UNIVERSITE D'AIX-MARSEILLE

MATHEMATIQUES ET INFORMATIQUE

LIS : Laboratoire d'Informatique et Systèmes

Thèse présentée pour obtenir le grade universitaire de docteur

Discipline : Informatique.

Rabah IGUERNAISSI

Comptage et suivi de personnes dans un réseau de caméras pour l'analyse comportementale.

Counting and tracking people in a cameras' network for behavioral analysis.

Soutenue le 11/12/2018 devant le jury :

M. Mohamed DAOUDI	Professeur - Institut Mines-Télécom Lille Douai	Rapporteur
M. Nicolas THOME	Professeur - Conservatoire national des arts et métiers	Rapporteur
Mme Séverine DUBUISSON	Maître de Conférences - Université d'Aix-Marseille, HDR	Examinatrice
M. Fakheddine ABABSA	Professeur - ENSAM, Arts et Métiers ParisTech	Examinateur
M. Omar AIT-AIDER	Maître de Conférences - Université Clermont Auvergne	Examinateur
M. Pierre DRAP	Chargé de Recherche - CNRS Laboratoire LIS - HDR	Directeur de thèse
M. Djamal MERAD	Professeur - Université d'Aix-Marseille	Co-directeur de thèse

Abstract

The study and the understanding of human behavior has become one of the major concerns for various reasons in different sectors of activity. This need to understand the habits of people led several big firms towards the use of videos surveillance for analyzing and interpreting behaviors. The exploitation of the huge amount of data produced by the surveillance's cameras requires the commitment and the involvement of several human operators. These reasons led to the emergence of research aimed at automating these procedures. As a result, the study of human behavior has become the main subject of several researches in the field of computer vision. Thus, a variety of behavior analysis solutions based on artificial intelligence emerged. These solutions include eye tracking, activity recognition, and counting and tracking people.

In this work, our objective is the proposal of a solution that enable the simultaneous track of several individuals in a multi-camera network in order to reconstruct their trajectories in the context of behavioral analysis. For this, we have proposed a system that is made of three main modules and a management module. The first module is a counting module based on the use of a Kinect to measure entries. This module is designed to use depth maps obtained from a ceiling kinect. The second module is a mono-camera tracking system that is based on the use of particle filtering to track individuals independently in each camera. The third module is a re-identification module which is based on the selection of salient regions for each individual. It enables the association of the individuals that are detected in the different cameras. To build this module, we proposed a new framework for the selection of salient patches (discriminant patches) for each of the tracked individuals; then, we used different matching methods to associate these patches and confirm the re-identification of individuals. The last module which is the management module is based on the use of ontologies for interpreting trajectories. This module is designed to create semantic trajectories from raw trajectories obtained previously. To do this, we have created ontologies that enable the segmentation of raw trajectories to episodes and then adding knowledge to these episodes in order to give them some sense.

Résumé

L'étude et la compréhension du comportement humain est devenue l'une des problématiques majeures dans différents secteurs d'activités. Ce besoin de comprendre les habitudes des individus a conduit plusieurs entreprises vers l'utilisation de vidéos pour l'analyse et l'interprétation des comportements. L'exploitation d'un grand nombre de données produites par des caméras de surveillance nécessite l'engagement et l'implication de plusieurs opérateurs humains. Ces raisons ont conduit à l'émergence de travaux de recherches qui ont pour objectif l'automatisation de ces procédures. De ce fait, l'étude du comportement humain est devenue l'un des principaux sujets de recherche dans le domaine de la vision par ordinateur, et de nombreuses solutions d'analyse du comportements basées sur l'utilisation de l'intelligence artificielle ont émergé. Parmi ces solutions, nous pouvons citer le suivi du regard, la reconnaissance d'activités, le comptage, et le suivi de personnes.

Dans ce travail, notre objectif est le développement d'un système qui va permettre de suivre simultanément plusieurs individus dans un réseau multi-caméras afin de reconstruire leurs trajectoires dans le contexte de l'analyse comportementale. Pour cela, nous avons proposé un système de suivi qui est composé de trois modules principaux et d'un module de gestion. Le premier est un module de comptage utilisant une Kinect pour mesurer les entrées. Ce module est conçu pour utiliser les cartes de profondeur obtenues d'une kinect installée au plafond. Le deuxième module, basé sur l'utilisation de filtres à particules, est un système de suivi mono-caméra destiné à suivre les individus indépendamment dans chacune des caméras. Le troisième module, basé sur la sélection des régions saillantes de chaque individu, sert à la ré-identification et permet d'associer les individus détectés dans les différentes caméras. Pour le construire, nous avons proposé un nouveau framework pour la sélection des patches saillants (patches discriminants) de chacun des individus suivis, et nous avons utilisé différentes méthodes d'appariement pour associer ces patches et confirmer la ré-identification des personnes. Enfin, le module de gestion est conçu pour créer des trajectoires sémantiques à partir des trajectoires brutes obtenues précédemment. Basé sur l'utilisation d'ontologies pour l'interprétation des trajectoires, il permet d'une part de segmenter les trajectoires en des épisodes, et d'autre part d'ajouter des connaissances à ces épisodes pour leur donner du sens.

Remerciements

Je remercie très chaleureusement Mr Nicolas THOME et Mr Mohamed DAOUDI pour l'honneur qu'ils m'ont accordé en acceptant d'être rapporteurs de cette thèse. Je remercie également Mme Séverine DUBUISSON, Mr Fakheddine ABABSA et Mr Omar AIT-AIDER d'avoir accepté d'être membres du jury qui a examiné ce travail.

Je souhaite remercier mes encadrants de thèse Mr Djamal Merad et Mr Pierre Drap pour m'avoir fait confiance et laissé des libertés d'initiative tout au long de la thèse et surtout pour leurs qualités humaines et j'aimerais leurs exprimer ma gratitude pour la confiance qu'ils m'ont accordés en me proposant ce sujet de thèse. J'aimerai exprimer mes remerciements en particulier à Djamal pour son attention et son écoute pendant les moments difficiles.

