

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

ECOLE NATIONALE POLYTECHNIQUE
Département de Génie Electrique



Projet de Master

En vue de l'obtention du grade de Master en Electronique

THEME :

*La reconnaissance de mouvement dans une
séquence d'images*

Présenté par :

Mr. Y. HAMMOUCHE

Proposé par :

Mr. C. LARBES

Mme. S. AIT DAOUD

Septembre 2012

Ecole Nationale Polytechnique
10, Avenue Hacén Badi, El-Harrach, Alger.

Remerciements

J'exprime ma reconnaissance à Monsieur C. LARBES d'avoir joué pleinement son rôle de promoteur en étant à mes côtés tout au long de l'étude de ce projet. Ses conseils et orientations m'ont guidé jusqu'à l'aboutissement de ce travail.

Je remercie ma promotrice madame S. AIT DAOUD pour ses remarques pertinentes qui ont apporté une amélioration certaine à mon travail. Je la remercie également d'avoir examiné pleinement mon travail et proposé des modifications constructives.

Je tiens de même à remercier Monsieur L. ABDELOUAL, MA-A à l'*Ecole Nationale Polytechnique d'Alger*, d'avoir accepté de présider le jury. Mes remerciements vont aussi à Monsieur M. O. TAGHI, MA-A à l'*Ecole Nationale Polytechnique d'Alger*, d'avoir bien voulu examiner mon travail.

Que tous ceux et toutes celles qui ont contribué de près ou de loin à l'aboutissement de ce travail trouvent ici mes sincères remerciements et l'expression de ma reconnaissance.

Dédicaces

Je dédie ce modeste travail à

Ma mère,

Mon père,

Mes frères,

Et ma sœur

qui m'ont soutenue sans relâche, dans toutes les circonstances, tout au long de mon parcours d'études.

Je dédie ce travail à toute ma famille pour leur soutien.

Je dédie ce travail à ma chère **A**mira et à tous mes ami(e)s pour leur soutien tout au long de mon parcours.

Y. HAMMOUCHE.

Résumé

La reconnaissance de mouvement est l'une des applications essentielles de la détection de mouvement, qui fait appel au traitement d'images. De ce fait, le travail proposé n'est qu'une extension du projet de fin d'études, intitulé « Etude, et simulation de la GMM adaptative pour la détection de mouvement ». Ce travail comportera une étude détaillée d'une méthode de reconnaissance de mouvement et de l'identité des personnes, à la fois automatique et en temps réel, dans un flux vidéo.

Mot clé : reconnaissance, détection, mouvement, démarche, vidéo-surveillance, ...

Abstract

Motion recognition is one of the essentials applications of motion detection, which uses image processing. Therefore, the proposed work is an extension of the graduation project, titled "Design, and simulation of adaptive GMM to motion detection." This work will include a detailed study of an approach for motion recognition and identity of persons, automatically and in real time, in a video stream.

Keyword : recognition, detection, motion, gait, video monitoring, ...

ملخص

الاعتراف بالحركة هي واحدة من أساسيات تطبيقات كشف الحركة، والتي تستخدم لمعالجة الصور، ولذلك فإن العمل المقترح هو امتداد لمشروع تخرج بعنوان "تصميم وتمثيل GMM لكشف الحركة، وهذا العمل يشمل دراسة تفصيلية إتباع لاعتراف الحركة ووهية الأشخاص، تلقائيا في الوقت الحاضر فيديو

الكلمات الرئيسية: الإعراف، الكشف، الحركة، الخطوة ومراقبة الفيديو.

Table des matières

Introduction générale	(1)
------------------------------------	-----

Chapitre I : Notions d'analyse de mouvement, et de détection de mouvement

I.1. Introduction	(3)
I.2. Notions de bases de vision et d'analyse de mouvement	(3)
I.2.1. La vision par ordinateur.....	(3)
I.2.2. L'analyse de mouvement.....	(4)
I.3. La détection de mouvement	(5)
I.3.1. Classification des méthodes de détection de mouvement.....	(5)
I.3.2. Revue des méthodes de détection basées sur la soustraction de fond.....	(7)
I.4. Conclusion	(12)

Chapitre II : La reconnaissance de mouvement dans une séquence vidéo

II.1. Introduction	(13)
II.2. Intérêts, et champs d'application de la reconnaissance de mouvement	(13)
II.2.1. La surveillance et identification des activités et comportement des personnes.....	(14)
II.2.2. La détection et la reconnaissance des personnes.....	(16)
II.3. Détection et identification des personnes par reconnaissance frontale de leur démarche	(18)
II.3.1. Etape un : Détection des zones en mouvement.....	(18)

II.3.2. Etape deux : Description et classification des silhouettes.....	(22)
II.3.2.1. Description des silhouettes.....	(23)
II.3.2.2. Classification des silhouettes.....	(25)
II.3.3. Etape trois : Identification des personnes détectées à base de leur démarche.....	(28)
II.4. Conclusion.....	(30)
 Conclusion générale.....	 (32)

Table des figures

Chapitre I :

Figure I.1 : Processus de traitement et d'analyse d'un système visuel artificiel.....(4)

Figure I.2 : Classification des méthodes de détection de mouvement.....(7)

Chapitre II :

Figure II.1 : Images de vidéo-surveillance montrant des vols de bagage dans un aéroport.....(15)

Figure II.2 : Images de vidéo-surveillance montrant une détection de comportements dangereux liés aux bagarres.....(15)

Figure II.3 : Image de vidéo-surveillance montrant un bagage abandonné susceptible de contenir une bombe.....(15)

Figure II.4 : Image d'un iris et d'une empreinte.....(16)

Figure II.5 : Reconnaissance de personnes en utilisant les techniques biométriques faciales.....(17)

Figure II.6 : Mélange de distributions Gaussiennes.....(19)

Figure II.7 : Classification des pixels soit faisant parti de l'avant ou l'arrière-plan.....(21)

Figure II.8 : Exemples de silhouettes extraites par la technique de soustraction de fond basée sur le mélange de Gaussiennes.....(21)

Figure II.9 : Architecture de l'algorithme de description et classification des silhouettes.....(22)

Figure II.10. : Exemple de rectangles maximaux coincés à l'intérieur d'une silhouette.....(24)

Figure II.11. : Exemples de distributions de dimension rectangles pour une forme humaine.....(25)

Figure II.12. : Exemple de classification positive (le rectangle en blanc, indique que le système à classer la forme détectée comme silhouette humaine).....(27)

Figure II.13. : Exemple de classification négative (dans ce cas de figure, soit le système n'a pas pu classer aucune forme d'objets détecter, ou bien la forme classer ne correspond pas à une silhouette humaine).....(28)

Figure II.14. : Etapes de l'algorithme de reconnaissance de la démarche.....(29)

Introduction générale

Avec la généralisation de l'utilisation d'images numériques, l'analyse du mouvement dans les séquences vidéo s'est révélée être un outil indispensable pour diverses applications telles la vidéo surveillance, l'imagerie médicale, la robotique...etc. L'analyse du mouvement dans une séquence d'images est un vaste sujet qui regroupe plusieurs problématiques. On peut notamment citer l'estimation du champ de mouvement, la détection de mouvement, le suivi de mouvement, et la reconnaissance de mouvement.

La problématique qui se pose dans ce manuscrit est la reconnaissance de mouvement dans une séquence vidéo, qui fait appel au traitement d'images. Le traitement d'images est un domaine basé sur plusieurs disciplines : les mathématiques, les probabilités, ... etc.

La reconnaissance de mouvement est une branche de la vision artificielle. Elle consiste à identifier des formes pré-décrites dans une image numérique, et par extension dans un flux vidéo numérique. En dépit de l'énorme progression en informatique et plus spécifiquement en traitement d'images, les systèmes appliquant les techniques de reconnaissance sont considérés comme un gros challenge très intéressant, qui a attiré les chercheurs de différents domaines : psychologie, identification de modèles, réseaux de neurones, vision par ordinateur, infographie ...

Ce manuscrit se décompose en deux grands chapitres :

Chapitre I « Notions d'analyse de mouvement, et de détection de mouvement »

Nous verrons d'une manière générale dans ce chapitre, les notions de bases de vision et d'analyse de mouvement. Par la suite, une revue des différents algorithmes de détection, basés sur la soustraction de fond, largement utilisés, sera faite.

Chapitre II « La reconnaissance de mouvement dans une séquence vidéo »

Ce chapitre est consacré uniquement à la reconnaissance de mouvement concernant les activités des sujets humains dans un milieu extérieur, ainsi que l'identification de ces sujets

Introduction générale

pour de raisons précises. Nous ferons par la suite, l'étude d'une méthode de reconnaissance, en temps réel, basée sur la démarche.

Chapitre I

Notions d'analyse de mouvement, et de détection de mouvement

I.1. Introduction :

La perception d'un mouvement dans une scène visuelle est un mécanisme très complexe d'un point de vue de sa compréhension. Pour le comprendre, des scientifiques ont cherché à le reproduire en développant des systèmes de vision artificielle par ordinateur capable d'observer et d'interpréter des scènes visuelles.

La vision par ordinateur est une reproduction du système d'interprétation visuelle humain sur une machine. Elle peut s'énoncer comme étant un processus de traitement et d'analyse d'une scène. L'analyse du mouvement est un problème fondamental, en particulier pour des domaines tels que la post-production, la compression vidéo, la vidéo surveillance, la reconnaissance, l'analyse sémantique de scènes, ...etc.

A ce jour, plusieurs modèles, méthodes d'analyse du mouvement ont été mis au point. Nous verrons dans ce chapitre les notions de base de la vision et d'analyse de mouvement. Par la suite, nous aborderons la notion de détection de mouvement dans une séquence d'images tout en citant les différentes méthodes et domaines d'application.

I.2. Notions de bases de vision et d'analyse de mouvement :

I.2.1. La vision par ordinateur :

La « Vision » n'est que, discerner et reconnaître les formes, les couleurs et les textures du monde qui nous entoure. On pourrait croire, que nous avons besoin uniquement de nos yeux pour voir, mais la réalité est tout autre : nos yeux ne servent qu'à capter le signal contenu dans les rayons lumineux qui viennent frapper nos rétines mais non pas à en extraire des informations. L'extraction de ces informations se fera à l'aide de notre cerveau (dans la zone appelée « le cortex visuel »). Cela veut dire que la vision, est l'association entre le sens de la vue et un ensemble de processus cognitifs pour lesquels nous avons besoin d'utiliser notre cerveau.

