 1.1 – Les trois modules d'un système de vidéosurveillance intelligente

Conclusion

Nous avons présenté dans ce chapitre une vue d'ensemble de ce qu'est la vidéosurveillance intelligente. Nous avons également pu découvrir l'étendue des usages de ces systèmes ainsi que les différentes manières dont ils pouvaient être exploités pour l'analyse de scène. Ainsi, nous avons évoqué l'évolution des recherches sur le développement de tels dispositifs. En effet, le passage à l'ère numérique permet notamment une surveillance plus intelligente ainsi qu'un pilotage et une surveillance à distance. Nous avons par ailleurs pu mettre en évidence la structure d'un système de vidéosurveillance intelligente dédié à l'analyse d'activité d'individus dans la scène. Le chapitre suivant dresse un état de l'art des méthodes utilisées dans le domaine de la vidéo surveillance et traite des techniques de détection et de suivi d'objets.

CHAPITRE 2

ÉTAT DE L'ART

Introduction

La détection automatique du mouvement dans les séquences vidéo est un sujet très actif depuis le début des années 1980. Depuis cette date, de nombreux travaux ont été menés et de nombreuses approches ont été mises en œuvre, et en particulier depuis le milieu des années 1990, lorsque la puissance des ordinateurs grand public a permis d'envisager sérieusement un traitement en temps réel des données vidéo, toutefois la plupart d'entre elles ne sont valables que dans certaines conditions spécifiques.

La section 2.1 est consacrée à la détection des objets en mouvement dans une séquence vidéo. la Section 2.2 fournit un aperçu de quelques techniques utilisées pour le suivi d'objets.

2.1 Détection d'objets en mouvement

La détection est la première étape d'un système de vidéosurveillance. Elle permet l'extraction des objets mobiles présents sur les images de la séquence vidéo en les séparant de l'arrière-plan. C'est une étape de traitement de bas niveau, critique et difficile car elle doit être robuste aux changements dynamiques de la scène (changements d'éclairage, rajout d'objets, ombres, etc). La littérature sur le sujet étant extrêmement abondante, nous n'énoncerons pas toutes les méthodes existantes. L'idée est plus de donner une description des méthodes les plus couramment utilisées pour la détection de mouvement en vision par ordinateur.

2.1.1 Points d'intérêt

Une première approche possible consiste à détecter des points d'intérêt, ce sont des points susceptibles d'être intéressants à suivre où apparaît un changement significatif dans l'image. Par exemple, les coins, les intersections, les points isolés et les points spécifiques sur les textures sont des points d'intérêt. Ces points auront de plus l'avantage d'être plus faciles à suivre. Ce type de détection est souvent utilisé dans des problèmes de vision stéréoscopique, cette technique étant en effet très robuste aux changements d'intensité lumineuse et d'angle de vue.

L'utilisation de cette méthode donnera la plupart du temps beaucoup de points d'intérêt mais ces derniers pourront aussi bien appartenir aux personnes qu'au décor (Figure 2.1).

Ces points d'intérêt sont issus d'une sélection des points obtenus par des détecteurs reconnus, à savoir le détecteur de coins de Harris [15], le détecteur SIFT [16] ou le détecteur KLT [17], etc.

Le détecteur de Harris est rapide, ce qui convient aux applications temps réel.



FIGURE 2.1 – Exemple de points d'intérêt extraits avec le détecteur de Harris [18]

2.1. DÉTECTION D'OBJETS EN MOUVEMENT

Harris et al. [15] utilisent la formule (2.1) qui permet de calculer en chaque pixel une valeur appelée valeur d'intérêt :

$$f(x, y) = \text{Det}(M(x, y)) - k \cdot \text{Tr}(M(x, y))^2 \quad (2.1)$$

Où :

- $k \in [0.04, 0.15]$
- $M(x, y)$: matrice de structure, capture la structure du voisinage local des pixels de l'image et s'exprime comme suit :

$$M(x, y) = \begin{pmatrix} (G_\sigma * (\frac{\delta I}{\delta x})^2)(x, y) & (G_\sigma * (\frac{\delta I}{\delta x} \cdot \frac{\delta I}{\delta y}))(x, y) \\ (G_\sigma * (\frac{\delta I}{\delta x} \cdot \frac{\delta I}{\delta y}))(x, y) & (G_\sigma * (\frac{\delta I}{\delta y})^2)(x, y) \end{pmatrix}$$

avec :

- $G_\sigma(x, y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2+y^2}{2\sigma^2}}$, où σ (déviatiion standard = racine carrée de la variance) est un paramètre généralement fixé à $\sigma = 2$.
- $\frac{\delta I}{\delta x}$ et $\frac{\delta I}{\delta y}$: représentent les dérivées partielles de la fonction intensité relativement aux axes des abscisses et des ordonnées, et s'expriment comme suit :

$$\begin{aligned} \frac{\delta I}{\delta x}(x, y) &\approx \frac{I(x+1, y) - I(x-1, y)}{2} \\ \frac{\delta I}{\delta y}(x, y) &\approx \frac{I(x, y+1) - I(x, y-1)}{2} \end{aligned}$$

Les points d'intérêt sont ensuite déterminés via un seuil S fixé par l'utilisateur. Si $f(x, y) > S$ alors le pixel (x, y) correspond à un point d'intérêt, sinon il sera rejeté.

2.1.2 Soustraction d'arrière-plan

Dans sa forme la plus simple, l'arrière-plan consiste en une image B dans laquelle la scène est représentée sans objet. La segmentation consiste à étudier chaque pixel de l'image et à les comparer avec l'image de référence B . Soit I_t l'intensité lumineuse de la t -ième image et (u, v) les coordonnées d'un pixel de cette image, si la différence est supérieure à un seuil τ , le pixel est classé en tant que **foreground**, sinon il est classé en tant que **background**.

2.1. DÉTECTION D'OBJETS EN MOUVEMENT

Le résultat de la segmentation (Figure 2.2) est une image binaire M donnée par :

$$M_t(u, v) = \begin{cases} 1 & \text{si } |I_t(u, v) - B(u, v)| \geq \tau \\ 0 & \text{sinon} \end{cases} \quad (2.2)$$

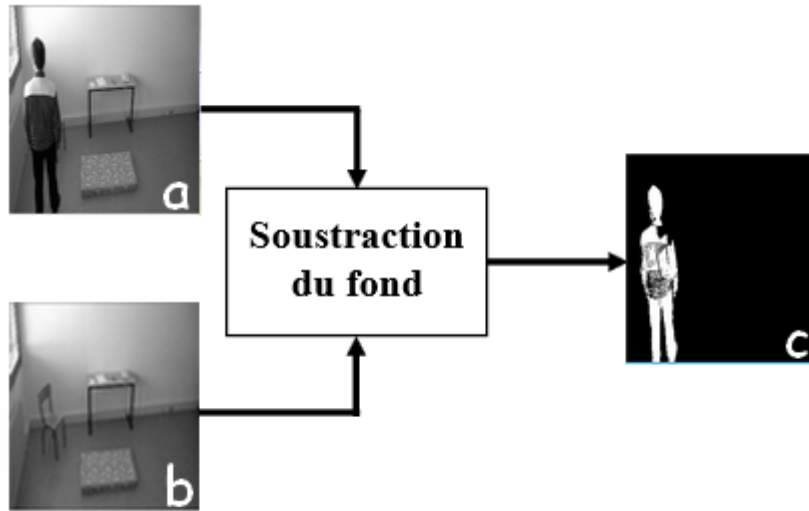


FIGURE 2.2 – Résultat de la segmentation par soustraction de fond. (a) Image courante. (b) Image du fond. (c) Segmentation d'objet

Cette méthode présente l'avantage de détecter aussi bien les mouvements lents que les mouvements rapides. Par ailleurs, même les objets momentanément immobiles sont détectés.

Les méthodes de soustraction de fond sont très efficaces lorsque le fond de la scène est statique et reste invariant ou légèrement changeant. Cependant une image unique de l'arrière-plan est souvent indisponible car il est modifié en continu par divers événements (ex. changements d'éclairage, rajout d'objets, ombres, etc.). Dans ces conditions, il est nécessaire de la mettre à jour régulièrement. Il existe différents algorithmes dans la littérature qui ont été conçus pour modéliser tout ce qui est statique, d'éliminer les ombres, et de récompenser l'évolution de l'arrière-plan. La performance de ces algorithmes, pour estimer un arrière-plan plus robuste, est variable d'un algorithme à l'autre [19].

2.1.3 Différence temporelle

Les approches basées sur la différence entre images (ou différence temporelle) ne nécessitent pas de modèles d'arrière-plan et extraient les régions en mouvement par analyse de la variation temporelle de l'intensité lumineuse des pixels. L'approche la plus simple consiste à observer la différence absolue entre deux images d'entrée. Un seuillage permet ensuite de déterminer les changements dans la scène observée. L'image des zones en mouvement (*foreground*), notée $M(u, v)$ est donnée par :

$$M_t(u, v) = \begin{cases} 1 & \text{si } |I_t(u, v) - I_{t-n}(u, v)| \geq \tau \\ 0 & \text{sinon} \end{cases} \quad (2.3)$$

Avec généralement une valeur faible pour n compris entre 1 et 5.

Cette approche s'adapte très rapidement aux changements de luminosité, mais est peu robuste face aux bruits d'acquisition. Cependant, lorsque les objets sont uniformes en intensité et/ou qu'ils se déplacent rapidement, cette méthode laisse apparaître un effet *fantôme* (surface anciennement recouverte par l'objet), et une zone d'ouverture au centre de l'objet comme illustré sur la Figure 2.3 :

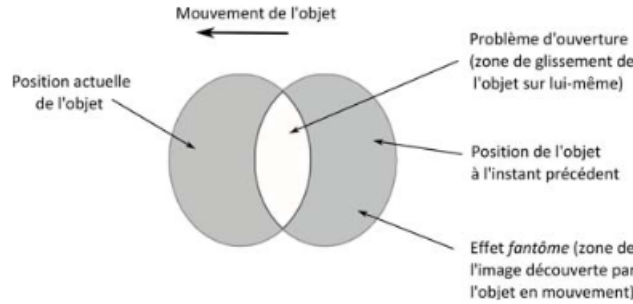


FIGURE 2.3 – Illustration du problème d'ouverture des algorithmes basés sur la différence entre images [20].