Un remerciement particulier à mes collègues de bureau Amine, Redouane et Raphaël, sans oublier bien sûr mes collègues et amis Lamia, Bilal, Jean-Philippe, Nawaf, Mohammed, Nicole, Florian, Manon, Joris et Nils pour les moments passés ensemble.

Je remercie aussi l'ensemble des stagiaires, doctorants, chercheurs, enseignants, personnels que j'ai côtoyé au sein du laboratoire LIS, en particulier : Matthew, Anna, Yannick, Yanal et Corinne, et ceux que j'ai côtoyé à l'IUT : Jean-Luc, Éric, Ivan, Delphine, Valérie et Ali, merci pour votre soutien et votre présence à mes côtés.

Je souhaite remercie mes deux très chers amis Ahmed et Soufiane qui ont toujours su me remonter le moral.

Je pense particulièrement à remercier Yann pour son aide et surtout sa participation à ce travail durant son stage (*ton aide m'a été très précieuse*).

Une pensée chaleureuse à mes parents, mes frères et mes sœurs.

Table des matières

ABSTRACT **2**

RESUME **3**

REMERCIEMENTS **4**

INTRODUCTION GENERALE **7**

CHAPITRE 1. ETAT DE L'ART ET DESCRIPTION GENERAL DE NOTRE METHODE. **11**

1.1. INTRODUCTION 11
1.2. ETAT DE L'ART 13
1.2.1. Approches centralisées *14*
1.2.1.1. Localisation sur le plan de sol 17
1.2.1.2. Reconstruction 3D 19
1.2.2. Approches distribuées *22*
1.2.2.1. Suivi mono-caméra de personnes 25
1.2.2.2. Ré-identification 25
1.2.2.3. Méthodes directes 26
1.2.2.4. Méthodes par apprentissage 29
1.2.3. Utilisation des différentes approches *31*
1.3. DESCRIPTION DE LA METHODOLOGIE PROPOSEE 32
1.3.1. Description générale *33*
1.3.2. Défis à relever *34*
1.3.3. Description générale des différents modules proposés *34*
1.3.3.1. Module de comptage 34
1.3.3.2. Module de suivi mono-camera 35
1.3.3.3. Module de Ré-identification 36
1.3.3.4. Module de gestion 37
1.4. CONCLUSION 37

CHAPITRE 2. COMPTAGE ET SUIVI MONO-CAMERA DE PERSONNES **39**

2.1. INTRODUCTION 39
2.2. COMPTAGE DE PERSONNES 40
2.2.1. État de l'art *41*
2.2.1.1. Comptage par rapport à une ligne d'intérêt (LOI counting) 42
2.2.1.2. Comptage dans une région d'intérêt (ROI counting) 43
2.2.2. Description de la méthodologie propose *44*
2.2.2.1. Détection de personnes 46
2.2.2.2. Suivi des personnes 51
2.3. SUIVI MONO-CAMERA DE PERSONNES 53
2.3.1. État de l'art *54*
2.3.2. Stratégie adoptée *57*
2.3.2.1. Détection des personnes 58
2.3.2.2. Suivi des personnes 61
2.3.2.3. Association des segments de trajectoires 63
2.4. EVALUATION 64
2.4.1. Comptage de personnes *64*
2.4.2. Suivi mono-caméra de personnes *68*
2.5. CONCLUSION 74

CHAPITRE 3. RE-IDENTIFICATION DE PERSONNES **76**

3.1. INTRODUCTION 76
3.2. RE-IDENTIFICATION DE PERSONNES 77
3.2.1. Ré-identification avec une seule caméra 78
3.2.2. Ré-identification dans un réseau de caméras 79
3.3. ÉTAT DE L'ART 79
3.3.1. Méthodes basées sur la signature discriminante 79
3.3.2. Méthodes basées sur la transformé des vecteurs caractéristiques 83
3.3.3. Méthodes basées sur l'apprentissage de métriques 85
3.4. DESCRIPTION DE LA METHODOLOGIE PROPOSEE 86
3.4.1. Sélection de patches discriminants 88
3.4.1.1. Introduction du framework utilisé 89
3.4.1.2. Sélection de patches 95
3.4.2. Appariement 97
3.4.2.1. Appariement direct 98
3.4.2.2. Appariement par apprentissage 100
3.5. EVALUATION 107
3.5.1. Bases de données utilisées 108
3.5.2. Dans le cadre d'un réseau de caméras 110
3.5.2.1. Résultats 112
3.5.2.2. Discussion 122
3.5.3. Gestion des occultations dans le champ de vision d'une seule caméra 124
3.5.3.1. Résultats 124
3.5.3.2. Discussion 129
3.6. CONCLUSION 130

CHAPITRE 4. MODELISATION SEMANTIQUE DES TRAJECTOIRES 131

4.1. INTRODUCTION 131
4.2. TRAJECTOIRES SEMANTIQUES 132
4.3. ÉTAT DE L'ART 133
4.4. DESCRIPTION DE LA METHODOLOGIE PROPOSEE 135
4.4.1. Modélisation de l'environnement 138
4.4.2. Modélisation des trajectoires 140
4.4.3. Modélisation des observations dans l'environnement 143
4.5. CONCLUSION 144

CONCLUSION GENERALE ET PERSPECTIVES 146

COMPTAGE ET SUIVI MONO-CAMERA DE PERSONNES 147
RE-IDENTIFICATION DE PERSONNES 148
REPRESENTATION SEMANTIQUE DES TRAJECTOIRES 148

ARTICLES 150

RÉFÉRENCES 151

Introduction générale

Avec l'arrivée du digital, les points de vente connaissent aujourd'hui une double révolution. La première concerne les parcours clients qui doivent maintenant être conçus de façon hybride en incorporant en magasin des éléments digitaux [1]. La deuxième, par les nombreuses self-service technologies disponibles, impacte le client qui voit ses activités monter en puissance, puisqu'en toute autonomie, il peut scanner ses produits, utiliser des caisses automatiques, comparer des produits sur son téléphone via des applications et ceci sans solliciter le personnel en contact [2].