Cela fait déjà bien longtemps que le sens de la vue a été donné aux ordinateurs grâce aux caméras numériques, mais si l'on se fie à la définition que nous venons de formuler, cela ne suffit pas à les doter de vision : il leur manque encore la capacité à extraire des informations des images et des flux vidéo, de manière à percevoir, analyser et comprendre le monde observé.

Grace aux intérêts des scientifiques en vue de comprendre l'anatomie et le fonctionnement du système visuel, et avec la naissance de machines de calcul puissantes et très sophistiquées en parallèle, de nouvelles questions se sont posées telles que : Est-il possible de construire un modèle de perception visuelle ? De ce fait, des théories ainsi que des modèles ont été mis au point, pour permettre à la vision par ordinateur de voir le jour.

La vision par ordinateur peut s'énoncer comme étant un processus de traitement et d'analyse d'une scène. En concordance avec le système visuel humain, le système de vision artificielle se compose de différentes fonctions, comme le montre la figure ci-dessous :

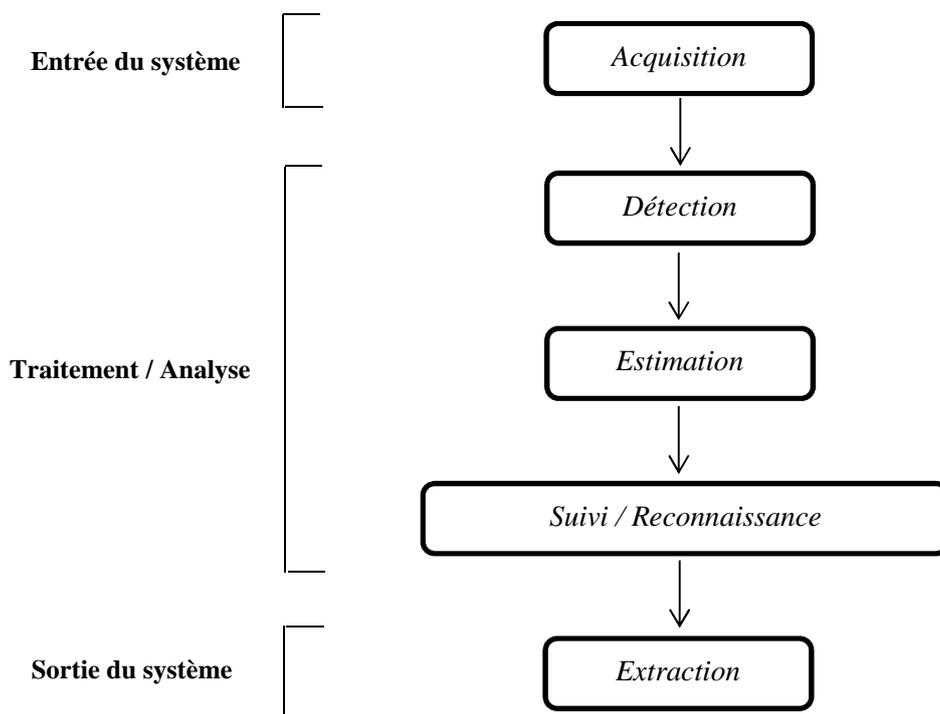


Figure .I.1 : Processus de traitement et d'analyse d'un système visuel artificielle. [11]

I.2.2. L'analyse de mouvement :

L'analyse du mouvement fait partie intégrante du traitement d'images. C'est un domaine apparu avec l'avènement de nouvelles technologies de l'information. Elle vise à analyser des séquences d'images numériques par un ensemble de méthodes, afin d'en tirer des informations potentielles. Ces informations seront exploitées pour répondre à des problèmes dans des domaines bien précis. [8]

La notion de mouvement dans une séquence visuelle est représentée par une variation d'intensité lumineuse des pixels constituant l'image au long de la séquence. On peut citer deux types de mouvement :

- **Mouvement réel** : mouvement tel qu'il est perçu par l'œil.
- **Mouvement apparent** : mouvement résultant de la projection du mouvement réel dans un plan d'image en 2D.

Nous pouvons rencontrer différents problèmes liés au mouvement dans une scène visuelle, tels que :

- La détection de mouvement : y a-t-il eu un mouvement dans la scène ? ;
- L'estimation du mouvement : problème de localisation (vitesse et position) ;
- La reconnaissance du mouvement : problème de reconnaissance du scénario correspondant au mouvement ;
- Le suivi de mouvement : problème de correspondance entre plusieurs images pour un même mouvement ;

I.3. La détection de mouvement :

La détection de mouvement est parmi les phases de traitement et d'analyse de mouvement les plus importantes. Utilisée dans plusieurs champs d'activité et domaines bien précis, la détection de mouvement indique uniquement que l'on parle d'une méthode qui a pour objet de trouver en quels points de l'image un mouvement a eu lieu. Un algorithme ayant cet objectif fournit en sortie une variable quantitative (quantité de mouvement) ou qualitative (booléenne) pour tout pixel de chaque image d'entrée. [8, 11]

I.3.1. Classification des méthodes de détection de mouvement : [12]

La plupart des états de l'art publiés jusqu'à présent sur la détection de mouvement, sont fortement orientés vers un domaine d'application précis. Avec la classification hiérarchisée des différentes méthodes de détection de mouvement, nous distinguerons quatre grandes familles de méthodes en fonction de la modélisation de l'arrière-plan.

- Détection sans modélisation de l'arrière-plan :

Ces méthodes consistent à détecter le mouvement par le calcul en tout point de l'image d'une quantité mathématique qui est fonction de l'intensité ou de la couleur de l'ensemble des

pixels et qui est censée refléter l'importance du mouvement visible dans la scène. Comme exemple on cite la dérivée temporelle de l'intensité lumineuse, l'entropie spatio-temporelle de l'image et la norme du flot optique. [12]

- Modélisation locale de l'arrière-plan :

Ces méthodes consistent à associer à tout point de l'image une valeur ou une fonction permettant de modéliser l'apparence de l'arrière-plan en ce point. Le modèle d'apparence de l'arrière-plan en un point ne dépend que des observations qui ont eu lieu en ce point.

Les autres pixels de l'image n'interviennent pas. La grande majorité des méthodes présentées dans la littérature bâtissent un modèle statistique (ensemble de paramètres d'une loi, ou ensemble d'échantillons), mais il peut s'agir d'un processus stochastique, d'un filtre prédictif ou simplement d'une valeur d'intensité.

- Modélisation semi-locale de l'arrière-plan :

Ces méthodes sont très semblables à celles de la catégorie précédente, à la différence près que la modélisation de l'arrière-plan en un point dépend des observations qui ont eu lieu dans un certain voisinage de ce point, ou dans la région de l'image à laquelle il appartient.

- Modélisation globale de l'arrière-plan :

Ces méthodes utilisent à chaque instant l'ensemble des observations pour construire un modèle de l'ensemble de l'arrière-plan.

La figure suivante résume cette classification :

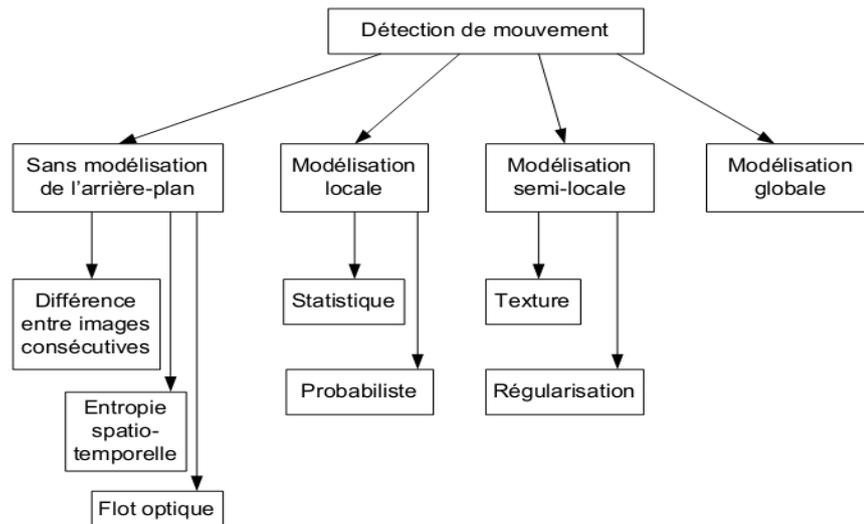


Figure I.2. : Classification des méthodes de détection de mouvement. [12]

Nous nous intéresserons dans ce qui va suivre aux méthodes de détection basées sur la modélisation du fond.

I.3.2. Revue des méthodes de détection basées sur la soustraction de fond :

Afin de détecter, segmenter, et suivre des objets automatiquement dans les vidéos, plusieurs approches sont possibles. Un simple algorithme de détection de mouvement compare une trame de fond statique avec l'image actuelle d'une scène vidéo, pixel par pixel. C'est le principe fondamental de soustraction de fond, qui peut être formulée comme une technique qui construit un modèle de fond et compare ce modèle avec l'image actuelle afin de détecter les zones où une différence significative se produit. Le but d'un algorithme de soustraction de fond est donc de distinguer les objets en mouvement (de l'avant-plan) d'être statique ou lent, des parties de la scène (appelé arrière-plan).

Bien qu'un modèle de fond statique puisse être approprié pour l'analyse de courtes séquences vidéo dans un environnement contraint à l'intérieur, le modèle est inefficace pour la plupart des situations pratiques, un modèle plus sophistiqué est donc nécessaire. En outre, la détection de mouvement n'est souvent que la première étape d'un processus de compréhension de la scène. Par exemple :

- la détection de sacs sans surveillance ;
- la reconnaissance de la démarche ;
- la détection de visage ;

- la surveillance du trafic routier ;

Le problème abordé par des techniques de soustraction de fond implique la comparaison d'une image observée avec une image estimée qui ne contient pas n'importe quel objet d'intérêt, ce qui est considéré comme le modèle de fond (ou image de fond) [14]. Ce processus de comparaison, appelée détection de l'avant-plan, divise l'image observée en deux ensembles complémentaires de pixels qui couvrent la totalité de l'image au premier plan qui contient les objets d'intérêt, et l'arrière-plan, son ensemble complémentaire. Comme indiqué dans [13], il est difficile de préciser une définition de référence de ce qu'est une technique de soustraction de fond, cette dernière se rapporte au niveau de l'application.