Pour pallier à ces problèmes, certains suggèrent l'utilisation d'une double différence à partir de trois images consécutives [21], en calculant dans un premier temps deux différences entre images et en effectuant un seuillage afin d'extraire deux masques binaires. Une opération *ET* logique permet ensuite d'obtenir l'image binaire de la double différence.

2.1.4 Apprentissage supervisé

L'idée est d'utiliser un modèle de l'objet à détecter et chercher à trouver les éléments se rapprochant de ce modèle. Ces méthodes comprennent la plupart du temps une phase d'apprentissage où on apprend par exemple à reconnaître certaines parties du corps humain tels les visages ou alors même tout le corps humain. On peut pour cela utiliser des modèles uniquement géométriques en essayant de repérer une silhouette humaine, ou bien utiliser des modèles basés sur les couleurs, les personnes ayant des couleurs assez différentes du fond.

Cette méthode requiert une grande collection d'exemples pour chaque classe d'objet. En outre, cette collection doit être étiquetée manuellement. Par conséquent, ces méthodes, assez performantes, sont coûteuses en calcul et assez beaucoup plus laborieuses à mettre en œuvre que les méthodes de soustraction de fond. Ainsi, il est difficile d'établir un modèle définitif d'une personne (grandes variations d'apparence).

Pour classifier les objets, différentes méthodes d'apprentissage existent comme les *réseaux de neurones* [22] et les machines à vecteur de support (*SVM*) [23].

2.2 Suivi d'objets

Le suivi des objets est une étape importante dans un système de vidéosurveillance. Il consiste à déterminer la position d'une cible de manière continue et fiable tout au long du flux vidéo afin de générer sa trajectoire qui représente un des moyens les plus utilisés pour décrire l'activité d'une personne. Les techniques de suivi sont très diverses, on peut suivre une partie du corps (tête, visage, jambes, etc) ou la totalité de la personne. Parfois, il est nécessaire de faire un suivi plus fin en cherchant à modéliser la personne par une représentation de type squelette, contour ou autre.

Parmi les algorithmes de suivi, les plus populaires sont le filtre de Kalman, le filtre à particules, le Mean-Shift et le flot optique. Ci-après, nous présentons chacun de ces algorithmes en citant tout d'abord les différentes façons pour représenter les objets d'intérêt dans la scène.

2.2.1 Représentation des objets

La représentation des objets fait partie intégrante du processus de suivi. Les objets peuvent être représentés de nombreuses façons (Figure 2.4) et le choix de la représentation d'un objet dépend fortement du domaine d'application.

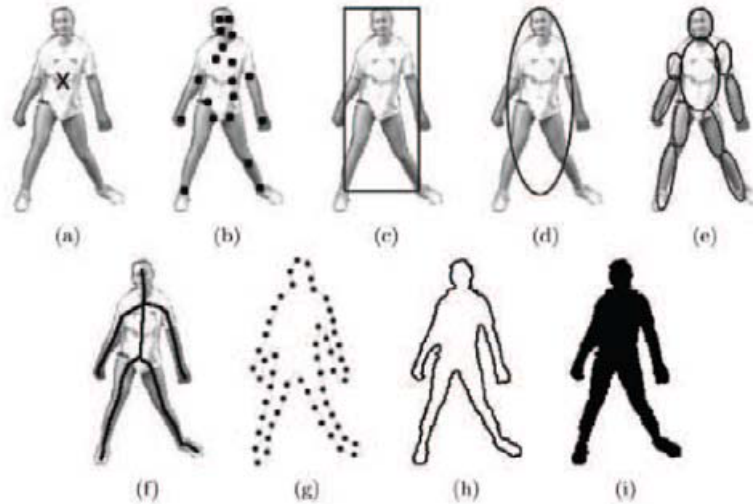


FIGURE 2.4 – Exemples de représentation d’objets [24] . (a)-(b) Représentation à l’aide d’un ensemble de points, (c)-(d)-(e) Représentation à l’aide d’un ensemble de formes géométriques, (f) Représentation à l’aide du squelette, (g)-(h)-(i) Représentation à l’aide du contour et de la silhouette d’un objet.

- **Points** - Un objet peut être représenté par un point (centre de l’objet [25]) ou par un ensemble de points [26]. Cette représentation est adaptée aux objets occupant une petite partie de l’image.
- **Formes géométriques** - L’objet est représenté par une forme géométrique, par exemple un rectangle ou une ellipse, permettant une description de la dimension de l’objet. Cette représentation est adaptée au suivi d’objets rigides (véhicules) mais peut également convenir pour des objets non rigides.
- **Contour et silhouette d’objet** - La représentation d’un objet par son contour permet une description plus complète de la forme d’un objet. La région interne du contour est appelée silhouette de l’objet et peut être utilisée conjointement à l’information de contour pour le suivi d’objets. Cette représentation est adaptée au suivi d’objets non rigides ayant une forme complexe [27].
- **Squelette** - On peut extraire le squelette d’un objet pour caractériser sa forme, ce modèle est utilisé en tant que descripteur de forme pour la

reconnaissance d'objets. Cette représentation est très adaptée au suivi d'objets déformables ainsi que des objets rigides.

2.2.2 Filtre de Kalman

Le filtre de Kalman est un outil récursif linéaire pour estimer l'état d'un système. Appliqué au problème du suivi de personnes, le filtre de Kalman sert à estimer la position d'une personne à l'instant t à partir de celle observée à l'instant $t - 1$. Pour chaque nouvelle position à estimer, le filtre de Kalman se déroule en 2 étapes de façon cyclique :

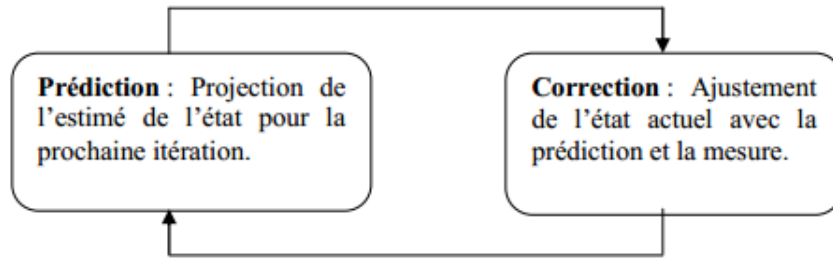


FIGURE 2.5 – Le cycle du filtre de Kalman.

La phase de prédiction consiste à estimer un état X_t à un instant courant t sachant l'estimation X_{t-1} de l'état précédent. La phase de correction permet ensuite de corriger l'état prédit à l'aide de la mesure actuelle.

Le modèle de Kalman est basé sur l'hypothèse que l'état réel X au temps t dépend de l'état précédent X_{t-1} suivant l'équation :

$$X_t = A.X_{t-1} + w_t \quad (2.4)$$

où :

- X_{t-1} : vecteur d'état contenant les paramètres à estimer,
- X_t : vecteur d'état prédit,
- A : matrice de transition ou d'évolution du système en absence du bruit,
- w_{t-1} : représente le bruit du processus supposé blanc de moyenne nulle, de covariance Q .

À l'instant t , une observation est réalisée, selon l'équation :

$$Z_t = H.X_t + v_t \quad (2.5)$$

2.2. SUIVI D'OBJETS

où :

- Z_t : vecteur d'état contenant les paramètres à estimer,
- H : matrice de mesure, est un modèle qui permet de passer de l'état réel à l'espace d'observation,
- v_t : bruit de mesure blanc de moyenne nulle indépendant, de covariance R .

Les équations de la prédiction :

La prédiction sur le vecteur d'état $X_{t|t-1}$ et la matrice de covariance $P_{t|t-1}$ sont données par :

$$\begin{cases} X_{t|t-1} = AX_{t-1|t-1} \\ P_{t|t-1} = AP_{t-1|t-1}A^T + Q \end{cases}$$

Les équations de la correction :

L'étape de correction consiste à mettre à jour la prédiction en prenant en compte l'observation Z_t à l'instant t , ses équations sont :

- Calcul du gain de Kalman K_t :

$$K_t = P_{t|t-1}H^T(HP_{t|t-1}H^T + R)^{-1} \quad (2.6)$$

- Correction de l'état :

$$X_{t|t} = X_{t|t-1} + K_t(Z_t - HX_{t|t-1}) \quad (2.7)$$

- Correction de la covariance :

$$P_{t|t} = (1 - K_tH)P_{t|t-1} \quad (2.8)$$

Le terme $(Z_t - HX_{t|t-1})$ dans l'équation (2.7) est l'innovation de l'observation, le terme $(HP_{t|t-1}H^T + R)$ dans l'équation (2.6) est la covariance de l'innovation.

Le filtre de Kalman a été utilisé avec succès pour suivre des personnes [28] ou des véhicules [29]. Néanmoins, ce filtre n'est pas aussi efficace lorsqu'il s'agit de trajectoires aléatoires de personnes dans le cas d'occlusions ou de changement brusque de direction, c'est à dire, lorsqu'il s'agit d'un système dont le modèle d'évolution est non linéaire.

2.2.3 Filtre à particules

Isard et al. [30] utilisent le filtre à particules pour le suivi de trajectoires avec des changements brusques et où les cibles peuvent être multiples. Le

2.2. SUIVI D'OBJETS

Le filtre à particules permet de lever la contrainte de linéarité et du modèle gaussien qu'impose le filtre de Kalman. Dans la méthode de filtre de particules, les états possibles de l'objet à l'instant t sont représentés par N particules $\{ s_t^{(n)} : n = 1, \dots, N \}$ générées aléatoirement. Une particule est définie par les valeurs de son état courant (p. ex. la position (x, y)) et le poids $\pi_t^{(n)}$ indiquant son importance.