Cette double révolution a radicalement modifié les parcours et trajectoires des clients en magasin, ce qui constitue une source importante de questions de recherche en marketing [3]. Il devient ainsi fondamental de mieux comprendre et d'expliquer les trajectoires et comportements des clients dans ces nouveaux contextes. Pour ce faire, les chercheurs et distributeurs se doivent de développer des méthodes performantes de collecte, de codage et d'analyse des comportements. A ce jour, les méthodes existantes se divisent en deux grandes familles. Les premières, où le client déclare son parcours à travers des questionnaires, sont remises en cause car fondées uniquement sur des éléments déclaratifs pouvant être incomplets ou subjectifs. Les deuxièmes, qui reposent sur les techniques d'observation par opérateurs ou caméras [4], sont à priori plus objectives mais souffrent cependant de carences importantes :

- L'analyse des comportements via des enregistrements nécessite au préalable une autorisation de la CNIL pour l'acquisition et la conservation temporaire de données. Ces procédures peuvent s'avérer lourdes, et le fait même d'être enregistré est généralement mal accepté par le consommateur.
- L'utilisation d'opérateurs humains est parfois source d'incomplétude et d'incohérence. Aux erreurs humaines peut s'ajouter la subjectivité de l'analyse, rendant plus difficile la comparaison d'informations relevées par différents

opérateurs. Par ailleurs, l'observation réalisée par opérateur mobilise un grand nombre de personnes pour réaliser des analyses quantitatives.

- L'échantillonnage temporel et non exhaustif des vidéos ne permet pas de capitaliser un grand nombre de comportements sur des intervalles de temps réduits.
- Le coût qui, entre autres conséquences, limite les études dans le temps et ne permet pas de les actualiser régulièrement.

Afin de palier à ces carences, et notamment pour améliorer et/ou automatiser la collecte, le codage et les analyses, de nouvelles solutions d'analyse comportementale, basées sur la vision par ordinateur et l'apprentissage, commencent à émerger. Ces solutions relèvent majoritairement de l'analyse des trajectoires, de la classification d'expressions faciales, de la reconnaissance des gestes et/ou du suivi de personnes [5-7]. L'un des sujets les plus étudiés dans ce domaine est le suivi des personnes dans un réseau de caméras. Cette problématique peut être considérée comme l'étape de base dans la conception de plusieurs applications dans ce domaine. Pour la résoudre, deux approches principales ont été proposées dans la littérature : les approches centralisées et les approches distribuées. Les approches centralisées sont basées sur l'exploitation de la relation entre différentes caméras pour effectuer un suivi, soit par la reconstruction 3D de la scène, soit par la récupération de la carte d'occupation du plan de sol. Les approches distribuées combinent généralement un module de suivi mono-caméra conçu pour la récupération des segments de trajectoire dans le champ de vision d'une seule caméra, et une stratégie de ré-identification conçue pour associer les individus suivis à travers les différentes caméras. Pour cela, la conception de stratégies de ré-identification adéquates est une tâche très importante dans ce genre d'approche.

Notre travail de recherche s'inscrit dans le cadre du projet MAGNUM (Mesure Analyse Gestion de flux Nativement Unifiée dans des Malls), financé par le Fond Unique Interministériels, dont l'objectif principal est la réalisation de solutions innovantes pour l'analyse du comportement client dans une surface de ventes. D'une manière générale, nous nous proposons de répondre à deux questions. La première est méthodologique : comment automatiser, pour les distributeurs, la collecte, le codage et l'analyse des informations ? La deuxième consiste, grâce à cette nouvelle méthodologie, à comprendre les trajectoires du client et ses antécédents dans le contexte de digitalisation des parcours et d'autonomie accrue du client.
Pour cela, nous avons développé un système permettant de suivre simultanément plusieurs individus dans un réseau multi-caméras afin de reconstruire leurs trajectoires

dans le contexte de l'analyse comportementale. Nous avons choisi une approche distribuée, à nos yeux plus adaptés, car réduisant les restrictions sur le réseau de caméras et permettant ainsi l'utilisation des réseaux déjà installés. Notre système de suivi est composé de trois modules principaux et d'un module de gestion. Le premier de comptage est basé sur l'utilisation d'une Kinect pour mesurer les entrées. Le deuxième de suivi mono-caméra utilise du filtrage à particules, et sert à suivre les individus indépendamment dans chacune des caméras. Le troisième de ré-identification sélectionne des régions saillantes, et permet d'associer les individus détectés dans les différentes caméras. Enfin, le module de gestion s'appuie sur des ontologies pour l'interprétation des trajectoires.

Ce manuscrit est articulé autour de quatre chapitres. Le premier chapitre détaillera l'état de l'art et présentera une vue d'ensemble de notre solution pour le suivi de personnes dans un réseau de caméras.

Le deuxième chapitre sera consacré aux deux premiers modules de notre système : le module de comptage et le module de suivi mono-caméra. Le module de comptage est basé sur l'exploitation des cartes de profondeur obtenues d'un kinect installée au plafond avec une position zénithale par rapport à la direction de mouvement afin de compter le nombre de personnes qui entrent et sortent d'un espace fermé. Pour cela nous avons conçu et développé deux sous-modules : l'un pour la détection de personnes dans les cartes de profondeur et l'autre pour le suivi. Ces deux sous-modules permettent de déterminer le nombre d'individus traversant une ligne d'intérêt dans le champ de vision de la kinect. Le module de suivi mono-caméra des personnes repose sur l'utilisation d'un filtre particulaire pour reconstruire les segments de trajectoires. Les séries de tests réalisés pour valider le fonctionnement de ces deux modules seront également présentées et analysées.