Beaucoup de techniques de soustraction de fond ont été proposées avec de nombreux modèles et des stratégies de segmentation, et plusieurs enquêtes sont consacrés à ce sujet (dans [9, 14, 15, 13] par exemple). Certains algorithmes se concentrent sur les exigences spécifiques qu'une technique idéale de soustraction de fond pourrait ou devrait remplir. Selon [15], une technique de soustraction de fond doit s'adapter aux changements d'illumination progressive ou rapide (temps de changement de la journée, nuages, etc), les changements de mouvement (oscillations de caméra), des objets de haute fréquence (mouvement de feuilles d'arbre ou les branches), et les changements dans la géométrie de l'arrière-plan (par exemple des voitures en stationnement). Certaines applications nécessitent des algorithmes de soustraction de fond incorporés dans l'appareil photo, de sorte que la charge de calcul devient la préoccupation majeure. Pour la surveillance des scènes en plein air, de la robustesse contre le bruit et l'adaptabilité aux changements d'illumination sont également essentiels.

La plupart des techniques décrites dans la littérature fonctionnent sur chaque pixel de façon indépendante. Ces techniques relèquent entièrement à des algorithmes de post-traitement dans la tâche est d'ajouter une certaine forme de cohérence spatiale aux résultats. Etant donné que les perturbations affectent souvent les pixels individuels, il en résulte des erreurs de classification locales. En revanche, la méthode décrite par Seiki et al. dans [16], est basée sur la supposition que les blocs adjacents de pixels d'arrière-plan doivent suivre les variations dans le temps similaires. Bien que cette hypothèse tienne la plupart du temps, en particulier pour les pixels appartenant à l'objet du même contexte, il devient problématique pour les pixels voisins situés à la frontière des objets de fond multiples. Malgré cet inconvénient, les pixels sont regroupés en blocs et chaque bloc $N \times N$ est traité comme un N^2 -composante du vecteur. Quelques échantillons sont ensuite recueillis au fil du temps et

utilisés pour former un modèle d'Analyse en Composantes Principales (ACP) pour chaque bloc. Un bloc d'une trame vidéo est classé comme arrière-plan si la configuration d'image observée est à proximité de ses coefficients de projection en utilisant les reconstructions de l'ACP de 8- blocs voisins. Bien que le modèle ACP soit également formé avec des échantillons de temps, le modèle résultant compte pour l'image entière. Les pixels individuels sont classés comme arrière-plan ou au premier plan en utilisant de simples seuillages qui est la différence entre l'image courante et la rétroprojection dans l'espace image de ses coefficients de l'ACP. Comme pour les autres méthodes basées sur la ACP, le processus d'initialisation et le mécanisme de mise à jour ne sont pas décrits. [8]

Un mécanisme à deux niveaux sur la base d'un classificateur est introduit dans [18]. Un classificateur détermine d'abord si un bloc d'image appartient à l'arrière-plan. Des mises à jour appropriées par blocs de l'image d'arrière-plan sont ensuite effectuées dans la deuxième étape, en fonction des résultats de la classification.

Des algorithmes basés sur le cadre de la détection compressée effectuent la soustraction de fond par l'apprentissage et l'adaptation d'une faible représentation tridimensionnelle comprimée de l'arrière-plan [17]. L'avantage majeur de cette approche réside dans le fait que la détection compressée estime les silhouettes des objets sans aucune reconstruction d'une image auxiliaire. D'autre part, les objets du premier plan doivent occuper seulement une petite partie de la vue de la caméra afin d'être correctement détectés.

L'estimation d'arrière-plan est formulée dans [22] en tant que problème d'étiquetage optimal dans lequel chaque pixel de l'image d'arrière-plan est marqué avec un numéro de trame, qui indique la couleur de la dernière qui doit être copiée. L'algorithme proposé par l'auteur produit une image de fond, qui est construite en copiant les zones à partir des trames d'entrée. Des résultats impressionnants sont présentés pour arrière-plans statiques, mais la méthode n'est pas conçue pour faire face à des objets se déplaçant lentement dans le fond, son résultat est une trame de fond statique unique.

Les auteurs de [19] ont été inspirés par le mécanisme biologique de mouvement basé sur le groupement perceptuel. Ils proposent un algorithme de saillance spatio-temporelle applicable aux scènes d'horizons très dynamiques, qui peuvent être utilisées pour effectuer la soustraction du fond. Les comparaisons de leur algorithme avec d'autres techniques montrent que leur algorithme réduit le taux d'erreur moyen, mais à un coût en temps de traitement

prohibitif (plusieurs secondes par image), ce qui le rend impropre à des applications en temps réel.

Les techniques de soustraction à base de pixels visent à compenser le manque de cohérence spatiale par une mise à jour constante de leurs paramètres du modèle. Les techniques les plus simples dans cette catégorie sont l'utilisation d'un cadre de fond statique (qui a récemment été utilisée), le fonctionnement moyen (pondéré), filtrage passe-bas du premier ordre, le filtrage médian temporel, et la modélisation de chaque pixel avec une Gaussienne. [7]

Les méthodes basées sur les filtres de détection de mouvement (sigma-delta) [10], sont très populaires pour le traitement embarqué. Comme dans le cas des convertisseurs analogique-numérique, un filtre de détection de mouvement (sigma-delta) se compose d'une simple approximation non-linéaire récursive de l'image de fond, qui est basée sur la comparaison et sur une augmentation élémentaire / décrétement (habituellement -1, 0 et 1 sont les seules valeurs possibles de mise à jour). Le filtre de détection de mouvement (sigma-delta) est donc bien adapté à de nombreux systèmes embarqués qui ne disposent pas d'une unité à virgule flottante. [7]

Toutes ces techniques uni-modales peuvent conduire à des résultats satisfaisants dans des environnements contrôlés, tout en restant rapides, faciles à mettre en œuvre, et simples. Cependant, des méthodes plus sophistiquées sont nécessaires lorsqu'il s'agit de vidéos capturées dans des environnements complexes où les mouvements de fond, mouvement de la caméra, et le bruit d'un capteur sont rencontrés [7].

Au fil des années, de plus en plus d'algorithmes complexes au niveau du pixel ont été proposés. Parmi ceux-ci, de loin le plus populaire est le modèle de mélange Gaussien (GMM). Présenté dans [20], ce modèle consiste à modéliser la distribution des valeurs observées dans le temps de chaque pixel par un mélange pondéré de Gaussiennes. Ce modèle est en mesure de faire face à la nature multimodale de nombreuses situations concrètes et donne de bons résultats pour des détections de fond répétitives, telles que les feuilles des arbres ou des branches. Depuis son introduction, le modèle a gagné beaucoup en popularité parmi la communauté de vision par ordinateur, et il soulève encore beaucoup d'intérêt que les auteurs continuent de revoir la méthode et de proposer des algorithmes améliorés [8, 11]. Dans [21], une méthode d'optimisation est proposée pour déterminer automatiquement les paramètres de l'algorithme GMM.

L'inconvénient de l'algorithme de GMM réside dans ses hypothèses fortes que le fond est plus souvent visible que le premier plan et que sa variance est significativement plus faible. Rien de tout cela n'est valable pour chaque fenêtre de temps. En outre, si la haute et basse fréquence des changements sont présents dans l'arrière-plan, sa sensibilité ne peut pas être à l'écoute et la précision du modèle peut s'adapter à des cibles elles-mêmes ou de manquer la détection de certaines cibles à haute vitesse. En outre, l'estimation des paramètres du modèle (en particulier la variance) peut devenir problématique dans les environnements bruyants du monde réel. Ce qui ne laisse souvent pas à d'autre choix que d'utiliser un écart fixe dans une implémentation matérielle. Enfin, il convient de noter que la pertinence statistique d'un modèle gaussien est discutable, car certains auteurs affirment que les images naturelles présentent des statistiques non gaussiennes.

Pour éviter la question difficile de trouver une forme appropriée pour la fonction de densité de probabilité, certains auteurs ont tourné leur attention vers des méthodes non paramétriques. Un des points forts des méthodes non paramétriques de densité estimée est leur capacité à contourner une partie de l'étape d'estimation des paramètres délicate en raison du fait qu'elles s'appuient sur des valeurs de pixels observés dans le passé. Pour chaque pixel, ces méthodes construisent un histogramme de valeurs d'arrière-plan en accumulant une série de valeurs réelles échantillonnées du passé récent du pixel. Puis estimer la fonction de densité de probabilité avec cet histogramme pour déterminer si oui ou non une valeur de pixel de l'image actuelle appartient à l'arrière-plan. Des méthodes non paramétriques de densité estimée peuvent fournir des réponses rapides à des événements à haute fréquence dans l'arrière-plan, y compris les valeurs nouvellement observées dans le modèle. Toutefois, la capacité de ces méthodes pour gérer avec succès l'évolution des événements concomitants à des vitesses différentes est contestable, car ils mettent à jour leurs modèles de pixels dans une manière première entrée, et première sortie. Cela a conduit certains auteurs à représenter les valeurs de fond avec deux séries de valeurs ou de modèles: un modèle à court terme et un modèle à long terme. Même si cela peut être une solution pratique pour certaines situations, il laisse la question ouverte de savoir comment déterminer l'intervalle de temps approprié. En termes pratiques, la manipulation de deux modèles augmente la difficulté d'affiner les valeurs des paramètres sous-jacents. [7]

I.4. Conclusion :

Nous avons abordé dans ce chapitre les notions de bases de vision et d'analyse de mouvement. Par la suite, un état de l'art a été fait en ce qui concerne la détection de mouvement dans une séquence vidéo. Il existe dans la littérature, différentes techniques ou méthodes de détection de mouvement, toutes pourvues d'avantages et d'inconvénients. Les méthodes de détection de mouvement basées sur la soustraction de fond (bien précisément celles des GMMs) ont donné de bons résultats.

Chapitre II

La reconnaissance de mouvement dans une séquence vidéo

II.1. Introduction :

La reconnaissance de l'homme et la compréhension de ses actions sont devenues des sujets très populaires dans le domaine de la vision par ordinateur et le domaine du traitement d'images. Un grand nombre d'applications de reconnaissance à partir des vidéos peut être trouvé : la vidéo-surveillance, l'interaction homme-machine et l'indexation des vidéos.