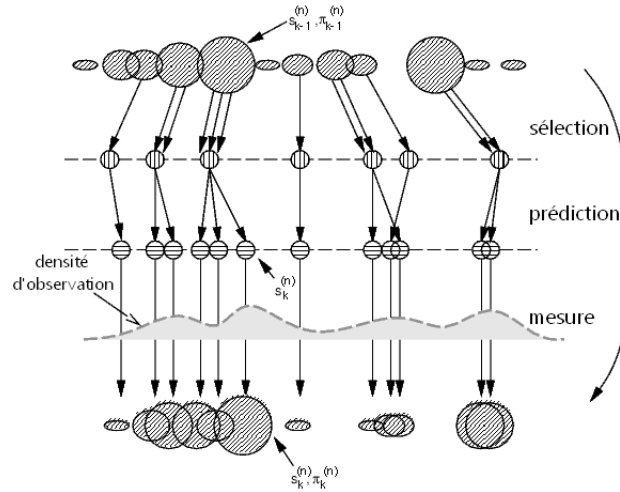


FIGURE 2.6 – Principe d'un filtre à particules [30] où l'on retrouve les trois grandes étapes : sélection, prédiction et mesure.

L'algorithme du filtre à particules se déroule en trois étapes :

- **Sélection** : Un nouvel ensemble de particules est construit aléatoirement en favorisant les meilleures particules de l'ensemble précédent.
- **Prédiction** : Chaque échantillon est propagé à l'aide d'un modèle dynamique stochastique.
- **Mesure** : Un poids est calculé pour chaque échantillon du nouvel ensemble.

Une fois que tous les échantillons ont été construits, le résultat du suivi est une estimation pondérée de l'ensemble des particules.

2.2.4 Méthode de recouvrement

Un des suivis les plus simples est le suivi par recouvrement. Cette méthode ne nécessite pas de prédiction de position, les objets sont appariés par

recouvrement de leur boîte englobante entre l'image $t - 1$ et l'image t .

Fuentes et al. [31] supposent une caméra fixe et une vitesse de déplacement des objets relativement faible par rapport à la fréquence de capture des caméras et que la scène contient peu de personnes en mouvement donc moins de problèmes d'occultation. Ils proposent d'abord de soustraire l'arrière plan en se basant sur le contraste de luminance de l'objet et celle de l'arrière-plan. La trajectoire de chaque blob est déterminée par l'intersection des boîtes englobantes sur deux images consécutives en tenant compte des séparations/fusions des blobs à travers deux images successives.

2.2.5 Mean-Shift

La procédure Mean-Shift est une autre approche largement utilisée par la communauté de vision par ordinateur pour le suivi d'objets non-rigides [32]. D'une manière générale, la procédure Mean-Shift utilise la distribution globale des couleurs ou des niveaux de gris à l'intérieur de la forme géométrique pour représenter l'apparence d'un objet. L'algorithme maximise la similarité en apparence itérativement en comparant l'histogramme de référence Q de l'objet à l'histogramme P calculé dans une fenêtre autour d'une position hypothétique de l'objet. Le vecteur de déplacement (appelé vecteur Mean-Shift) est calculé à chaque itération de manière à améliorer cette similarité. Ce processus est répété généralement de quatre à six fois avant d'atteindre la convergence définie par un déplacement minimum.

2.2.6 CamShift

L'algorithme du CamShift (Continuously Adaptive Mean Shift) encapsule celui du Mean-Shift dans une boucle faisant fluctuer la taille de la fenêtre jusqu'à convergence. A chaque itération, le mean shift est appliqué à l'image, avec une taille de fenêtre fixe jusqu'à convergence, puis la taille de la fenêtre s'oriente et s'agrandit en fonction de la distribution des couleurs.

2.2.7 Flot optique

Une autre approche possible consiste à utiliser ce qu'on appelle les flots optiques. Les algorithmes de flot optique ont pour but d'estimer le champ vectoriel, qui décrit les mouvements géographiques de chaque point image au fil du temps, et fournit des informations importantes pour l'analyse de mouvement [33]. Autrement dit, associé à chaque point physique élémentaire de l'image, un vecteur vitesse 2D, comme étant la projection sur le plan, de la vitesse dans l'espace 3D de ce point. Le flot optique est particulièrement

utile lorsque la caméra est en mouvement, mais ce n'est pas le cas dans notre projet. Le calcul du flux optique est souvent complexe en temps de calcul, ce qui fait qu'il est peu utilisé dans les applications temps réel. De plus, le flot est sensible au bruit dans l'image.



FIGURE 2.7 – Le flux optique [33]. (a et b) Région de la bouche sur deux images consécutives. (c) Champ de vecteur de vitesses.

Conclusion

Nous avons vu dans ce chapitre les méthodes et techniques couramment utilisées pour les deux premières étapes d'un système de vidéo-surveillance. La Section 2.1 a été consacrée à la détection des objets dans une séquence vidéo à l'aide de méthodes telles que la différence temporelle, la soustraction d'arrière-plan ou la détection des points d'intérêt, ou encore des méthodes d'apprentissage supervisé qui utilisent un modèle de l'objet à détecter et cherchent à trouver les éléments se rapprochant de ce modèle. Appliquées dans un contexte de vidéo-surveillance, ces méthodes sont généralement employées pour détecter les régions dans l'image susceptibles de contenir des objets d'intérêt.

Quant à la section 2.2, elle a été consacrée aux méthodes les plus utilisées pour le suivi d'objets. Le processus de suivi consiste à conserver l'identité des objets détectés et de suivre leurs évolutions au cours du temps. Le suivi d'objets est par conséquent dépendant de la représentation des objets. Parmi les approches existantes, on retrouve les méthodes basées sur un modèle prédictif (filtre de Kalman ou à particules), d'autres basées sur l'apparence des objets (Mean-Shift, CamShift), d'autres basées sur les caractéristiques de la région objet (méthode de recouvrement) et d'autres utilisent le champ des vecteurs vitesses pour suivre les points d'intérêt dans des séquences d'images

2.2. SUIVI D'OBJETS

(flot optique). Dans le chapitre suivant nous allons présenter les résultats de l'implémentation de quelques méthodes discutées dans l'état de l'art ainsi qu'une comparaison entre ces approches.

CHAPITRE 3

IMPLÉMENTATION ET COMPARAISON

Introduction

Ce chapitre consacré aux résultats expérimentaux se divise en deux parties. Tout d'abord, la détection d'objets dans une séquence vidéo et plus particulièrement une seule personne, puisque notre projet consiste à surveiller des personnes vivant seules dans un appartement, nous donnerons une étude comparative qualitative entre les différentes méthodes implémentées. Ensuite, nous discutons les résultats obtenus en testant une implémentation des méthodes proposées pour le suivi de personnes.

L'implémentation est faite en exploitant la librairie *OpenCV* (Open Computer Vision) en C++ vue la rapidité d'exécution. Pour réaliser nos tests, nous avons utilisé des vidéos issues du "*Fall detection Dataset*" [34]. Cette base de données contient 191 vidéos dans un cadre réaliste de la vidéosurveillance utilisant une seule caméra. Le taux de trame est de 25 images/s. Ces séquences vidéo illustrent les principales difficultés que nous pouvons trouver dans un environnement domiciliaire des personnes âgées, elles contiennent un éclairage variable et des difficultés typiques comme des occlusions ou fond encombré et texturé. Les acteurs effectuent diverses activités quotidiennes normales et des chutes.

Environnement du travail

Afin de mener à bien ce projet, nous avons utilisé un ensemble de matériels dont les principales caractéristiques sont les suivantes :

- Processeur : Intel(R) Core(TM) i3-2370M CPU @ 2.41GHz
- RAM : 4.00 GO
- OS : Windows 8.1 Professionnel (32 bits)

OpenCV

OpenCV (Open Source Computer vision) est une bibliothèque graphique libre, initialement développée par Intel et maintenant soutenu par la société de robotique *Willow Garage*, spécialisée dans le traitement d'images en temps réel. Toutes ses fonctionnalités ont été programmées en langage C/C++ puis exportées vers une multitude de langages comme Python, JAVA et Matlab. La bibliothèque OpenCV met à disposition de nombreuses fonctionnalités très diversifiées, elle propose la plupart des opérations classiques en traitement d'images :

- lecture, écriture et affichage d'image ;
- calcul de l'histogramme des niveaux de gris ou d'histogramme de couleur ;
- issage, filtrage ; *etc..*

Cette bibliothèque s'est imposée comme un standard dans le domaine de la recherche parce qu'elle propose un nombre important d'outils en traitement vidéo :

- lecture, écriture et affichage d'une vidéo depuis un fichier ou une caméra ;
- détection de visage, détection de mouvements, poursuite d'objets ;
- triangulation de Delaunay ; *etc..*

Certains algorithmes classiques dans le domaine de l'apprentissage artificiel sont aussi disponibles :

- K-means ;
- Réseau de neurones artificiels ;

-
- Séparateur à vaste marge ; *etc.*.

Elle met également à disposition quelques fonctions d’interfaces graphiques, comme les curseurs à glissière, les contrôles associés aux événements souris, ou bien l’incrustation de texte dans une image.

Séquences de test

Afin d’évaluer les performances des différentes techniques de détection et de suivi, de nombreux tests ont été réalisés sur plusieurs séquences vidéo de différentes tailles et de différentes résolutions. Le Tableau 3, illustre les caractéristiques des six séquences utilisées.

Vidéo	Taille (Octets)	Résolution (Pixels)	Nombre d’images	Images par seconde
Office (12)	70786188	320*240	298	25
Office (14)	41098188	320*240	173	25
Office (2)	61523532	320*240	259	25
Home_01 (2)	55538012	320*240	240	24
Home_01 (15)	41366480	320*180	238	24
Home_01 (29)	47100782	320*180	271	24

TABLE 3.1 – Séquences vidéo d’expérimentation.