Dans le troisième chapitre, nous détaillerons le module de ré-identification utilisé pour associer les individus suivis dans les différentes caméras, afin de fusionner les segments de trajectoires obtenus précédemment et reconstruire les trajectoires globales des personnes suivies durant leurs présences dans l'espace surveillé. Notre méthode de ré-identification repose sur l'exploitation des différences locales entre les individus suivis. Un nouveau framework, permettant la sélection des régions saillantes pour chacun des individus suivis, a été mis en place. Ensuite ces patches saillants sont utilisés pour associer les individus à l'aide de trois différentes approches d'appariement. Là encore, nous

présenterons et analyserons la multitude de tests effectués pour valider le fonctionnement de ce module avec les différents appariements.

Le quatrième et dernier chapitre sera consacré au développement du module de gestion. Ce module est utilisé pour l'ajout de connaissance aux trajectoires brutes pour faciliter leurs interprétations par les opérateurs humains. Une façon naturelle pour ajouter ces connaissances consiste à utiliser des ontologies. Dans notre cas, trois ontologies ont été définies : une ontologie de modélisation de l'environnement pour modéliser les parties statiques de l'espace surveillé, une ontologie de modélisation des trajectoires pour la construction des trajectoires sémantiques, et une ontologie d'observation dans l'environnement pour relier les trajectoires à l'environnement.

Ce manuscrit se terminera par une conclusion résumant les apports principaux, une synthèse des travaux réalisés, ainsi que des perspectives pour la continuité de ce travail.

Chapitre 1. Etat de l'art et description général de notre méthode.

1.1. Introduction

De nos jours, les réseaux de caméras de surveillance sont installés dans tous types d'infrastructures tels que les petits magasins, les surfaces de ventes jusqu'aux grandes infrastructures comme les aéroports et les gares. Ces réseaux de caméras génèrent une quantité faramineuse de données qui sont généralement sous exploitées et qui sont seulement utilisées dans le cadre de la surveillance. Cette quantité phénoménale de données peut être exploitée de manière plus efficace dans d'autres applications qui sont actuellement moins étudiés tel que la fouille de données et/ou l'extraction des connaissances. De nouvelles thématiques de recherches ont émergé durant les dernières années, il s'agit, généralement d'analyser le comportement des individus à partir de données issues des caméras pour des applications en marketing, sécurité, médical,…

Pour des raisons de performance le besoin de processus automatiques d'analyse pour des études de comportement devient indispensable. Les applications implémentées doivent se baser sur des algorithmes d'analyse d'images/vidéos qui sont capables d'interpréter le comportement des personnes présentes dans la scène étudiée (surveillée) sans l'intervention d'opérateurs humains.

Dans ce chapitre, nous allons nous consacrer à l'une des étapes importantes de l'analyse du comportement qui consiste à suivre des personnes dans un réseau de caméras.

La problématique de suivi de personnes a été traitée dans deux contextes. Premièrement le suivi de personnes dans le champ de vision d'une seule caméra qui consiste dans la récupération de la trajectoire de l'individu en question durant sa présence dans le champ de vision de chacune des caméras indépendamment des autres. Cette problématique est communément connue comme le suivi mono-caméra des personnes. Une illustration de cette problématique est montrée dans la Figure 1.

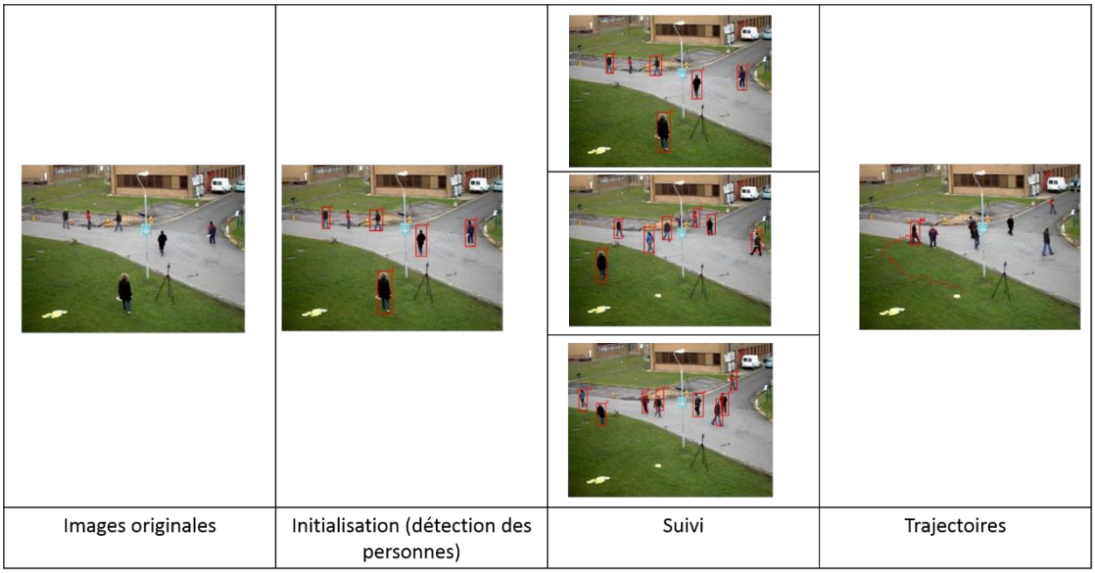


Figure 1. Illustration pour un système de suivi mono-caméra.