Beaucoup de techniques et de méthodes sont développées dans le but d'étudier et d'analyser le mouvement. De nos jours, les techniques sont de plus en plus robustes grâce à la puissance croissante de la technologie et des machines qui ont permis d'évaluer des performances surtout dans le domaine de la vidéo-surveillance. Un progrès très important est réalisé, pour lequel l'analyse du mouvement est faite sur des séquences vidéo. Cependant, la quantité de mouvements et des gestes recueillis en une seule minute sont innombrables et par la suite difficiles à analyser. Pour cela, les outils offrant ce type d'analyse automatique en temps réels sont rares. Un autre problème réside dans le coût élevé des systèmes de capture du mouvement du fait que ces systèmes nécessitent une mise en place et un environnement assez contraignants.

Nous verrons dans ce chapitre, différents exemples de reconnaissance de mouvement concernant l'être humain ainsi que ses activités dans un flux vidéo. Nous verrons par la suite, les différentes techniques de reconnaissance largement utilisées. Et pour finir, nous verrons une méthode de reconnaissance de mouvement en détails, capable de reconnaître une personne humaine dans un flux vidéo ainsi que son identité.

II.2. Intérêts, et champs d'application de la reconnaissance de mouvement :

Nous nous intéresserons dès maintenant uniquement aux types de reconnaissance de mouvement concernant l'être humain, et plus précisément, à l'activité de ce dernier dans un environnement externe. L'intérêt majeur de la reconnaissance de mouvement, est de comprendre (ou voir) ce qui se passe dans un flux vidéo (identité et activité des individus, comportement de foules, gestes, ...) et d'essayer d'extraire les informations potentielles, nécessaires à la compréhension des différents scénarios préalables. [7]

La reconnaissance de mouvement rentre principalement dans le domaine de la vidéo-surveillance. L'objectif serait de développer un système de vidéo-surveillance intelligent capable de remplacer le système de vidéo-surveillance passif traditionnel qui se révèle inefficace lorsque le nombre de caméras dépasse la capacité de l'opérateur humain à les

surveiller. En bref, le but de la vidéo-surveillance n'est pas seulement d'installer des caméras à la place des yeux humains, mais aussi d'accomplir la tâche de surveillance tout aussi automatique que possible.

La vidéo-surveillance des scènes dynamiques dispose d'un large éventail d'applications potentielles, tels : la surveillance des activités et comportement humains dans les endroits sensibles, la détection et reconnaissance des personnes lors des contrôles d'accès, détection de chutes, et gestes, ... etc [23]

Nous pouvons classer ces applications en deux catégories de reconnaissance :

II.2.1. La surveillance et identification des activités et comportement des personnes :

L'identification des personnes à distance par un système de surveillance intelligent peut aider par exemple la police à attraper les suspects. La police peut construire une base de données caractéristique biométrique des suspects et placer des systèmes de vidéo-surveillance à des endroits où les suspects apparaissent habituellement comme par exemple, les stations de métro, les casinos, ... etc. Ces systèmes reconnaissent automatiquement et jugent si oui ou non les gens observés sont suspects. Si oui, des alarmes sont déclenchées immédiatement. Ce type de systèmes utilise des techniques biométriques. [23]

Ces systèmes peuvent aussi être utilisés pour :

- la reconnaissance lors d'un cas de vol (*Fig II.1.*) ou d'agression (*Fig II.2.*), et la reconnaissance des colis dangereux (*Fig II.3*) ;
- la reconnaissance des activités humaines (sauts d'athlètes dans [3]) et des gestes dans [4] ;

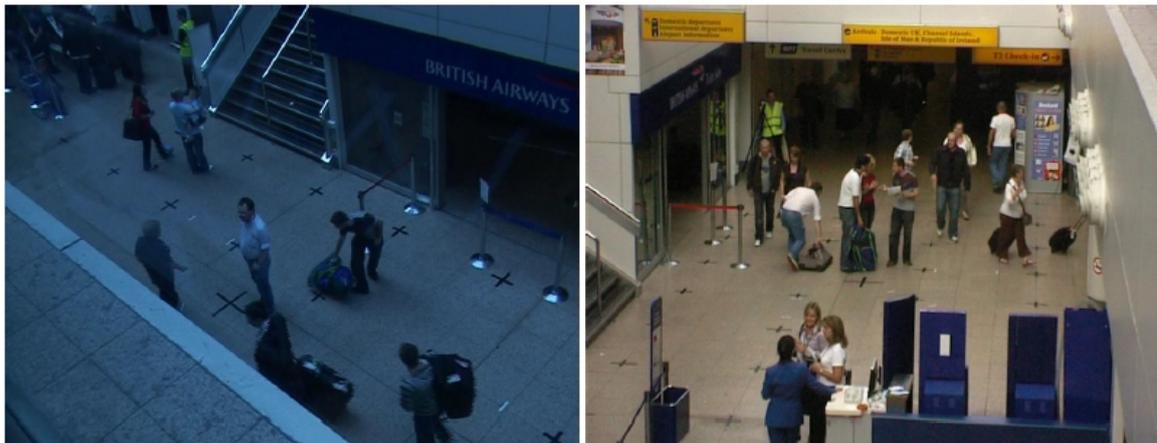
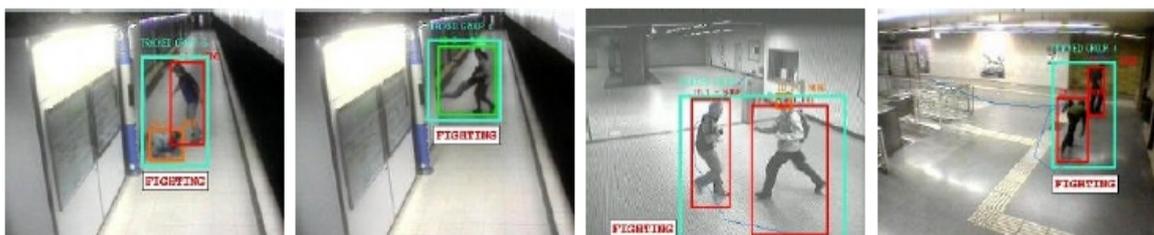


Figure II.1. : Images de vidéo-surveillance montrant des vols de bagage dans un aéroport. [1]



(a) Personne couchée au sol

(b) Variation dans la largeur du groupe

(c) Séparation de personnes dans un groupe

(d) Variation dans la trajectoire du groupe

Figure II.2. : Images de vidéo-surveillance montrant une détection de comportements dangereux liés aux bagarres. [16]



Figure II.3. : Image de vidéo-surveillance montrant un bagage abandonné susceptible de contenir une bombe. [1]

II.2.2. La détection et la reconnaissance des personnes :

Dans certains endroits sensibles en matière de la sécurité telles que les bases militaires et les unités gouvernementales importantes, seules les personnes avec une identité particulière sont autorisées à entrer. Une base de données caractéristique biométrique est construite en utilisant des techniques biométriques. Quand quelqu'un est sur le point d'entrer, le système pourrait obtenir automatiquement les caractéristiques du visiteur, comme la taille, l'apparence du visage et de sa démarche à partir d'images prises en temps réel, et de décider ensuite si le visiteur peut être autorisé à entrer. [2, 23]

La reconnaissance ou l'identification humaine ne peut être réalisée qu'à travers une variété de la biométrie, en utilisant différents types de capteurs: lecteurs d'empreintes digitales, scanners d'iris, de microphones pour la reconnaissance vocale et des caméras vidéo. Les caméras vidéo sont avantageuses parce que, d'une part elles ne sont pas intrusives, et d'autre part les sujets peuvent être filmés sans leur coopération.

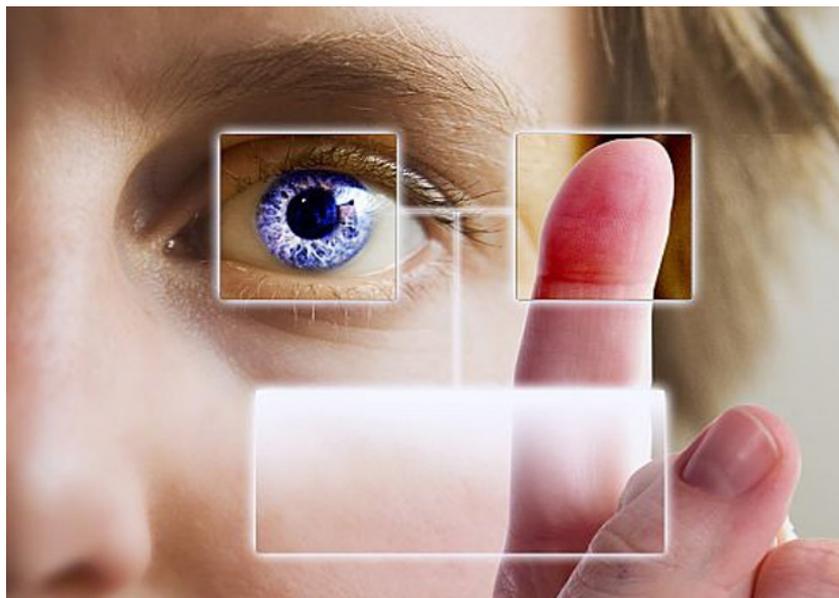


Figure II.4. : Image d'un iris et d'une empreinte.