3.1 Détection de personne

Comme il a déjà été mentionné, la première étape d'un système de vidéosurveillance est la détection des objets mobiles présents sur les images de la séquence vidéo en les séparant de l'arrière-plan pour des analyses ultérieures. Nous avons implémenté trois méthodes parmi les quatre décrites dans la section 2.1. La première méthode faisait la détection des points d'intérêt en utilisant le détecteur de coins Harris. La deuxième méthode utilisait la soustraction d'arrière plan en utilisant, dans un premier temps, les images en niveau de gris et dans un second temps les images en couleur (RGB). La troisième méthode utilise la différence temporelle, premièrement entre deux images successives, deuxièmement entre trois images successives (double différence temporelle). Dans cette section, nous tentons de comparer ces méthodes afin de déterminer si une des trois méthodes est préférable.

3.1.1 Résultats

- **La première méthode** consiste à détecter des points d'intérêt, ce sont des points susceptibles d'être intéressants à suivre comme des zones avec des textures intéressantes. Pour cela nous avons utilisé le détecteur de coins Harris.

La figure 3.1 montre un exemple des points d'intérêt de Harris détectés sur image représentant une personne en train de marcher. Ce détecteur nous permet ainsi d'avoir des pixels qui seront suivis plus facilement dans les images suivantes. Le suivi est effectué grâce aux algorithmes de flot optique que nous avons abordé dans la sous-section 2.2.7.

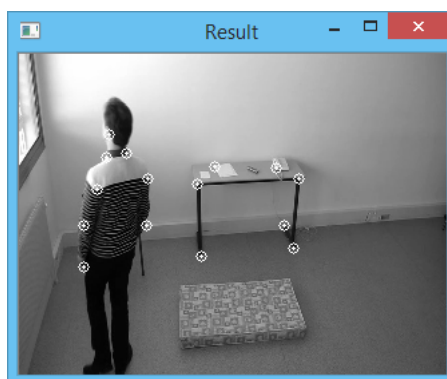


FIGURE 3.1 – Résultat de la détection des points d'intérêt en utilisant le détecteur de Harris.

3.1. DÉTECTION DE PERSONNE

Pour faciliter le suivi et réduire l'espace de recherche, on peut améliorer l'algorithme pour ne détecter que les points qui appartiennent à l'objet en ajoutant un masque de l'objet qui représente une matrice spécifiant les pixels de l'image à traiter représentés par les éléments non-nuls de la matrice. Le masque utilisé est obtenu par la soustraction d'arrière-plan comme illustré dans la figure suivante :

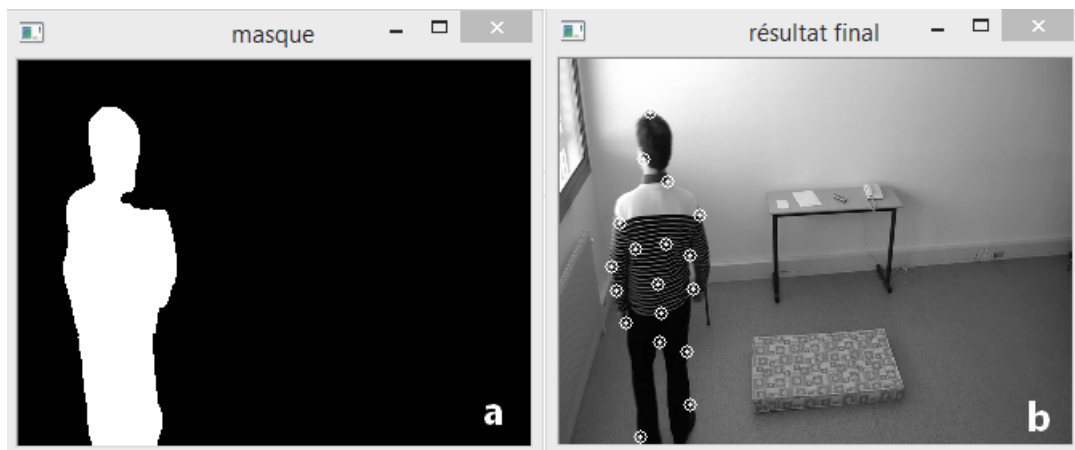


FIGURE 3.2 – Résultat de la détection des points d'intérêt en utilisant le détecteur de Harris. (a) masque utilisé, (b) les points détectés.

- **La deuxième méthode** utilise une image de fond. Cette méthode est très utilisée et elle est adoptée par plusieurs applications. Cette façon détecte la région de mouvement en soustrayant pixel par pixel l'image courante à l'image de fond. Cette technique se limite aux caméras en position fixe. Il existe trois manières pour créer l'image de fond :

- Choisir l'image de fond qui a peu de changement ou qu'il n'y a pas de mouvement pendant une longue durée.
- Choisir la première image dans le flux vidéo.
- Utiliser un modèle de fond adaptatif qui est continuellement mis à jour à l'aide des nouvelles images de la séquence traitée.

Dans notre cas, nous surveillons une personne dans son appartement, alors la caméra est fixe et le fond de la scène est statique et reste invariant ou légèrement changeant, pour cela nous choisissons la première image dans le flux vidéo qui ne contient pas d'objets mobiles comme image de fond. Nous utilisons dans un premier temps les image au format couleur *RGB*, le

3.1. DÉTECTION DE PERSONNE

résultat de la soustraction est illustré dans la figure suivante :

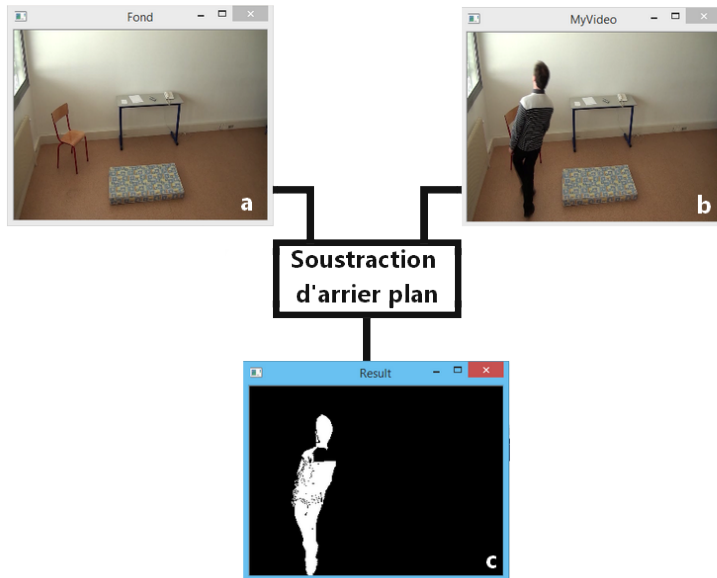


FIGURE 3.3 – Résultat de la segmentation par soustraction de fond. (a) Image du fond. (b) Image courante. (c) Segmentation d’objet.

Souvent l’image obtenue après une opération de soustraction de fond est bruitée (Figure 3.4 (a)). Pour diminuer l’effet de ce bruit, nous avons appliqué un lissage avec noyau de taille 12x12 sur l’image où nous avons constaté que les résultats obtenus avec cette dimension sont suffisants.

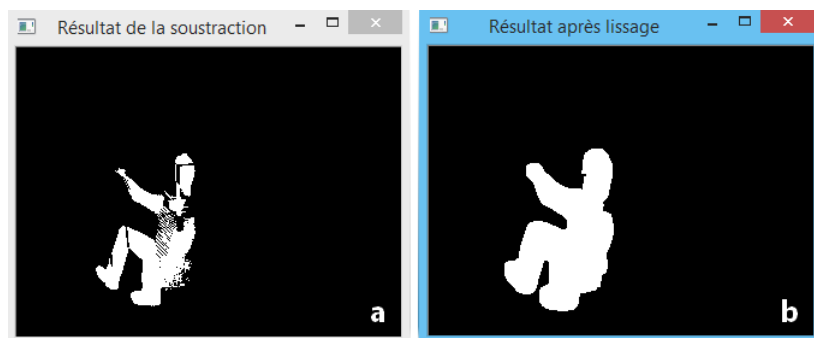


FIGURE 3.4 – Résultat de la segmentation par soustraction de fond. (a) Résultat sans lissage. (b) Résultat après lissage.

3.1. DÉTECTION DE PERSONNE

Cette technique nécessite également le calcul de la différence absolue entre les pixels de l'image de fond et ceux de la nouvelle image pour chaque canal (R,G et B) puis comparer cette valeur avec un seuil ce qui allonge le temps de calcul, pour réduire ce dernier il suffit d'utiliser les images en niveaux de gris.

- **La troisième méthode** ne nécessite pas une image d'arrière-plan et détecte la région de mouvement grâce à la différence de pixel par pixel de deux trames dans un flux vidéo. Le résultat obtenu est illustré par la figure 3.5.

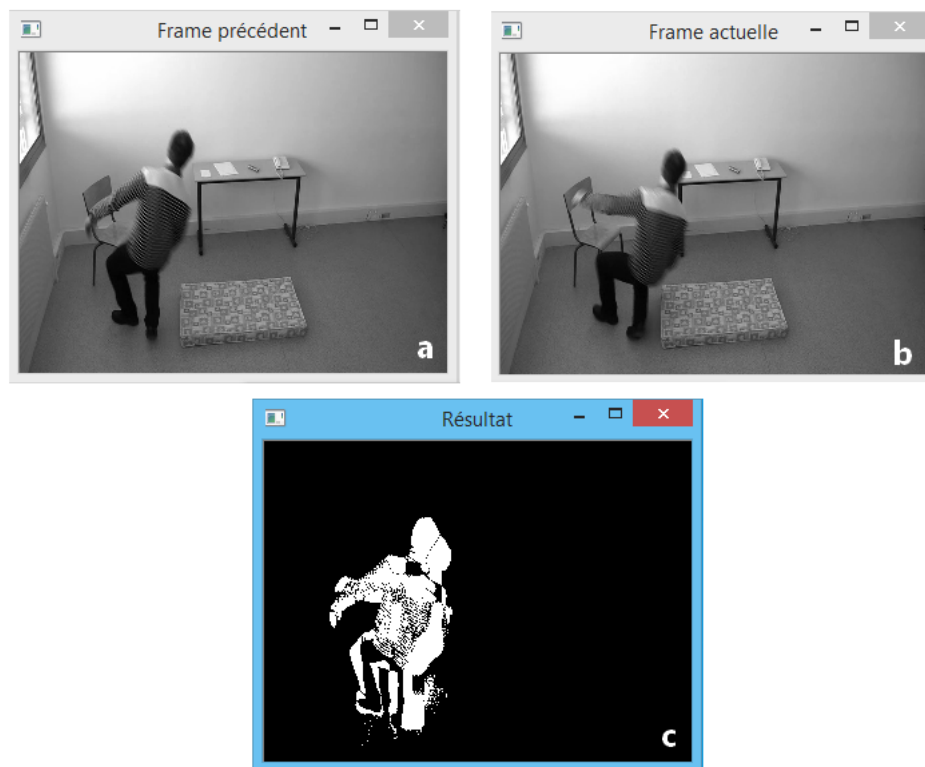


FIGURE 3.5 – Résultats de la méthode Différence Temporelle. (a) 100^{ème} trame de la sequence. (b) 103^{ème} trame de la sequence. (c) Résultat de la différence.