Une fois que les trajectoires de la personne suivie sont récupérées pour chacune des caméras du réseau, une deuxième problématique s'impose. Cette nouvelle problématique se résume dans la récupération de la trajectoire de l'individu en question mais cette fois d'une façon globale sur l'ensemble des caméras du réseau. Cette problématique est communément connue comme le suivi multi-caméras de personnes. Une illustration de cette problématique est montrée dans la *Figure 2*. Les deux types de suivis peuvent être réalisés soit en temps réel online tracking soit sur des vidéos préenregistrer offline tracking. [8]

Figure 2. Illustration pour un système de suivi multi-caméra.

Nous pouvons facilement déduire que le problème de suivi multi-caméras de personnes n'est qu'une extension du problème de suivi mono-caméra. De ce fait, la majeure partie d'algorithmes utilisés dans le cadre de suivi multi-caméras de personnes sont basés sur des méthodes déjà utilisées dans le cadre du suivi mono-caméra comme la détection de mouvement, filtrage, ou encore le flow optique.

L'utilisation d'un système multi-caméras a plusieurs avantages par rapport à l'utilisation d'un système mono-caméra. Parmi ces avantages on peut citer la gestion des occultations et de grands espaces. Cependant, le suivi multi-caméras nécessite en contrepartie des algorithmes plus complexes et des charges de calculs beaucoup plus grandes. Du fait du défi que cette problématique représente, le suivi multi-caméras de personnes est l'une des problématiques de recherche les plus abordée durant ces dernières années dans le cadre de l'analyse comportementale basée sur la vision par ordinateur [8],[9].

1.2. Etat de l'art

Le suivi multi-caméras a été fréquemment abordé pour son importance, ces larges perspectives d'utilisations ainsi que le défi scientifique qu'il relève. Plusieurs études et recherches ont été effectuées dans ce cadre et une variété d'algorithmes et de méthodes ont été proposées pour résoudre cette problématique. Ces méthodes peuvent être classées de différentes manières en se basant sur différents critères. Elles peuvent être classées par

rapport aux méthodes de détection de mouvement qui sont utilisées. De la même manière, elles peuvent être classées en se basant sur la configuration du réseau de caméras utilisé (caméras avec ou sans chevauchement). Dans cet état de l'art, nous avons choisi de classer ces méthodes en se basant sur l'étape dans laquelle la fusion de données entre les différentes caméras intervient. Avec ce classement, les travaux qui ont été effectués dans le cadre du suivi multi-caméras sont séparés en deux catégories principales qui sont les approches centralisées et les approches distribuées.

Dans les approches centralisées la fusion de données entre l'ensemble des caméras intervient avant le suivi tandis que pour les approches distribuées la fusion de données entre l'ensemble des caméras intervient après le suivi pour combiner les segments de trajectoires obtenues dans chacun des champs de vision des caméras [8-10]. La combinaison de ces deux approches nous donne une nouvelle catégorie qui est les approches décentralisées. Dans cette catégorie la fusion de données intervient avant le suivi pour chaque groupe de caméras et la fusion finale entre l'ensemble des groupes intervient après le suivi. Ce classement est illustré dans la *Figure 3*.

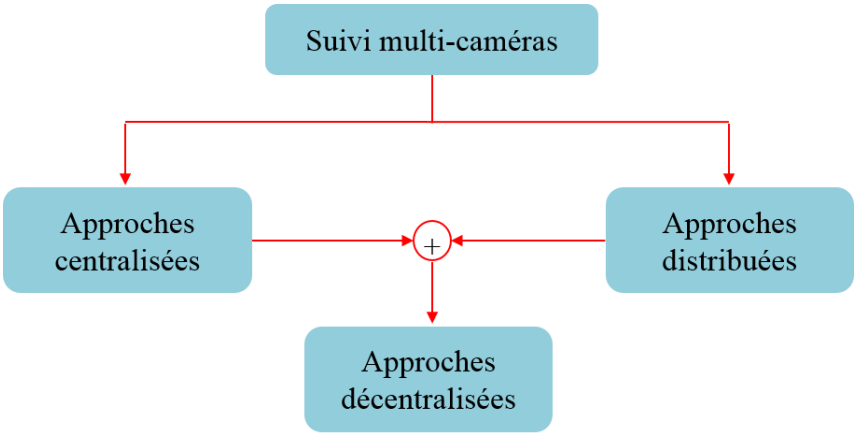


Figure 3. Les différentes approches de suivi multi-caméras.

1.2.1. Approches centralisées

Une variété de méthodes de suivi multi-caméras basées sur l'utilisation d'approche centralisée ont été proposées. Ces méthodes sont généralement utilisées dans les réseaux de caméras avec des champs de vision qui se chevauchent. Elles sont utilisées dans les scènes de foules avec beaucoup d'occultations pour la gestion des petites surfaces. Elles se basent sur la fusion d'informations qui parviennent de l'ensemble des caméras du

réseau avant le suivi pour résoudre les problèmes d'occultation et d'observations bruitées [10, 11].

D'une manière générale, toutes les méthodes basées sur l'utilisation d'une approche centralisée ont le même pipeline de base. Elles sont réalisées en plusieurs étapes qui commencent par la détection des individus présents dans l'ensemble des caméras séparément. Une fois les individus détectés, l'étape suivante consiste dans la fusion de ces détections pour localiser les individus dans la scène et finalement un suivi est effectué le suivi pour récupérer la trajectoire de chaque individu. Ce pipeline de base est illustré dans la *Figure 4*.