La reconnaissance faciale par l'utilisation d'une caméra vidéo est une biométrie largement utilisée et qui a une longue histoire. Elle a été développée par Benton et Van Allen en 1968 pour évaluer la capacité d'identification des visages non familiers. L'utilisation des techniques de reconnaissance faciale a connu un développement à grande échelle depuis le milieu des années 90, avec l'utilisation efficace de nouvelles technologies, notamment

l'ordinateur et sa capacité de traitement d'images ce qui lui a donné la possibilité de contrôler l'identité des personnes sous toutes ses formes :

- contrôle d'accès des utilisateurs aux réseaux informatiques et à l'Internet ;
- contrôle d'accès de personnels aux zones contrôlées comme les aéroports ;
- recherche de criminels et des personnes disparues ;

De plus, la reconnaissance faciale va permettre l'établissement de l'interaction entre machines et humains par la détection de l'identité, comme on l'a vu dernièrement avec le téléphone portable muni d'un système capable de reconnaître son utilisateur. [2]

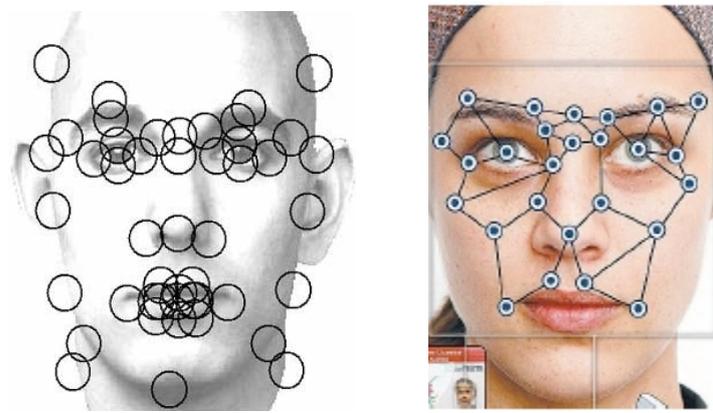


Figure II.5. : Reconnaissance de personnes en utilisant les techniques biométriques faciales. [1]

Bien que son efficacité relativement limitée soit conditionnée par la nécessité d'une image du visage de la personne, demander à une personne de coopérer peut être un problème d'autant que ne n'est pas tout le monde qui va aider le système. La reconnaissance de la démarche est une alternative viable, dans ce cas, et il n'est pas nécessaire de restreindre le champ de vision limité à l'environnement, ni à demander la coopération. La reconnaissance de la démarche n'est pas (encore) aussi efficace que la reconnaissance faciale, mais elle pourrait renforcer la décision rendue dans un système biométrique multimodal.

La démarche, comme une donnée biométrique est un sujet de discussion récent, qui a gagné en popularité. Sa robustesse contre les conditions d'imagerie pauvres la rend applicable à un large éventail de scénarios du monde réel. En outre, aucune sorte de coopération de la part des sujets n'est nécessaire. La démarche est également difficile à simuler. Pourtant, les techniques de reconnaissance de la démarche ne sont toujours pas assez précises pour utiliser la démarche comme élément biométrique unique d'un système de surveillance réelle. Ces

techniques de reconnaissance sont mieux utilisées pour renforcer une décision dans un système biométrique multimodal. [7]

Nous allons voir dans ce qui va suivre, une approche de reconnaissance de personnes utilisant leur démarche.

II.3. Détection et identification des personnes par reconnaissance frontale de leur démarche : [7]

Le travail porte sur la conception d'un système complet permettant la reconnaissance en temps réel de personnes dans des séquences vidéo acquises par une caméra fixe. En allant des images fournies par la caméra jusqu'à l'identité de la personne filmée, la méthode élaborée par les auteurs de [7] se décompose en trois étapes. La première utilisant un algorithme de détection de mouvement basé sur la soustraction de l'arrière-plan dont l'objectif est d'identifier les zones de l'image qui correspondent à des objets en mouvement. La seconde étape (description et classification des silhouettes), permet de déterminer parmi les objets détectés dans la première étape, les zones qui correspondent à des êtres humains. La troisième et dernière étape utilise un algorithme qui a pour but l'identification des personnes détectées à base de leur démarche.

II.3.1. Etape un : Détection des zones en mouvement :

La première étape du système réside dans la segmentation du flux vidéo d'entrée basée sur une méthode de soustraction de fond adaptative afin de produire des images binaires (1 pour les objets en mouvement et 0 pour les objets statiques), qui seront utilisées dans la deuxième étape.

Les méthodes de segmentation d'arrière-plan sont nombreuses. Parmi elles, la méthode basée sur une modélisation adaptative de chaque pixel en tant que mélange de Gaussiennes (*Fig II.6.*), dont chaque Gaussienne correspond à la probabilité d'observer une intensité ou une couleur particulière pour ce pixel. Dans chaque groupe de Gaussienne, la moyenne représente la couleur moyenne ou de l'intensité du pixel, alors que la variance est utilisée à des variations d'éclairage et le bruit de la caméra.

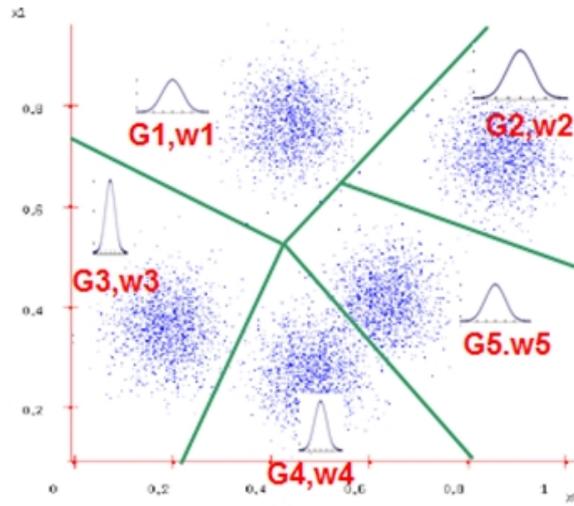


Figure. II.6 : Mélange de distributions Gaussiennes. [11]

L'algorithme général repose sur l'hypothèse que l'arrière-plan est visible plus fréquemment que le premier plan et que la variance de l'arrière-plan est relativement faible. Il s'agit d'hypothèses communes pour la plupart des techniques de soustraction d'arrière-plan. Une description détaillée de l'algorithme peut être trouvée dans [11]. La description technique est donnée ci-après.

Si X_t est la valeur de couleur ou de l'intensité observée à l'instant t pour un pixel particulier de l'image, l'historique du pixel $\{X_1, \dots, X_t\}$ est modélisé comme un mélange de K distributions Gaussiennes. La probabilité d'observer une couleur ou d'une valeur d'intensité au temps t est exprimé en :

$$P(X_t) = \sum_{i=1}^K w_{i,t} \cdot \eta(X_t, \mu_{i,t}, \Sigma_{i,t}) \quad (\text{II.1})$$

Où :

- K est le nombre de classes Gaussiennes utilisées pour modéliser l'histogramme du pixel ;
- $w_{i,t}$ est le poids associé au groupe i à l'instant t , il modélise la quantité de données représentée par la i -ème-gaussienne ;
- $\mu_{i,t}, \sigma_{i,t}$ sont la matrice moyenne et la covariance de la Gaussienne i ;
- η est une fonction de densité de probabilité d'une distribution Gaussienne, donnée par :

$$\eta(X_t, \mu_{i,t}, \Sigma_{i,t}) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \cdot e^{-\frac{1}{2}(X_t - \mu)\Sigma^{-1}(X_t - \mu)} \quad (\text{II.2})$$

Pour des raisons d'efficacité de calcul, la matrice de covariance $\sigma_{I,t}$ est supposée isotrope et diagonale.

$$\sigma_{I,t} = \omega_{I,t}^2 \mathbf{I} \quad (\text{II.3})$$

Les distributions Gaussiennes sont classées par ordre décroissant selon le rapport $\omega_{I,t}$ sur $\sigma_{I,t}$. Les b premières Gaussiennes sont considérées comme Gaussiennes de l'arrière-plan, tandis que le reste d'entre elles représente l'avant-plan. Le facteur dynamique B est estimé selon la relation ci-dessous, en accumulant les valeurs $\omega_{I,t}$ selon l'ordre de la Gaussienne calculé, jusqu'à atteindre une valeur de seuil donnée par T .

$$B = \text{argmin}_b \left(\sum_{i=1}^b \omega_{I,t} > T \right) \quad (\text{II.4})$$

Où T est un seuil de décision entre les distributions du fond et celles des objets en mouvement.

Comme indiqué précédemment, cet algorithme suppose que l'arrière-plan est visible le plus souvent par rapport au premier plan et que la variance de l'arrière-plan est relativement faible. Chaque nouvelle valeur du pixel est comparée aux K distributions jusqu'à ce qu'une correspondance soit trouvée, auquel le pixel reçoit une étiquette de classe de la distribution (de fond ou plan) selon la correspondance. Une correspondance est définie comme une valeur de pixel à l'intérieur de 2,5 fois l'écart type d'une distribution, comme montré dans cette relation :

$$SQRT \left((X_{t+1} - \mu_{i,t})^T \Sigma_{i,t}^{-1} (X_{t+1} - \mu_{i,t}) \right) > 2.5 \quad (\text{II.5})$$

Si aucune correspondance n'est trouvée, le pixel est considéré comme appartenant à l'avant-plan (*Fig II.7*). Dans ce cas, une nouvelle distribution, centrée sur le pixel de couleur ou d'intensité, est initialisée en supprimant la distribution qui représente le plus faible poids dans le modèle de mélange. Cette nouvelle répartition aura une variance initiale élevée et un poids (le plus faible d'avant).

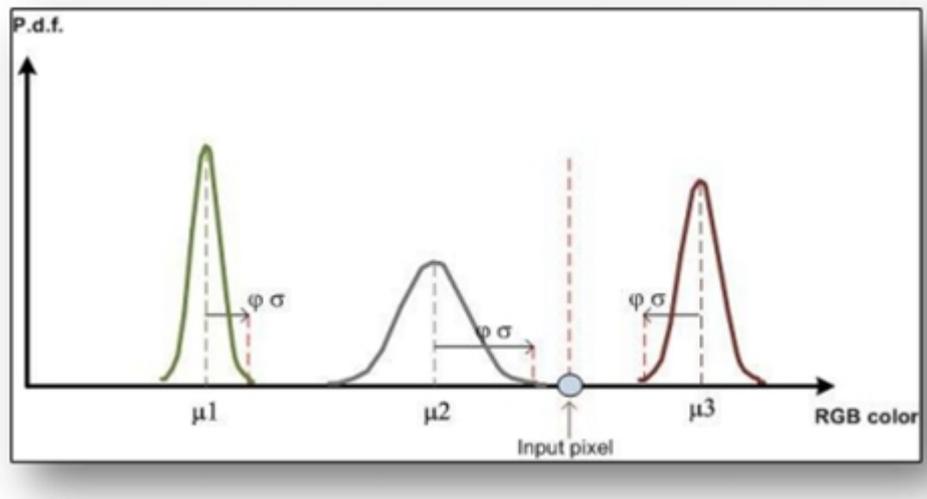


Figure II.7 : Classification des pixels faisant parti soit de l'avant ou de l'arrière-plan. [11]

Une fois la nouvelle valeur du pixel classée, le modèle doit être mis à jour. Une méthode standard consiste à utiliser l'algorithme d'espérance et maximisation (EM). Malheureusement, ce serait beaucoup trop coûteux en ressources informatiques. Dans [20], Stauffer et Grimson donnent une approximation du K-means qui est assez efficace pour être exécuté en temps réel.