On peut constater que le résultat de la détection n'est pas satisfaisant. Ceci est dû au fait que la différence est calculée entre deux trames distantes (75^{ème} et 78^{ème}) et ce pour simuler un mouvement important. On peut donc conclure que cette méthode n'est pas efficace pour les séquences de mouvement important.

3.1. DÉTECTION DE PERSONNE

Pour améliorer la détection, nous avons utilisé la méthode double différence temporelle. Comme c'est déjà expliqué précédemment, trois trames consécutives sont utilisées dans cette méthode. Un *AND* logique est opéré entre les différences des trames $(t-1, t)$ et $(t, t+1)$ comme illustré sur la figure 3.6.

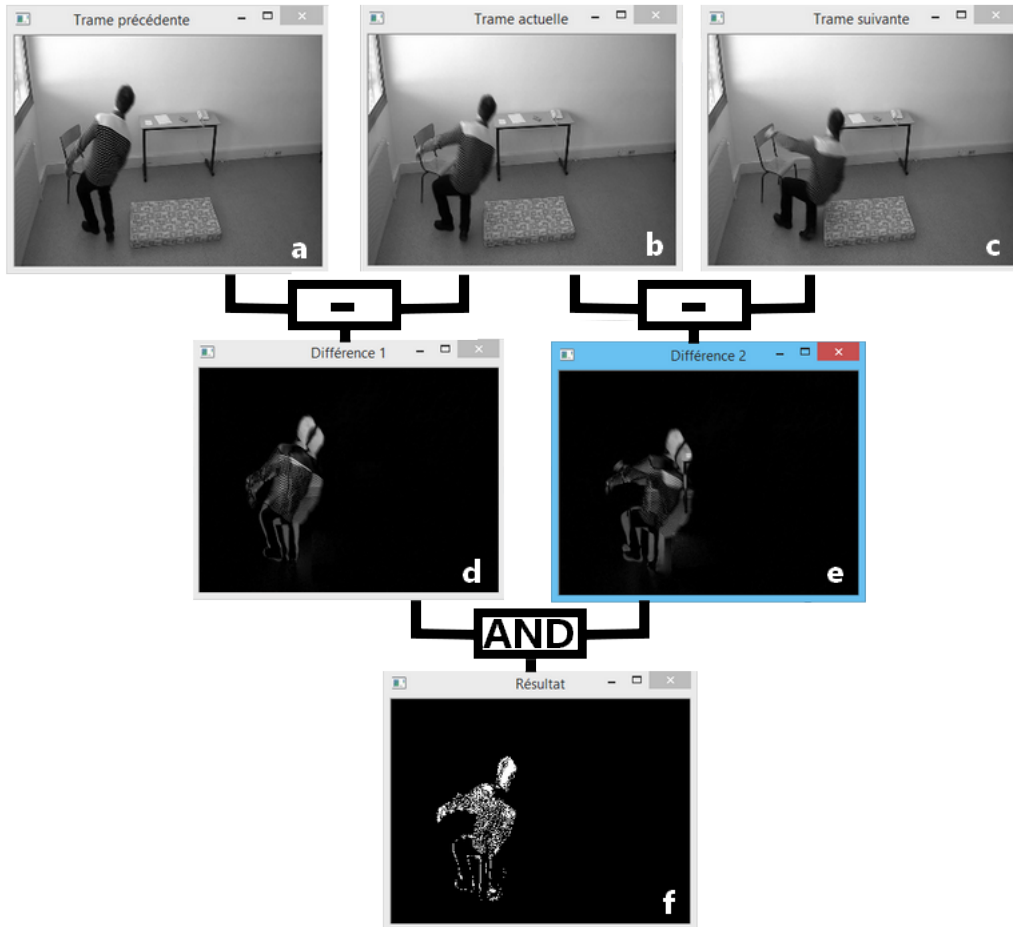


FIGURE 3.6 – Résultats de la méthode Double Différence Temporelle. (a) 72^{ème} trame de la sequence. (b) 75^{ème} trame de la sequence. (c) 78^{ème} trame de la sequence. (d) la différence entre (a) et (b). (e) la différence entre (c) et (d). (f) Résultat final de la détection de la personne en mouvement.

On peut constater que le résultat est nettement meilleur que celui de la différence temporelle entre deux images.

3.1. DÉTECTION DE PERSONNE

3.1.2 Discussion

D'après les résultats rapportés dans la sous-section précédente, un résumé global a été mis en place et présenté dans le tableau 3.2. Ce tableau souligne les principaux conclusions de cette comparaison entre les méthodes de détection de mouvement.

	Temps de calcul	Détection d'objets immobiles	adaptation aux changements de la scène	Détection complète des objets
Points d'intérêt	**	**	***	**
Soustraction d'arrière-plan (RGB)	*	***	*	***
Soustraction d'arrière-plan (niveau de gris)	**	***	*	***
Différence temporelle (niveau de gris)	***	*	***	*
Double différence temporelle (niveau de gris)	***	*	***	**

TABLE 3.2 – Résumé des résultats présentés, où le nombre d'étoiles est proportionnelle à l'efficacité de la méthode.

La détection des points d'intérêt présente plusieurs avantages : elle est très robuste aux changements d'intensité lumineuse et d'angle de vue, elle n'utilise

3.1. DÉTECTION DE PERSONNE

lise aucun modèle global des objets à détecter ce qui rend l'apprentissage pratiquement inexistant, enfin, elle est simple à mettre en œuvre et fonctionne très bien sur des objets texturés. Cependant, lorsqu'on visualise temporellement les points d'intérêt, il apparaît en effet un fourmillement lié à la stabilité imparfaite de la détection. Des points apparaissent ou disparaissent, et cela nous donne l'impression visuelle de mouvements qui ne correspondent pas à des mouvements réels. En outre, le temps de calcul en utilisant cette méthode dépend du nombre de points d'intérêt à détecter. Généralement, utiliser des points d'intérêt au lieu de tous les points dans l'image réduit le temps de calcul pour le suivi d'objets par rapport aux méthodes utilisant toute l'image.

La soustraction d'arrière-plan détecte les mouvements lents ainsi que les mouvements rapides, même les objets immobiles sont détectés. Elle Permet d'extraire la silhouette d'une personne facilitant grandement sa détection et son analyse. Ces méthodes de soustraction de fond présentent l'intérêt d'être relativement simples et efficaces dans des environnements en intérieur pour lesquels le fond présente des caractéristiques suffisamment stables. En revanche, cette méthode est très sensible au changement dynamique de la scène (changement de la lumière, ajout d'objets, ombres, etc.). Aussi, pour une meilleure tolérance aux variations légères du fond, des méthodes de soustraction de fond avec modèle adaptatif ont été proposées. Dans ces méthodes, le modèle de fond est continuellement mis à jour à l'aide des nouvelles images de la séquence traitée. Elles permettent d'adapter le modèle de fond en intégrant de nouveaux éléments stationnaires apparus dans la scène et qui demeurent statiques durant une période suffisamment longue. Nous avons également constaté que l'utilisation du simple espace RGB donnait, dans nos conditions opératoires, des résultats de segmentation satisfaisants. De plus, la transformation des images de l'espace RGB (qui est le format de nos images) en niveaux de gris réduit le temps de calcul.

Les résultats montrent que la simple différence est très avantageuse du point de vue temps de calcul, toutefois lorsque l'objet se déplace lentement et qu'il possède une couleur uniforme, la soustraction laisse apparaître un effet fantôme (surface anciennement recouverte par l'objet), et une zone d'ouverture au centre de l'objet. Ce phénomène s'aggrave lorsque, les objets sont rapides et/ou la fréquence d'acquisition des images est faible. Notamment, cette méthode ne détecte pas les objets immobiles. De surcroît, un mauvais réglage du seuil peut conduire à un taux élevé de bruit et par conséquent, à une fausse détection ou à une absence de détection bien que le mouvement existe. L'application de cette méthode révèle donc plusieurs limites.

La double différence temporelle est également avantageuse en temps de calcul, et donne de bons résultats pour des mouvements d'objets rapides suffisamment texturés. Toutefois, en présence de mouvements lents ou en absence de mouvements pendant une certaine période, cette méthode pourrait souffrir d'imprécision et conduire dans certains cas au même problème de découpage de l'objet, apparu avec la méthode simple différence.

3.2 Suivi de personne

Dans cette deuxième partie, nous allons présenter les résultats obtenus par quelques méthodes de suivi d'objets, cette étape peut être définie comme étant la localisation spatiotemporelle d'un objet en mouvement au cours d'une séquence vidéo afin de générer sa trajectoire. nous commençons par les résultats de la méthode mean-shift. Une comparaison entre ces résultats termine cette section.