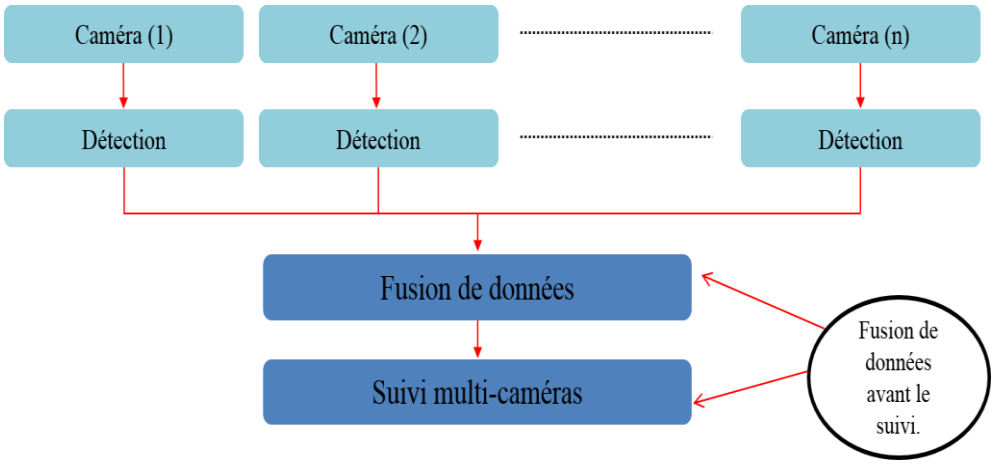


Figure 4. Le pipeline de base pour une approche centralisée.

Dans la première étape, les individus présents dans le champ de vision de chacune des caméras sont détectés à l'aide d'un algorithme classique de détection de mouvement comme le flow optique ou la soustraction de fond. Une fois la détection réalisée, l'étape suivant consiste dans la combinaison des détections enregistrées sur l'ensemble des caméras en s'assurant que les frames traités représentent le même instant, d'où la nécessité d'avoir un réseau de caméras synchronisées. Ces détections sont utilisées pour soit trouver la localisation 3D de l'individu suivi dans la scène, soit pour établir un plan d'occupation pour le plan de sol de la scène. Cette étape nécessite d'avoir la relation géométrique entre les positions de l'ensemble des caméras. Cette dernière étape est répétée à plusieurs reprises pour chaque nouvelle frame sur une certaine période ce qui permet d'enregistrer les positions des individus suivis dans le temps et de pouvoir

reconstruire leur trajectoire durant cette période. Une synthèse des méthodes utilisant une approche centralisée est présentée dans le Tableau 1.

Référence	**Vecteur de caractéristique**	**Méthode d'association**	**Calibration**
Kim et al Du et al	Binary map	Intersection de l'axe vertical central	Oui
Fleuret et al Huang et al	Binary map	Plan d'occupation	Oui
Santos et al Khan et al	Binary map	Contraintes Homographiques	Non
Liem et al Guan et al Arsic et al	Couleur	Reconstruction 3D	Oui
Chen et al	Tracklets	Optimisation de graph	Non
Kroeger et al	Couleur	SFM	Non
Yao et al	Couleur et apparence	Modèle 3D	Oui
Hofmann et al Leal-Taixe et al	Couleur	Modèle déformable et projection en 3D	Oui
Bredereck et al	Couleur HSV et positions	Positions 3D	Oui
Jiaung et al Balteiri et al	Couleur et apparence	Modèle déformable et distances en 3D	Oui
Wen et al	Apparence et continuité de mouvement	Positions 3D	Oui
Du et al	Couleur et apparence	Algorithme probabiliste	Non

Zhang and Cheng	haar-like features, HOG, et LBP	Combinaison des champs de vision	Oui

Tableau 1. Synthèse des méthodes utilisant une approche centralisée.

Les algorithmes de suivi qui sont basés sur l'utilisation d'une approche centralisée se fient à la fusion de données obtenues par l'ensemble des caméras avant le suivi. De ce fait, l'étape principale de ces algorithmes est l'étape de fusion de données qui fait la différence entre une méthode et une autre.

Parmi ces algorithmes, deux catégories principales peuvent être distinguées en se basant sur la méthode de fusion de données adoptée. La première catégorie est celle où la fusion de données est faite pour établir le plan d'occupation pour le plan de sol [12-16] et la deuxième catégorie est celle où la fusion de données est fait pour obtenir des informations sur la position 3D des individus suivis dans la scène [11, 15, 17-21].

1.2.1.1. Localisation sur le plan de sol

Les méthodes de suivi multi-caméras qui se basent sur la localisation des individus sur le plan de sol sont généralement réalisées en trois étapes principales. Dans la première étape, elles utilisent un algorithme de détection de mouvement et une segmentation adaptée pour détecter les personnes dans le champ de vision de chacune des caméras qui couvre la scène. Une fois les frames obtenues à partir des différentes caméras analysées pour la détection des personnes l'étape d'après consiste dans l'exploitation de la relation géométrique entre les positions des caméras pour obtenir la position de chaque individu détecté sur le plan de sol. La dernière étape consiste dans la répétition des deux premières étapes sur une période de temps et l'enregistrement des différentes positions occupées par chaque individu pour obtenir sa trajectoire. Ce principe de fonctionnement est illustré dans la *Figure 5*.

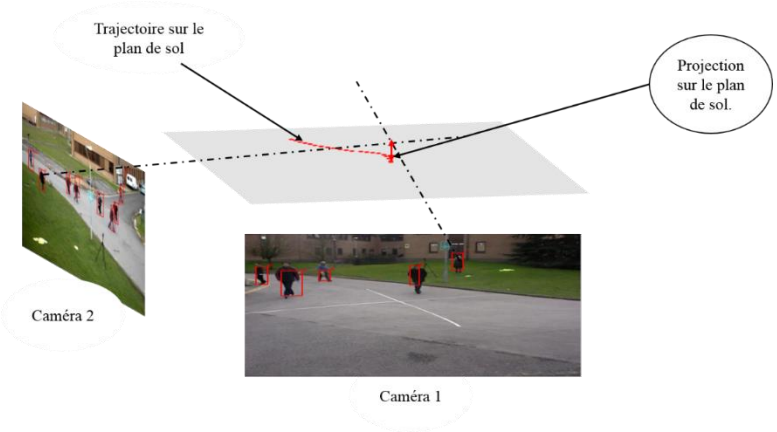


Figure 5. Localisation sur le plan de sol.