Après le calcul de l'avant-plan, les pixels de l'avant-plan sont regroupés sur un algorithme de 8-composantes reliées. Ce qui garantit qu'une étiquette unique est attribuée à chaque région liée. Alors chaque région liée est considérée comme une entrée distincte pour la phase de description et classification des silhouettes. Des exemples de silhouettes candidates extraites par l'algorithme de détection par mélange de Gaussiennes sont présentés dans la figure suivante :



Figure II.8. : Exemples de silhouettes extraites par la technique de soustraction de fond basée sur le mélange de Gaussiennes. [7]

II.3.2. Etape deux : Description et classification des silhouettes :

Dans le module de description des silhouettes, chaque silhouette candidate sera traitée comme si elle était l'unique région dans l'image. Il y'a donc autant de silhouettes que de régions liées pour lesquelles un algorithme doit décider si oui ou non la silhouette appartient à un modèle de forme connue. Soit la figure suivante :

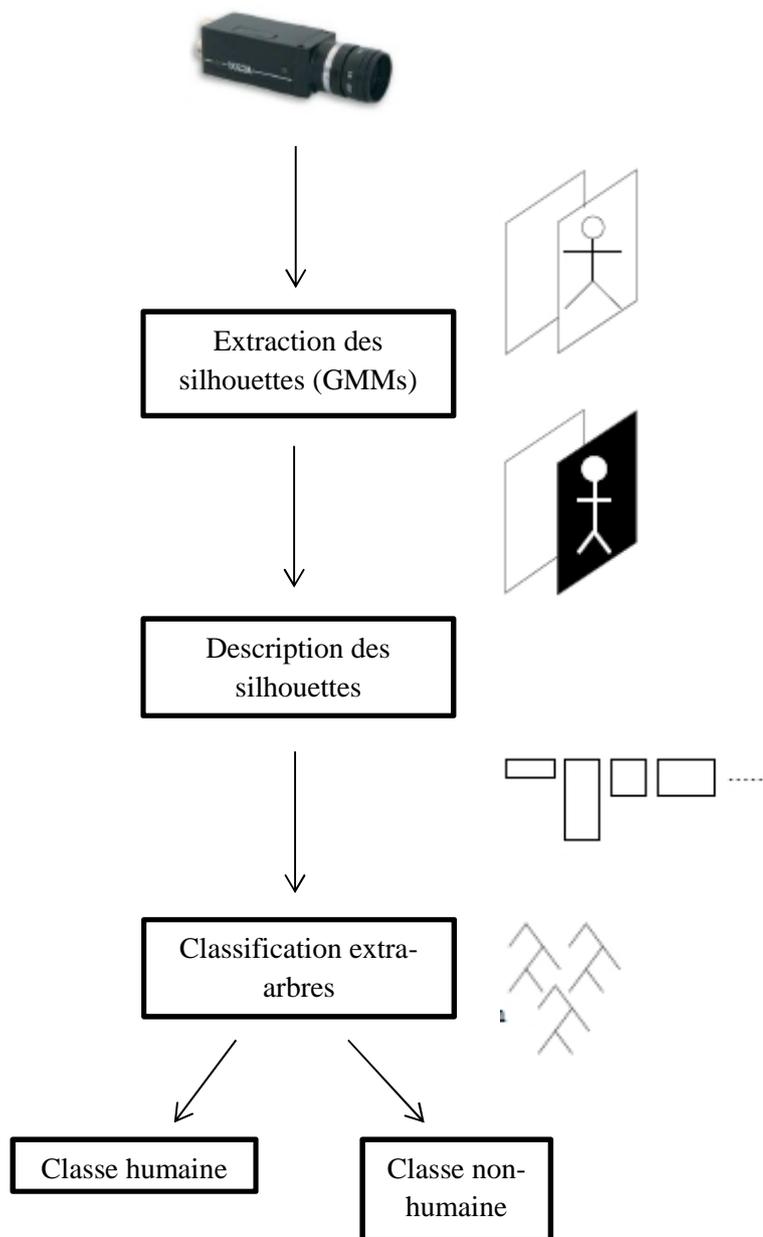


Figure II.9. : Architecture de l'algorithme de description et classification des silhouettes.

II.3.2.1. Description des silhouettes :

L'une des difficultés majeures dans la description consiste à trouver un ensemble de caractéristiques de silhouettes appropriées. La plupart des descripteurs surfaciques peuvent être décrits en termes de la théorie de la morphologie mathématique. Dans [7], les auteurs ont introduit un nouvel opérateur « granulométrique » et non morphologique pour produire un ensemble de caractéristiques décrivant les silhouettes extraites.

Le concept de la granulométrie, est basé sur la définition suivante :

Soit $\psi = (\psi_\lambda)_{\lambda \geq 0}$ est une famille de transformations d'images, dépendante du paramètre λ , ce qui s'applique le mappage suivant: $\varepsilon \rightarrow \varepsilon : x \rightarrow \psi_\lambda(x)$. Cette famille constitue une granulométrie si et seulement si les propriétés suivantes sont satisfaites :

$$\begin{aligned}
 \forall \lambda \geq 0, \quad \psi_\lambda & \text{ Augmente} \\
 \forall \lambda \geq 0, \quad \psi_\lambda & \text{ Anti-extensive} \\
 \forall \lambda \geq 0, \mu \geq 0, \quad \psi_\lambda \psi_\mu & = \psi_\mu \psi_\lambda = \psi_{\max(\lambda, \mu)}.
 \end{aligned} \tag{II.6}$$

La troisième propriété implique que, pour chaque $\lambda \geq 0$ est une transformation idempotente, qui est $\psi_\lambda \psi_\lambda = \psi_\lambda$. Comme ces propriétés reflètent ceux d'une ouverture, les ouvertures se rangent bien dans ce cadre aussi longtemps que nous pouvons commander les ouvertures avec un scalaire. Par exemple, supposons que X or B est l'ouverture par une boule de rayon r . Alors $\psi = (\psi_r)_{r \geq 0} = (X \text{ or } B)_{r \geq 0}$ est une granulométrie. D'intérêt particulier, les granulométries sont générées par les ouvertures par des versions mises à l'échelle d'un élément structurant convexe.

Plusieurs mesures utiles existent pour les granulométries définies par un facteur d'échelle unique: la distribution de la taille et de la configuration du spectre. La distribution de la taille est une courbe qui donne la probabilité d'un point appartenant à un objet de rester dans cet objet après l'ouverture par rapport à un facteur de taille. Le spectre d'aspect est défini aussi comme la dérivée de la distribution de la taille. Toutes ces mesures sont prises sur les résidus des opérateurs entraînés par un critère unidimensionnel. Ils ne sont ni applicables à une famille d'éléments structurants arbitraires, ni en mesure de produire plusieurs caractéristiques dimensionnelles non corrélés. Par conséquent, un nouvel opérateur est défini qui produit une couverture :

Soit S une famille de I éléments structurants arbitraires $S = \{S^{i \leq I}\}$. La couverture d'un ensemble X de S est défini comme l'union des éléments traduits de S qui sont inclus dans X tel que :

$$C(X) = \{S_z^j \mid z \in \varepsilon \text{ et } S^j \subset X\} \quad (\text{II.7})$$

Une ouverture à deux dimensions est préférable et qui est d'un intérêt pratique et maniable pratiquement : une ouverture par un élément structurant rectangulaire $B = mH \cup nV$ où m, n sont respectivement la largeur horizontale et verticale à l'échelle des segments. Basée sur la famille B de toutes les tailles des rectangles possibles, $C(X)$ sera l'union de tous les rectangles maximaux inclus dans X . Ces rectangles sont illustrés dans la figure II.9.

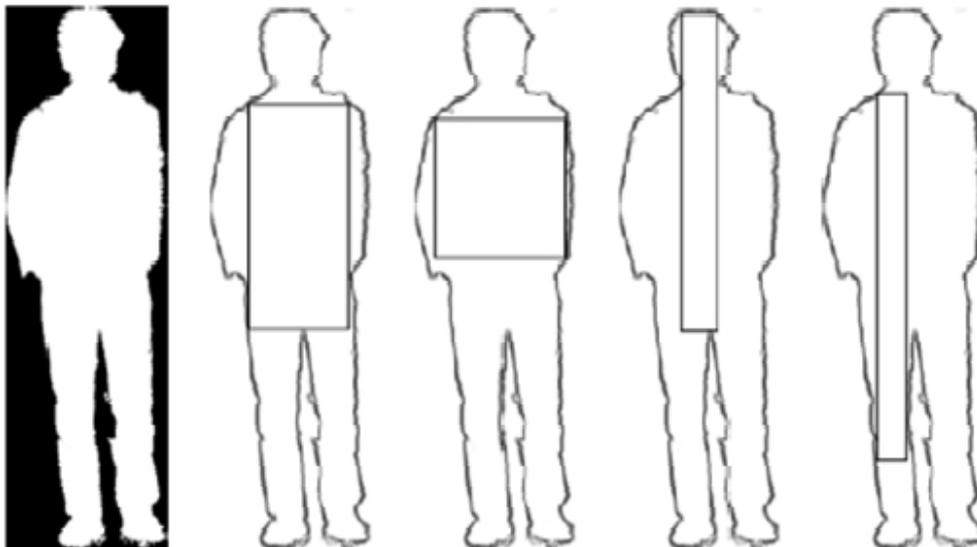


Figure II.10. : Exemple de rectangles maximaux coincés à l'intérieur d'une silhouette. [7]

L'avantage principal de l'utilisation d'une couverture, c'est que nous avons une famille d'éléments structurants, décrivant la surface d'un objet X , dont les membres peuvent se chevaucher, mais uniquement chacun d'entre eux se tiennent quelque part dans X .