3.2.1 Résultats

- **Le Mean-Shift** : Comme il est déjà mentionné, cette méthode localise l'objet en maximisant itérativement la similarité d'apparence en comparant les histogrammes de l'objet et une fenêtre candidate autour d'une position hypothétique de l'objet. Cela est fait en 3 étapes :

1- Dessiner un rectangle (boîte englobante) sur une région de l'objet d'intérêt (Figure 3.7 (a)) puis calculer l'histogramme de couleur de cette région dans l'espace HSV.

Pour chaque nouvelle trame :

2- Calculer la rétro-projection de l'histogramme précédent pour détecter les régions ayant la même couleur que la région définie précédemment. Cette rétro-projection est construite en associant à chaque pixel la valeur de l'histogramme de l'image précédente.

3- Mettre-à-jour la position de la boîte englobant l'objet (au dessus de la région la plus proche de la position précédente ayant une densité maximale) en utilisant Mean-Shift.

3.2. SUIVI DE PERSONNE

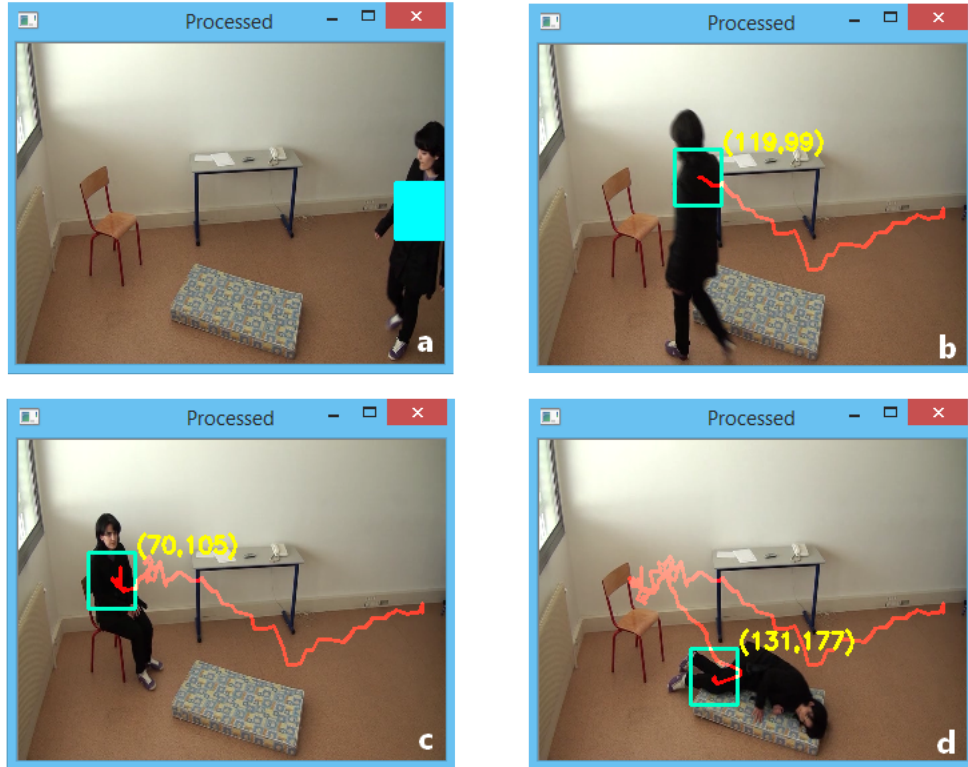


FIGURE 3.7 – Résultats du suivi avec le Mean-shift. (a) Définition de la région d'intérêt. (b) 88^{ème} trame de la sequence. (c) 145^{ème} trame de la sequence. (d) 217^{ème} trame de la sequence.

Nous avons initialisé la boîte englobante sur une région d'intérêt et pas sur l'objet entier pour éviter les pixels n'appartenant pas à l'objet. En effet Si l'histogramme de référence inclut des pixels parasites (fond, autre objet...), ils vont influencer la recherche lorsque l'objet est perdu : si l'objet se déplace et qu'il n'appartient plus au voisinage du point de recherche, les pixels les plus attractifs seront ceux qui ressemblent aux pixels parasites, ce qui peut éloigner le centre de recherche de la position réelle de l'objet comme illustré dans la figure 3.8.

3.2. SUIVI DE PERSONNE

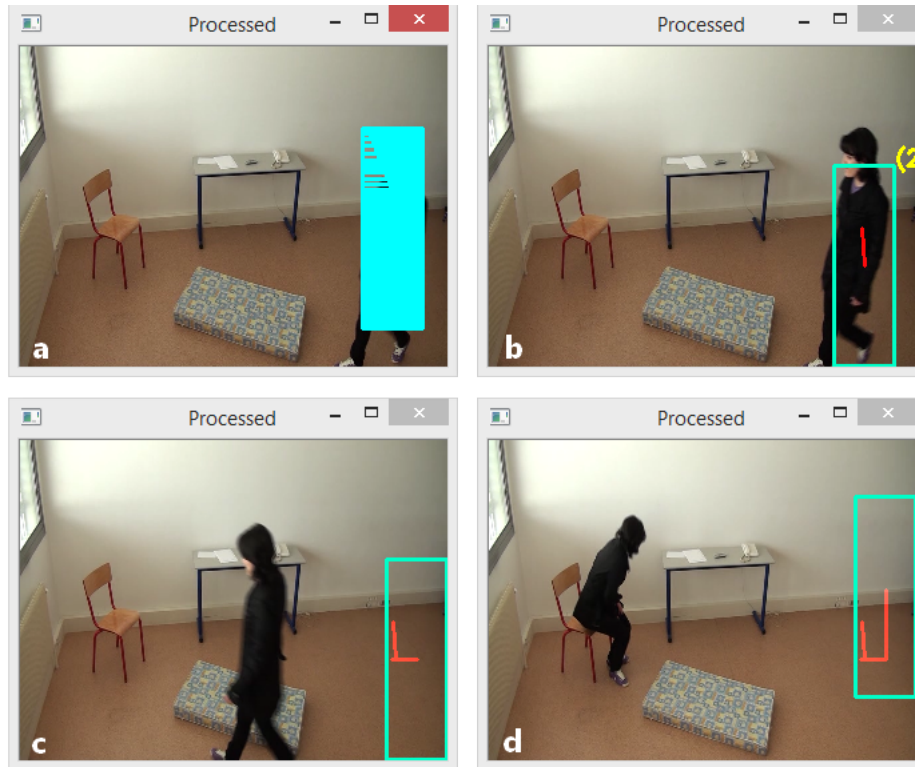


FIGURE 3.8 – Erreur du suivi en utilisant Mean-Shift. (a) Définition de l'objet d'intérêt en incluant des pixels du fond. (b) 46^{ème} trame de la sequence. (c) 71^{ème} trame de la sequence. (d) 119^{ème} trame de la sequence.

- **Le CamShift** : exploite habilement l'algorithme du mean-shift en modifiant la taille de la fenêtre lorsque ce dernier est arrivé à la convergence. le résultat de cette méthode est illustré par la figure suivante :

3.2. SUIVI DE PERSONNE

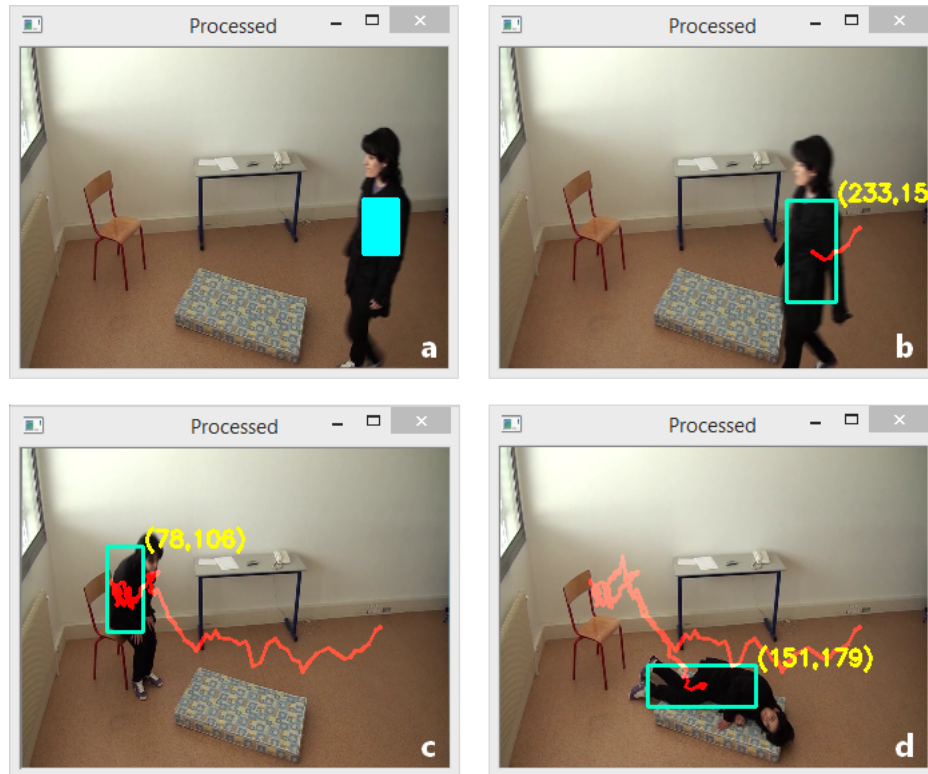


FIGURE 3.9 – Résultats du suivi avec le CamShift. (a) Définition de la région d'intérêt. (b) 59^{ème} trame de la sequence. (c) 177^{ème} trame de la sequence. (d) 219^{ème} trame de la sequence.

Ces deux méthodes utilisent la boîte englobante définie manuellement par l'utilisateur sur la région d'intérêt pour la chercher dans les trames suivantes. Nous avons automatisé cette tâche en appliquant une soustraction d'arrière-plan, une fois l'objet d'intérêt est détecté on l'isole dans une zone rectangulaire (boîte englobante), puis on calcule l'histogramme de cette région en utilisant que les pixels appartenant à l'objet à l'aide du masque obtenu par la soustraction d'arrière-plan, cela pour éviter la difficulté citée précédemment, qui réside dans la création de l'histogramme de l'objet de référence.