Dans cette catégorie d'algorithmes de suivi multi-cameras, on peut citer la méthode proposée par Kim et Davis [12]. Cette méthode est basée sur l'utilisation d'un algorithme standard de soustraction de fond pour obtenir le premier plan pour chacune des caméras du réseau. Les pixels du premier plan obtenus dans l'étape précédente sont classés à l'aide d'un algorithme de classification bayésienne pour obtenir des blobs qui représentent les individus présents dans la scène. Ces blobs sont combinés entre les différentes caméras pour obtenir les localisations des individus sur le plan de sol en reconstruisant la vue de dessus de la scène. Les localisations sont obtenues par l'intersection entre les axes verticaux des blobs. De la même façon, Du et Piater [22] ont utilisé l'intersection entre les axes principaux des cibles pour combiner les informations entre l'ensemble des caméras et ils ont utilisé un filtre particulaire pour suivre les individus dans le plan de sol et dans chacune des caméras.

Dans le même contexte, Fleuret et al [13] ont proposé une méthode qui est basée sur la combinaison entre un plan d'occupation probabiliste qui donne l'estimation de l'occupation du plan de sol pour chaque frame des vidéos analysées et une optimisation globale des trajectoires pour les individus suivis sur des patches de vidéos d'une centaine de frames chacun. L'estimation de plan d'occupation est faite par une méthode standard de soustraction de fonds suivi par l'utilisation d'un modèle générative qui est basé sur la couleur et un modèle de mouvement. Dans la même logique, Huang and Wang [23] ont définis une probabilité de détection pour estimer la probabilité d'avoir un objet en mouvement pour chaque position dans le plan de sol en se basant sur les images de premier plan obtenues dans les différentes caméras.

Toujours dans l'utilisation de localisation sur le plan de sol, Santos et Morinoto [14] ont proposé une méthode de suivi multi-caméras qui commence par une méthode standard de soustraction de fonds pour détecter les objets en mouvement dans la scène. Une fois les individus en mouvement détectés, l'étape suivante consiste dans le calcul de la masse supportée par chaque pixel du plan de sol pour toutes les caméras. Ce calcul de masse est utilisé pour localiser les individus dans la scène en supposant que les individus se trouvent dans les maximums locaux. Une fois détectées, ces nouvelles positions sont associées aux individus suivis dans les frames précédentes pour obtenir les trajectoires. Dans le même contexte, Khan et Shah [24] ont utilisé une homographie planaire pour obtenir les positions des individus dans le plan de sol, tandis que Eshel et Moses [25] ont choisi de suivre seulement les têtes des individus en les représentant par un seul point caractéristique qui est projeté sur le plan de sol afin d'obtenir les différentes positions des individus durant leurs présence dans la scène.

1.2.1.2. Reconstruction 3D

La deuxième catégorie de méthodes utilisant une approche centralisée est celle basée sur la reconstruction 3D soit de la scène en question soit des individus suivis. Ces méthodes sont basées sur l'utilisation d'informations 3D obtenues de l'ensemble des caméras pour localiser les individus suivis dans un environnement 3D. Pour ce faire, ces algorithmes se font généralement en trois étapes. Dans la première étape, un algorithme standard de détection de mouvement est utilisé pour détecter les personnes en mouvement dans chacune des caméras du réseau et obtenir leurs localisations dans le champ de vision de chacune de ces caméras. Une fois les positions 2D des individus suivis sont récupérées, la deuxième étape consiste dans la projection de ces positions dans un environnement 3D pour obtenir les positions 3D des individus suivis. La troisième et dernière étape consiste dans la répétition des deux premières étapes sur une période de temps pour obtenir un vecteur de position 3D qui représente la trajectoire de la personne suivie dans un environnement 3D. Ce principe de fonctionnement est illustré dans la *Figure 6*.

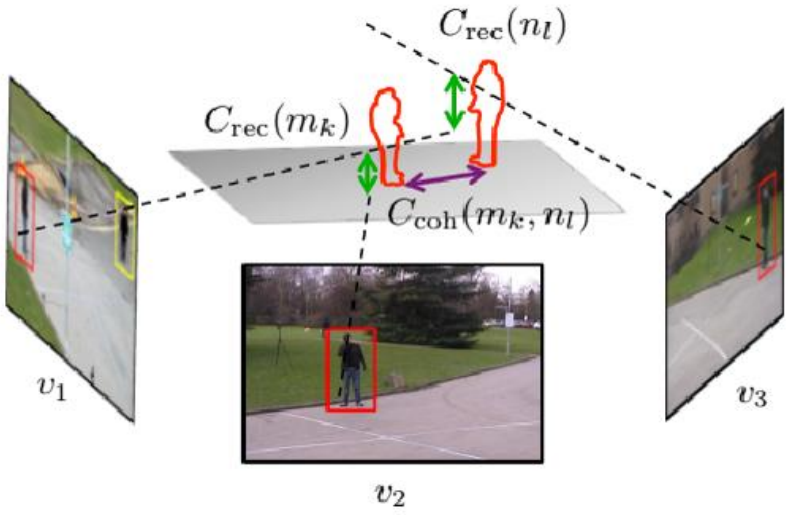


Figure 6. Utilisation d'informations 3D.