L'étape suivante consiste à extraire les caractéristiques de la couverture, telles : la largeur, la hauteur, le périmètre de tous les rectangles. Toutefois, à des fins de classification, des précautions doivent être prises pour éviter des caractéristiques redondantes. Pour cela, les positions des éléments de $C(X)$ ont été prises par rapport au centre de X . Et enfin, 6 caractéristiques ont été sélectionnées sur tout élément de $C(X)$:

- 2 coordonnées du centre du rectangle, calculé par rapport au barycentre de la silhouette ;
- largeur du rectangle, qui est normalisée par rapport à la largeur de la zone de délimitation de la silhouette ;
- la hauteur du rectangle, qui est normalisée par rapport à la hauteur de la zone de délimitation de la silhouette ;
- l'aire du rectangle, qui est normalisé par rapport à la surface de l'ensemble de la silhouette ;
- le pourcentage de la surface du rectangle considéré des pixels qui sont couverts exclusivement par le rectangle considéré ;



Figure II.11. : Exemples de distributions de dimension rectangles pour une forme humaine. [7]

II.3.2.2. Classification des silhouettes :

Une fois que l'ensemble de toutes les caractéristiques décrivant une silhouette ont été extraites (voir la figure II.10), il devient possible d'exploiter un algorithme d'apprentissage automatique pour cartographier cet ensemble dans une classe. En effet, de telles correspondances sont particulièrement difficiles à obtenir à la main et doivent être apprises par le système. Comme nous nous intéressons à la détection des silhouettes humaines, seules deux classes d'intérêt sont prises en compte: la classe des silhouettes humaines, et la classe de toute autre silhouette. Un tel algorithme est formé sur un ensemble de données, prises à partir de la base de données COIL-100 [24], qui contiennent les deux silhouettes humaines et les

silhouettes non-humaines. Ces données alimenteront l'ensemble d'apprentissage avec un grand nombre de cas pour chacune de ces deux classes.

L'approche d'apprentissage automatique nécessite de prendre en compte deux difficultés :

- il existe un besoin d'un classificateur qui a d'excellentes capacités de généralisation et qui n'est pas soumis à un sur-ajustement ;
- il faut définir un moyen d'appliquer ce classificateur sur un ensemble de rectangles, dont le nombre peut varier considérablement entre les silhouettes ;

Dans [7], les auteurs ont proposé d'utiliser les arbres extra-aléatoires (extra-arbres). L'algorithme extra-arbres est un moyen rapide, précis et polyvalent. Les raisons de son utilisation sont :

- qu'il a prouvé son efficacité pour résoudre certaines tâches de classification d'images en couleur ;
- qu'il forme une architecture non-paramétrique de fonctions d'approximations, qui n'a pas besoin de connaissances préalables ;
- qu'il a un faible biais, une faible variance et de bonnes performances en généralisation ;

Nous allons décrire dans ce qui va suivre, comment extra-arbres peut être utilisé pour mapper un simple rectangle à une classe. Nous expliquerons ensuite comment mapper un ensemble de rectangles à une classe.

Extra-arbres se compose d'une forêt de N arbres de décisions binaires indépendantes. Ces arbres sont construits avec un algorithme d'induction hautement (extra) aléatoire. Chacun de leurs nœuds internes est marqué par un seuil sur l'un des attributs d'entrée (les 6 caractéristiques des silhouettes décrites précédemment), qui doit être testé dans ce nœud. Quant aux feuilles, elles sont marquées par les étiquettes de la classification de sortie. Pour classer un rectangle avec un modèle extra-arbres, ce rectangle est indépendamment classé par chaque arbre. Ceci est réalisé en commençant par le nœud racine, puis en progressant vers le bas de l'arbre en fonction du résultat des tests sur le seuil trouvé au cours de la descente, jusqu'à atteindre une feuille. Ce faisant, chaque arbre vote pour une classe. Enfin, la classe qui obtient la majorité des votes est affectée au rectangle.

Nous venons de décrire le processus de classification d'un rectangle. Mais nous décrivons une silhouette X par sa couverture $C(X)$, ce qui est un ensemble de rectangles. Par ailleurs, deux silhouettes distinctes peuvent avoir un nombre différent de rectangles dans leur couverture. Nous devons donc mettre en place une métarègle sur l'extra-arbres pour cartographier un ensemble $C(X)$ à une classe.

Soit M un entier positif fixé. Compte tenu de l'ensemble $C(X)$ de rectangles qui façonnent la silhouette X , nous sélectionnons M rectangles à l'intérieur de cet ensemble, ce qui induit un sous ensemble $C_M(X) \subset C(X)$. Ensuite, nous appliquons le modèle extra-arbres sur chaque rectangle à l'intérieur de $C_M(X)$. Ce processus génère un vote par rectangle. Enfin, la silhouette X est affectée à la classe qui a obtenu la majorité des votes, ou un rapport suffisant de votes. Ce ratio minimum peut être choisi pour être supérieur à l'inverse du nombre de classes, si un niveau de confiance élevé est requis.

Les deux figures suivantes, montrent des exemples de classification à la fois positive et négative :

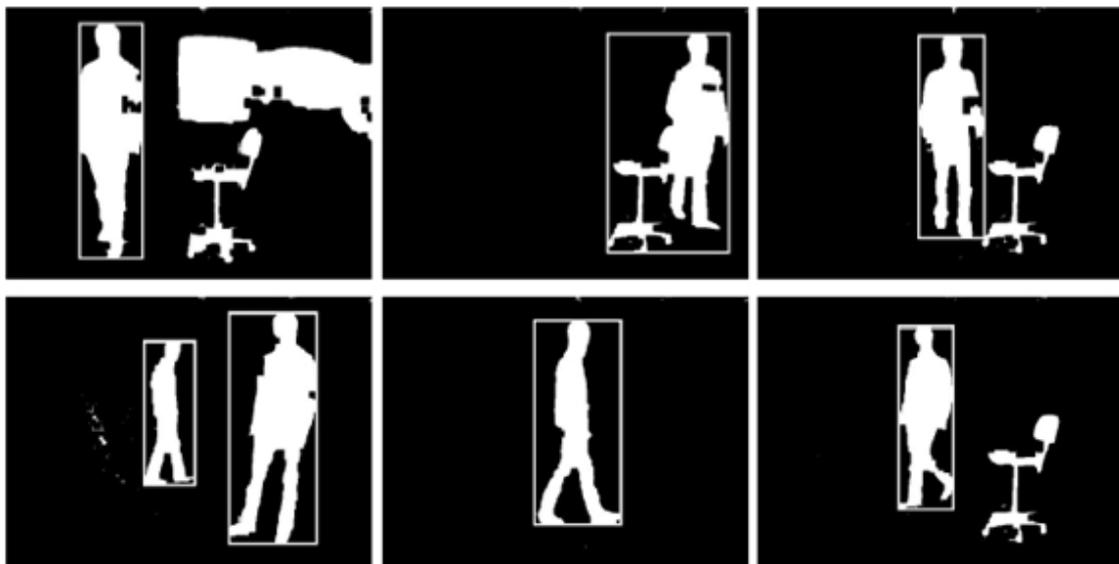


Figure II.12. : Exemple de classification positive (le rectangle en blanc, indique que le système à classer la forme détectée comme silhouette humaine). [7]



Figure II.13. : Exemple de classification négative (dans ce cas de figure, soit le système n'a pu classer aucune forme d'objets détectés, ou bien la forme classée ne correspond pas à une silhouette humaine).

[7]

II.3.3. Etape trois : Identification des personnes détectées à base de leur démarche :

Les techniques de reconnaissance de la démarche sont généralement classées en deux catégories : les approches à base de modèles et les approches globales. [7]

L'approche à base de modèles fait usage des modèles explicites de la démarche dont les paramètres doivent être estimés par traitement de la séquence d'images. L'identification est effectuée uniquement sur la base des valeurs estimées du modèle explicite. Les approches à base de modèles sont généralement invariantes en échelle, dans la mesure où l'estimation des paramètres est possible compte tenu de la configuration d'images. Ceci est un avantage majeur, car les conditions d'apprentissage sont susceptibles de différer des conditions d'utilisation pratique. D'autre part, ces méthodes ont souvent besoin d'images en haute définition afin de fonctionner correctement. Ils présentent aussi un coût sensiblement plus élevé en terme de calcul. Nous pouvons trouver dans cette catégorie :

- une modélisation articulée de la silhouette ;
- une modélisation des cuisses comme une paire de lignes épaisses ;
- une modélisation de la silhouette d'une personne qui marche comme un groupe de sept points de suspension, ou la modélisation des jambes comme deux pendules reliées en série ;

Des approches globales n'assument aucun modèle explicite pour la démarche de l'être humain. Elles extraient des informations de la démarche directement à partir de séquences d'images. Les signatures de la démarche sont, par exemple dérivées des séries chronologiques de silhouettes binaires extraites à partir de la séquence originale avec un algorithme de

soustraction de fond. Ceci apporte une invariance adaptée aux conditions de couleur, de texture ou de l'éclairage (en supposant que l'algorithme de soustraction fond utilisé soit robuste). Nous pouvons trouver dans cette catégorie ces deux exemples d'approche globale :

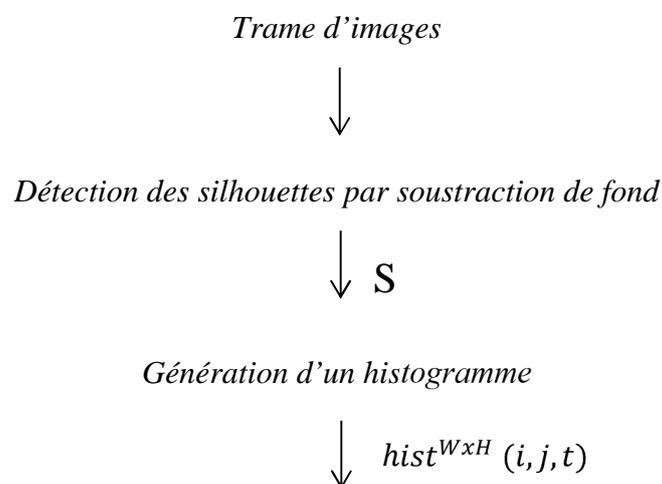
- une approche simple qui utilise des zones redimensionnées de silhouettes comme une signature de démarche est décrite ;
- une approche où les silhouettes sont moyennées sur des cycles de démarche complets pour calculer l'énergie de la démarche (GEI) ;