- **Filtre de Kalman** : Dans le processus du suivi, le filtre de Kalman regarde un objet quand il se déplace, c'est à dire il prend les informations sur l'état de l'objet à ce moment-la. Puis il utilise ces informations pour prédire où se trouve l'objet dans la prochaine trame. La position de l'objet est corrigée ensuite en considérant la prédiction et aussi l'observation. Pour les tests, nous avons considéré la position de l'objet comme étant le centre de gravité de son rectangle englobant, ce dernier est obtenu après une soustraction d'arrière-plan.

Le vecteur d'état composé seulement de la position et de la vitesse supposée constante, est présenté par :

$$X_t = \begin{pmatrix} x_t \\ y_t \\ v_{x_t} \\ v_{y_t} \end{pmatrix} \quad (3.1)$$

où x_t est l'abscisse de la cible, y_t son ordonné, v_{x_t} la vitesse en abscisse de la cible et v_{y_t} sa vitesse en ordonné.

La mesure (l'observation) est composée seulement de la position (centre de gravité) de l'objet mobile. Le vecteur de mesure est donné par :

$$Y_t = \begin{pmatrix} x_t \\ y_t \end{pmatrix} \quad (3.2)$$

La matrice de transition est donnée par :

$$F = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (3.3)$$

La matrice de mesure est :

$$H = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix} \quad (3.4)$$

Le résultat de cette méthode est illustré par la figure 3.10, le point blanc est la position réelle du centre de gravité de la personne et le cercle rouge est la position estimée.

3.2. SUIVI DE PERSONNE

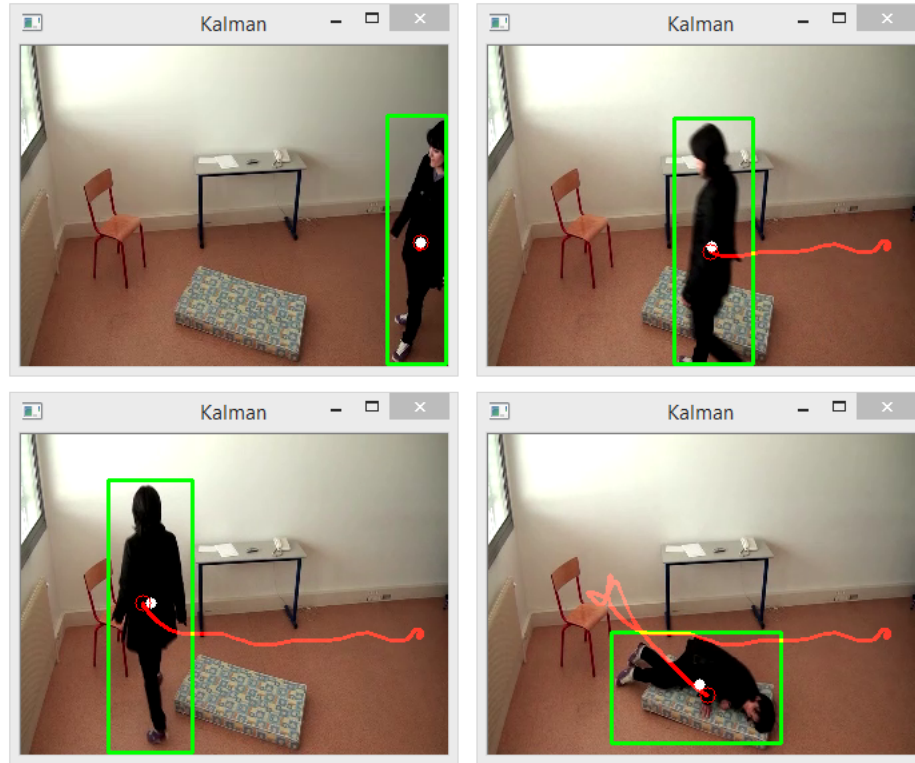


FIGURE 3.10 – Résultats du suivi avec le filtre de Kalman. (a) 25^{ème} trame de la sequence. (b) 74^{ème} trame de la sequence. (c) 99^{ème} trame de la sequence. (d) 216^{ème} trame de la sequence.



FIGURE 3.11 – Trajectoire mesurée/estimée de la personne en mouvement.

Sur la figure 3.11 la trajectoire mesurée (blanche) et celle estimée (rouge) sont conjointement illustrées sur l'image. On peut constater que la différence ou l'erreur commise sur l'estimation de la trajectoire est faible.

3.2.2 Discussion

Ici, nous présentons un tableau comparatif des différents algorithmes de suivi. On a classé toutes les méthodes selon plusieurs critères. Nous représentons cette classification par le nombre de signes « * ». Plus ce nombre est grand, plus la méthode est intéressante dans ce critère par rapport aux autres méthodes, et plus elle répond à nos besoins.

	Précision	Robustesse	Complexité	Implémen- tation	Rapidité
Mean-Shift	*	**	***	***	***
CamShift	**	***	***	***	***
Filtre de Kalman	***	***	**	**	**

TABLE 3.3 – Tableau comparatif, où le nombre d'étoiles est proportionnelle à l'efficacité de la méthode.

Le principal avantage des méthodes Mean-Shift et CamShift est qu'elles permettent d'estimer la position de l'objet en un nombre restreint d'itérations. En outre, leurs faibles complexités ainsi que leurs faibles temps de calcul permettent d'envisager un suivi en temps réel. Toutefois, dans le cas d'une scène contenant des objets ayant les mêmes distributions de couleurs et occupant

des régions très proches, l'estimation sera ambiguë. La précision dépend du nombre de bins choisi lors de l'initialisation. En effet, Le nombre de bins doit être égal pour la distribution de référence et pour la distribution candidate. Utiliser plus de bins revient à augmenter la résolution de l'histogramme et donc la précision de la poursuite. En contre partie, cela augmente la sensibilité au changement de luminosité.

Le filtre de Kalman donne des résultats satisfaisants. Cependant, comme tout processus itératif sa sensibilité aux valeurs initiales est grande. En effet, La convergence du filtre dépend de l'état initial à définir ce qui augmente le temps de calcul en cas d'une mauvaise initialisation. Un autre inconvénient se situe dans la délicate mise en oeuvre du procédé.

Conclusion

Dans ce chapitre, nous avons présenté les résultats expérimentaux de nos implémentations pour comparer les différentes méthodes de détection et de suivi de personnes en mouvement. Dans un premier temps, nous avons présenté les résultats de la détection des objets d'intérêt dans l'image. Nous avons conclu que la soustraction de fond reste l'approche la plus utilisée pour la détection et ceci est dû notamment au fait que, en plus d'être plus rapide et plus simple, elle exploite certaines techniques permettant de modéliser aisément les changements d'éclairage, le bruit, et les mouvements périodiques de régions dans l'image de fond. Ces capacités permettent d'effectuer une détection plus précise et plus adaptée dans une grande variété de situations. Les résultats présentés dans la deuxième partie consacrée au suivi de personnes montrent l'intérêt de l'utilisation du CamShift pour le suivi temps réel. Son succès provient de son faible temps de calcul, de sa faible complexité calculatoire associée à une bonne robustesse au bruit et de son invariance aux rotations et aux changements d'échelle.

CONCLUSION GÉNÉRALE ET PERSPECTIVES

Les personnes âgées en perte d'autonomie, vivant seules à domicile, font face à de nombreux risques, tels que des risques de chutes, de mauvaise médication ou de troubles cognitifs. Ces risques peuvent être diminués par l'utilisation d'un système de vidéosurveillance qui permet de surveiller la personne âgée. L'idée était de placer un réseau de caméras dans l'appartement de la personne pour le suivre, analyser ses comportements et par suite générer une alerte en cas de détection d'une situations de danger.

Dans un tel cadre, nous avons pu constater qu'il y avait trois points clés pour l'analyse de scène et l'interprétation d'activités : la détection d'individus d'intérêt, le suivi de ces individus et l'analyse de leurs trajectoires ainsi que de leurs actions.

Dans ce travail, on s'est intéressé à deux des problèmes majeurs d'analyse de mouvement dans une séquence d'images ; la détection et le suivi de personnes dans des scènes à l'intérieur où la caméra est fixe et le fond est statique. Détecter un individu dans une image consiste à séparer sa région de celle qui correspond au reste de l'image. Suivre un individu sur une séquence d'images est une opération qui permet de localiser cet objet au fil du temps afin de générer sa trajectoire. La représentation de la trajectoire permet alors de décrire les déplacements de cet individu, et peut être considérée comme une caractéristique déterminante qui aide à la discrimination entre des activités normales et anormales.

Dans l'état de l'art réalisé dans le deuxième chapitre, nous avons présenté, dans un premier temps, les diverses approches d'extraction des objets mobiles de l'arrière plan. Nous différencions la méthode de détection des points d'intérêt, les méthodes de la différence inter-images, celles reposant

sur une image de référence (soustraction d'arrière-plan) et celles sans image de référence (différence temporelle) et les méthodes basées sur un apprentissage supervisé. Le choix de l'approche dépend de l'application et du type de scène analysée. Il est également possible de combiner ces différentes approches. Par exemple, on peut détecter seulement les points d'intérêt situant dans l'objet en après une soustraction d'arrière-plan.

La deuxième partie de cet état de l'art a été consacrée aux méthodes couramment utilisées pour le suivi objet dans une séquence vidéo, où un objet est une zone de l'image qui peut être modélisée par des contours, silhouettes, primitives géométriques (rectangle englobant l'objet d'intérêt) ou encore par le point central. Chaque méthode tente de résoudre un problème bien déterminé en utilisant une modélisation d'objet adéquate qui dépend des propriétés utilisées (mouvement, histogramme de couleurs. etc).