Les auteurs de [7], ont proposé un algorithme de reconnaissance de la démarche, capable de reconnaître les personnes à partir de trames d'images capturées en temps réel à partir de caméras de surveillance situées dans des couloirs. Contrairement aux nombreuses techniques présentes dans la littérature qui traitent des séquences complètes de démarche, leur algorithme identifie une personne déjà connue (c'est-à-dire, une personne présente dans leur base de données) dès qu'il obtient un cycle de marche complet, ce qui représente environ 1 seconde ou 25 images (si la fréquence d'images est de 25 images/seconde). Comme montré de la *fig .II.14*, cet algorithme comporte trois étapes :

1- l'extraction d'une silhouette par une technique de soustraction d'arrière-plan à l'instant t (proposée dans la section II.3.1) ;

2- le calcul d'un histogramme à l'instant t , qui est utilisé pour mettre à jour la signature de la démarche ;

3- la classification de la signature de la démarche par un algorithme d'apprentissage automatique qui délivre l'identité de l'une des personnes connus du système ;



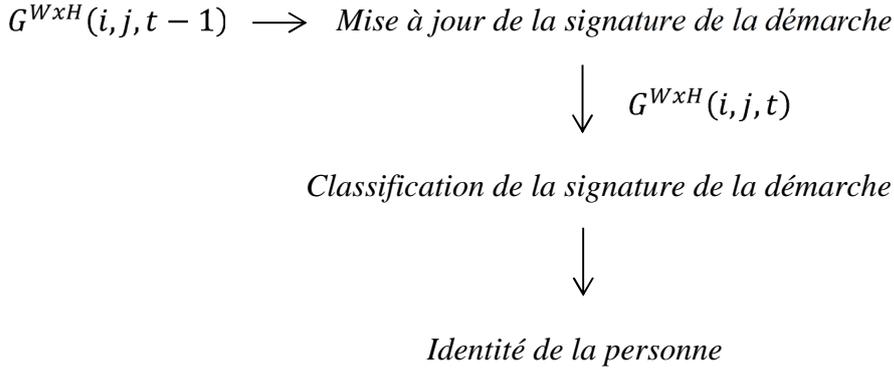


Figure II.14. : Etapes de l'algorithme de reconnaissance de la démarche. [7]

Avec :

$$\mathbf{hist}^{WxH}(i, j) = \frac{1}{\alpha} \#\{\mathbf{R}_d | \mathbf{w}_d \in \mathbf{B}^W(i), \mathbf{h}_d \in \mathbf{B}^H(j)\} \quad (\text{II.8})$$

$$\mathbf{G}^{WxH}(i, j, t) = \{\mathbf{hist}^{WxH}(i, j, 1), \dots, \mathbf{hist}^{WxH}(i, j, t - 1), \mathbf{hist}^{WxH}(i, j, t)\} \quad (\text{II.9})$$

Afin d'identifier une personne qui marche, une série chronologique $\mathbf{hist}^{WxH}(i, j, t)$ de ses silhouettes sont extraites des images vidéo brutes, à raison d'une silhouette par image. Pour chaque trame, la silhouette est convertie en un ensemble de caractéristiques (proposée dans la section II.3.2.), qui sont utilisées pour mettre à jour la signature de la démarche $G^{WxH}(i, j, t - 1)$. La signature de la démarche est introduite dans un classificateur qui va afficher le label de la classe correspondante à une personne en particulier.

La signature de la démarche obtenue au temps t est l'ensemble des fonctionnalités utilisé pour la reconnaissance. Il n'y a pas de difficulté particulière associée à l'élaboration d'une signature de la démarche pour l'étiqueter à une classe, ne serait-ce qu'elle doit être rapide, souple et précise. Un autre critère pour le classificateur est sa capacité à gérer des ensembles de caractéristiques ayant une dimensionnalité élevée. Pour cela, l'algorithme extra-arbres a été choisi pour cette en raison de sa capacité à gérer les espaces caractéristiques de haute dimensionnalité (proposée dans la section II.3.2.2).

II.4. Conclusion :

L'analyse des mouvements en temps réel nécessite une précision importante. Les algorithmes de traitement d'images manquent d'efficacité pour reconnaître et analyser le mouvement. Les études sur la connaissance du mouvement utilisant des systèmes à base d'apprentissage, sont focalisées sur le mouvement des parties du corps sans s'intéresser à sa

globalité. Nous avons présenté dans ce chapitre, une méthode de reconnaissance de mouvement basée sur la démarche. Cette méthode a prouvé son efficacité, car :

- le traitement et l'analyse des données se fait en temps réel ;
- l'algorithme de soustraction de fond est robuste, et ;
- l'approche de reconnaissance à base de la démarche est solide et ne nécessite aucune coopération des sujets à identifier ;

Conclusion générale

Dans ce manuscrit, une introduction a été faite dans le domaine de la capture du mouvement d'un corps humain, et plus particulièrement dans la description du comportement ou de l'activité de ce corps en mouvement dans un milieu extérieur. Par la suite, une étude détaillée d'une méthode de reconnaissance a été faite dans le domaine de la détection et surtout l'identification d'un corps humain en mouvement. Cette méthode basée sur la démarche, consiste après la détection et la classification des corps humains, à déterminer un ensemble de caractéristiques spécifiques à la démarche de chaque individu pour permettre ensuite au système d'identifier l'individu détecté.

Au cours de ces dernières années, il est devenu irréaliste de traiter manuellement ou même visuellement la quantité gigantesque d'informations recueillies par les caméras de surveillance. Il faudrait donc un système de vidéo-surveillance robuste et rapide capable de supporter une telle quantité d'informations et de les traiter automatiquement. Le système étudié, permet justement de traiter et d'analyser cette quantité d'informations automatiquement et en temps réel.

Enfin, nous espérons que ce modeste travail contribuera si peu soit-il à la réalisation des systèmes de télésurveillance et aidera les étudiants à mieux comprendre ce vaste domaine.

Liste des références

- [1] : C. WOLF, Analyse de séquence vidéo, Liris CNRS, INSA Lyon.
- [2] : R. HAMZAOUI, Reconnaissance des personnes par le visage dans des séquences vidéo, Mémoire de fin d'étude à l'Institut Supérieur des Sciences Appliquées et des Technologies de Sousse, 2010-2011.
- [3] : E. RAMASSO, D. PELLERIN et M. ROMBAUT, Reconnaissance de sauts d'athlète dans les vidéos : fusion du mouvement de caméra et de la trajectoire de points caractéristiques, Laboratoire des Images et des Signaux, Grenoble, France.
- [4] : A. DESLANDES, Reconnaissance du geste humain par vision artificielle : application à la langue des signes, Rapport de stage dans le cadre du DEA IARFA, Institut National des Télécommunications, 2002.
- [5] : H. TABIA, M. GOUIFFES et L. LACASSAGNE, Reconnaissance des activités humaines à partir des vecteurs de mouvement quantifiés, IEF, Institut d'Electronique Fondamentale, France.
- [6] : Web, <http://dailygeekshow.com/2012/08/01/des-chercheurs-ont-cree-une-technologie-capable-dimiter-liris-de-loeil-humain/>
- [7] : O. Barnich, Motion detection and human recognition in video sequences, Thèse de doctorat en Sciences de l'Ingénieur (électricité et électronique) à l'Université de Liège, Faculté des Sciences Appliquées, 2009-2010.
- [8] : Y.S. KORI & L. HOCINI, Détection de mouvement dans une séquence vidéo par une méthode multimodale, Mémoire de fin d'études, ESI ex : INI, 2009.
- [9] : Y. Benezeth et al, Review and Evaluation of Commonly Implemented Background Subtraction Algorithms, Institut PRISME, Université d'Orléans, France.
- [10] : L. NIZAR & L.M. SEDDIK, Détection d'un Mouvement dans une Séquence Vidéo par Filtres Morphologiques, Mémoire d'ingénieur en électronique ENP, 2009.
- [11] : Y. HAMMOUCHE & B. YADOUGHI, Etude, et simulation de la GMM adaptative pour la détection de mouvement, Mémoire d'ingénieur en électronique ENP, 2012.
- [12] : B. YADOUGHI, Suivi d'objets en mouvement dans les séquence vidéo, Master en électronique ENP, 2012.
- [13] : R. J. RADKE et al, Image Change Detection Algorithms : A Systematic Survey, Department of Electrical, Computer, and Systems Engineering Rensselaer Polytechnic Institute, USA.
- [14] : A. M. M^CIVOR, Background Subtraction Techniques, Reveal Ltd, PO Box 128-221, Remuera, Auckland, New Zealand.

- [16] : M. SEKI et al, Background Subtraction based on Cooccurrence of Image Variations, Advanced Technology R&D Center, Mitsubishi Electric Corporation, Hyogo, 661-8661, Japan.
- [17] : V. CEVHER et al, Compressive Sensing for Background Subtraction, Rice University, ECE, Houston TX 77005.
- [18] : H-H. LIN & J-H. CHUANG, Learning a Scene Background Model via Classification, IEEE Transactions on Signal Processing, Vol. 57, No. 5, May 2009.
- [19] : V. MAHADEVAN & N. VASCONCELOS, Background Subtraction in Highly Dynamic Scenes, Department of Electrical and Computer Engineering, University of California, San Diego.
- [20] : C. STAUFFER & W.E.L. GRIMSON. Learning patterns of activity using real-time tracking, Artificial Intelligence Laboratory Massachusetts Institute of Technology Cambridge.
- [21] : B. WHITE & M. SHAH, Automatically Tuning Background subtraction Parameters Using Particle Swarm Potimization, University of Central Florida School of Electrical Engineering and Computer Science.
- [22] : S. COHEN, Background Estimation as a Labeling Problem, Office of Technology, Adobe Systems Inc., San Jose.
- [23] : W. HU, T TAN, A Survey on Visual Surveillance of Object Motion and Behaviors, IEEE Transactions on Systems, Man, And Cybernetics - Part C : Applications and Reviews, Vol. 34, No. 3, August 2004.
- [24] : S. NENE, S. NAYAR, and H. MURASE. Columbia object image library (COIL-100), Technical report, Columbia University, February 1996.