Le troisième chapitre a été consacré à la comparaison des résultats obtenus par des différentes méthodes de détection et de suivi de personnes présentées dans l'état de l'art. L'objectif est déterminer les meilleurs méthodes pour avoir un système fiable et rapide. Pour ce qui concerne la détection, nous avons conclu que les résultats obtenus par la détection d'arrière-plan sont les plus satisfaisants, cette méthode est largement utilisée pour la détection des objets mobiles dans des séquences vidéo et ceci est dû notamment au fait que, en plus d'être plus rapide et plus simple, elle exploite certaines techniques permettant de modéliser aisément la dynamique de l'arrière-plan. La deuxième partie des expérimentations était dédiée au suivi de personne, nous avons présenté les résultats obtenus en utilisant le Mean-shift, CamShift et le filtre de Kalman. Nous avons conclu que le CamShift génère les meilleurs résultats de suivi parmi les techniques étudiées grâce à sa robustesse et sa faible complexité.

Les perspectives de ce travail sont les suivantes :

- Pour la détection de mouvement, nous avons choisi la première trame d'une séquences vidéos comme étant l'image de fond à utiliser pour la soustractions d'arrière-plan, cela produit des erreurs de détections lorsque l'individu est présent dans la première trame (nous avons obtenu deux blobs dans l'image résultante de la soustraction, un pour la première position de l'individu et l'autre pour sa position courante). Afin d'améliorer la robustesse de cette méthode, il faut faire une étude sur les méthodes permettant d'obtenir automatiquement un arrière-plan de la scène qui soit le plus robuste aux changements de l'éclairage, aux ombres, et au bruit présent dans la séquence vidéo.

- Une fois les blobs détectés, il nous faut trouver dans ces blobs une caractéristique, tel que la peau, qui nous permette d'affirmer que les éléments auxquels ils correspondent sont bien des humains et non pas des objets afin de lancer par la suite le processus du suivi. Pour cela, il serait intéressant de faire un état de l'art des méthodes de la détection de la peau dans une image.
- Pour le suivi, aucune étude concernant le choix des caractéristiques n'a été mise en place (espace de couleurs, taille d'histogramme, nombre de canaux utilisés par histogramme, etc.). Il serait nécessaire d'étudier l'influence du choix des caractéristiques afin de ne conserver que celles les plus discriminantes pour l'application de suivi.
- Le suivi utilisant CamShift offre des résultats acceptables mais reste assujéti à de nombreuses contraintes comme l'occultation de la cible ou un changement trop important de son apparence. Pour augmenter la robustesse l'estimation du déplacement, nous pouvons intégrer un filtre de Kalman dans l'algorithme CamShift. Lorsque le coefficient de similarité entre l'objet de référence et les objets candidats est faible, on peut considérer l'objet comme étant perdu. Dans ce cas, le filtre de Kalman peut prendre le relais afin d'estimer la nouvelle position de l'objet en fonction de ces états précédents.
- L'idéal c'est d'avoir un système de vidéosurveillance entièrement intelligent. Pour atteindre ce but, il faudra aborder le module de l'analyse d'activités pour avoir énormément d'information sur la personne surveillée. En plus de connaître sa localisation dans une pièce, son mouvement et ses actions, un système de vidéosurveillance pourrait permettre d'analyser le comportement de la personne, à savoir si elle prend correctement ses médicaments, si elle mange et dort à des heures régulières et combien de temps dure son sommeil. Il serait aussi possible de vérifier la durée et le moment d'autres activités quotidiennes telles que les repas, la télévision, la salle de bain, etc. Toutes ces données vont être des indicateurs pour mieux prédire les accidents.

BIBLIOGRAPHIE

- [1] A. Hampapur, L. Brown, J. Connell, S. Pankanti, A. Senior, Y. Tian. Smart Surveillance : Applications, Technologies and Implications. *2003 Joint Conference of the Fourth International Conference on Information, Communications & Signal Processing and Fourth Pacific-Rim Conference on Multimedia*, vol. 2, pp. 1133-1138, New York, United States, December 15-18, 2003.
- [2] J. K. PETERSEN : *Understanding Surveillance Technologies*. CRC Press ; 1 edition, 2000.
- [3] N.T. Nguyen, D.Q. Phung, S. Venkatesh, H.H. Bui. Learning and detecting activities from movement trajectories using the hierarchical hidden markov models. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 2, pp. 955-960, 2005.
- [4] A. Corradini. Dynamic time warping for off-line recognition of a small gesture vocabulary. *IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, pp. 82-89, 2001.
- [5] K. Sobottka, I. Pitas. Face localization and facial feature extraction based on shape and color information. *International Conference on Image Processing (ICIP)*, vol. 3, pp. 483-486, 1996.
- [6] D. Anderson, R.H. Luke, J.M. Keller, M. Skubic, M. Rantz, M. Aud. Linguistic summarization of video for fall detection using voxel person and fuzzy logic. *Computer Vision and Image Understanding*, vol. 113, no. 1, pp. 80-89, January 2009.

- [7] K. Dimitropoulos, T. Semertzidis, N. Grammalidis. Video and signal based surveillance for airport applications. *Proceedings of the 6th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 170-175, Genova, Italy, 2009.
- [8] F. Cupillard, A. Avanzi, F. Brémond, M. Thonnat. Video understanding for metro surveillance. *The IEEE ICNSC 2004 in the special session on Intelligent Transportation Systems*, vol. 1, pp. 186-191, 2004.
- [9] J. Owens, A. Hunter. Application of the self-organizing map to trajectory classification. *Proceedings of the Third IEEE International Workshop on Visual Surveillance (VS'2000)*, pp. 77-83, 2000.
- [10] A. Leykin, M. Tuceryan. A vision system for automated customer tracking for marketing analysis : Low level feature extraction. *Proceedings of the International Workshop on Human Activity Recognition and Modelling, HAREM'05*, pp. 1-7, 2005.
- [11] T. Wassantachat, Z. Li, J. Chen, Y. Wang, E. Tan. Traffic density estimation with on-line SVM classifier. *Proceedings of the 6th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 13-18, Genova, Italy, 2009.
- [12] J. R. Wang, 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.
- [13] P. Zhu, W. Hu, X. Li, L. Li. Segment model based vehicle motion analysis. *Proceedings of the 6th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 514-517, Genova, Italy, 2009.
- [14] Y. Zhai, A. Hampapur. Virtual boundary crossing detection without explicit object tracking. *Proceedings of the 6th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 518-522, Genova, Italy, 2009.
- [15] C. Harris, M. Stephens. A combined corner and edge detector. *In 6th Alvey Vision Conference*, pp. 147-151, 1988.
- [16] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, November 2004.

- [17] T. Kanade, C. Tomasi. Detection and tracking of point feature. *Technical report CMU-CS-91-132*, Carnegie Mellon University, April 1991.
- [18] Y. Benabbas. Analyse du comportement humain a partir de la vidéo en étudiant l'orientation du mouvement. *Thèse de doctorat*, Lille I, 2012.
- [19] Y. Benezeth, P.M. Jodoin, B. Emile, H. Laurent, C. Rosenberger. Comparative Study of Background Subtraction Algorithm, *Journal of Electronic Imaging*, vol. 19, no. 3, 2010.
- [20] D.A. Migliore, M. Matteucci, M. Naccari. A revaluation of frame difference in fast and robust motion detection. *In Proceedings of the 4th ACM international workshop on Video surveillance and sensor networks*, pp. 215-218. ACM, 2006.
- [21] Y. Kameda, M. Minoh. A human motion estimation method using 3-successive video frames. *In International Conference on Virtual Systems and Multimedia*, pp. 135-140, 1996.
- [22] H. Rowley, S. Baluja, T. Andkanade. Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 23-38, 1998.
- [23] C. Papageorgiou, M. Oren, T. Poggio. A general framework for object detection. *In IEEE International Conference on Computer Vision (ICCV)*, pp. 555-562, 1998.
- [24] A. Yilmaz, O. Javed, M. Shah. Object tracking : A survey. *Acm Computing Surveys (CSUR)*, vol. 38, no. 4, pp. 13, 2006.
- [25] C. J. Veenman, M. J. T. Reinders, E. Backer. Resolving motion correspondence for densely moving points. *IEEE Transactions Pattern Analysis Machine Intelligence*, vol. 23, no. 1, pp. 54-72, 2001.
- [26] D. Serby, E. Koller-Meier, L. Van Gool. Probabilistic object tracking using multiple features. *In International Conference on Pattern Recognition*, pp. 184-187, Washington, DC, USA, 2004.
- [27] A. Yilmaz, X. Li, M. Shah. Contour-based object tracking with occlusion handling in video acquired using mobile cameras. *IEEE Transactions Pattern Analysis Machine Intelligence*, vol. 26, no. 11, pp. 1531-1536, 2004.
- [28] D.-S. Jang, S.-W. Jang, H. Choi. Tracking a partially occluded target with a cluster of kalman filters. *International Journal of Intelligent Systems*, vol. 17, no. 6, pp. 595-607, 2002.

- [29] L. Jiao, Yi Wu, G. Wu, E.Y. Chang, Y.-F. Wang. Anatomy of a multi-camera video surveillance system. *Multimedia Systems*, vol. 10, no. 2, pp. 144-163, August 2004.
- [30] M. Isard A. Black. Condensation – Conditional Density Propagation for Visual Tracking. *International Journal of Computer Vision*, vol. 29, no. 1, pp. 5-28, 1998.
- [31] L. Fuentes et Sergio A. Velastin. People Tracking in Surveillance Applications. *Image and Vision Computing*, vol. 24, pp. 1165-1171, 2006.
- [32] D. Comaniciu, V. Ramesh, and P. Meer, Real-time tracking of non-rigid objects using mean shift. *In Computer Vision and Pattern Recognition. Proceedings. IEEE Conference on*, vol. 2, pp. 142-149, IEEE, 2000.
- [33] D. Fleet et Y. Weiss. Optical Flow Estimation. *Handbook of Mathematical Models in Computer Vision*, pp. 239-258, 2005.
- [34] [http ://le2i.cnrs.fr/Fall-detection-Dataset](http://le2i.cnrs.fr/Fall-detection-Dataset).