Dans ce contexte d'utilisation d'informations 3D pour le suivi multi-caméras de personnes, une variété de méthodes ont été proposées [11, 26-29]. Parmi ces méthodes on peut citer celle proposée par Liem et Gavrila [11]. Cette méthode est basée sur la reconstruction volumétrique de la scène en 3D dans un réseau de caméras avec recouvrement. L'algorithme commence par la détection des premiers plans pour chacune des caméras avec un algorithme classique de soustraction de fond. Une fois les blobs des premiers plans détectés pour l'ensemble des caméras, ils sont utilisés pour le calcul du volume occupé par chaque blob dans la scène en 3D. Ces blobs en 3D sont projetés sur le plan de sol pour estimer le nombre de personnes présentes dans chaque blob en se basant sur la surface que les projections de ces blobs occupent. À la fin, l'association de ces nouvelles détections avec les individus déjà suivis dans le système permet le recouvrement de leurs trajectoires. De la même manière, Guan et al [26] ont utilisé une méthode de reconstruction dynamique de scène en se basant sur les formes des silhouettes d'objets obtenues de leurs détection en 2D dans les champs de vision de l'ensemble des caméras. Arsic et al [27] ont utilisé une transformation homographique sur plusieurs couches pour obtenir la reconstruction de la scène en 3D, et Kroeger et al [28] ont combiné des méthodes de suivi multi-caméras par détection avec des méthodes de localisation développer pour le SFM. Cette combinaison permet d'effectuer un suivi dans des caméras dynamiques sans restriction sur le mouvement des individus suivis. Toujours dans la même logique Yao et Odobez [29] ont adopté une approche de suivi qui est basée sur la représentation des individus par des modèles dans l'environnement 3D.

Toujours dans les méthodes qui se basent sur l'utilisation d'information 3D, Hofmann et al [17] ont proposé un algorithme qui combine les problèmes de la reconstruction 3D et le suivi dans un seul framework. Cette méthode commence par la détection des individus dans chacune des caméras en utilisant le modèle déformable décrit dans [30]. Une fois ces détections 2D obtenues pour l'ensemble des caméras, elles sont ensuite projetées dans un système commun de coordonnées 3D pour avoir la reconstruction 3D des objets suivis. La liste ordonnée de ces reconstructions est considérée comme la trajectoire de l'objet en question. De la même manière, Leal-Taixe et al [31] ont combiné la reconstruction 3D et le suivi dans un seul problème d'optimisation dans lequel les détections 2D sont combinées pour avoir la localisation et le suivi dans un environnement 3D.

Jiang et al [32] ont utilisé ces informations 3D pour le suivi mais d'une autre manière. Ils ont proposé une méthode qui est basée sur l'utilisation de deux graphs. Dans le premier graph, ils imitent des hypothèses sur le plan de sol en utilisant les apparences 2D et les distances 3D des individus suivis pour extraire un ensemble de tracklets. Ensuite le deuxième graph est utilisé pour associer ces tracklets et obtenir les trajectoires globales des personnes suivis. Del-Blanco et al [18] ont proposé un algorithme basé sur l'utilisation d'une méthode standard de soustraction de fond pour détecter le objets en mouvement pour chacune des caméras, en même temps ils ont extrait les descripteurs de couleur pour ces objets. Ces informations sont par la suite utilisées dans l'environnement 3D pour estimer la densité de probabilité d'occupation volumétrique pour chaque individu dans le temps en utilisant un filtre particulaire.

En plus des deux catégories citées auparavant pour les approches centralisées. Il y a certaines méthodes hybrides comme celle proposée par Balteiri et al [15]. Dans cette méthode, les informations 3D obtenues de l'ensemble des caméras sont utilisées pour détecter les individus et leurs localisations dans le plan de sol. Pour cela, ils ont approximé les individus par des cylindres. Ensuite ils ont effectué le suivi sur une courte période de temps en exploitant les informations géométriques en 3D ainsi que les contraintes entre les caméras. Ces segments de trajectoires sont combinés par la suite pour avoir les trajectoires globales pour les personnes suivies. L'association entre les différents segments est faite avec une méthode de ré-identification basée sur l'utilisation du modèle 3D de corps présenté dans [33] qui prends en considération la couleur et l'apparence des individus. Une autre méthode qu'on peut citer ici est celle proposé par Du et Piater [34] qui se base sur un partage limité d'informations entre les caméras. Cet algorithme est basé sur l'utilisation d'un filtre particulaire pour suivre les individus dans

le champ de vision de chacune des caméras avant l'utilisation d'un algorithme de propagation de croyances (belief propagation) pour le partage d'information entre les trackeurs.

1.2.2. Approches distribuées

La deuxième grande catégorie de méthodes de suivi multi-caméras de personnes est celle des méthodes basée sur l'utilisation d'une approche distribuée. Une variété d'algorithmes a été proposée dans cette catégorie. Ces approches sont généralement utilisées pour la gestion des grandes scènes avec beaucoup de caméras qui peuvent ne pas avoir des champs de vision qui se chevauchent. Elles peuvent aussi être utilisées dans le cas où la relation géométrique entre les caméras est inconnue ou difficile à estimer. Un autre avantage non négligeable pour cette approche est le fait qu'il n'y ait pas de restriction sur la synchronisation ou la position des caméras ce qui rend cette approche plus adaptée pour les réseaux de caméras qui sont déjà installés [10, 35].

Les méthodes de suivi qui utilisent une approche distribuée se basent sur un partage limité d'informations entre les caméras. En général, ces algorithmes se font en trois étapes principales pour obtenir les trajectoires globales des individus suivis sur toute la scène surveillée. Ils commencent par la détection et le suivi de personnes dans chacune des caméras indépendamment des autres. Ensuite vient l'étape de ré-identification dans laquelle les individus suivis par l'ensemble des caméras sont associés les uns aux autres pour avoir les correspondances entre ces individus. A la fin vient l'étape où les segments de trajectoires obtenus dans la première étape sont combinés à l'aide des correspondances obtenues dans la deuxième étape pour avoir les trajectoires des personnes sur toute la scène. Ce pipeline de base est illustré dans la *Figure 7*.

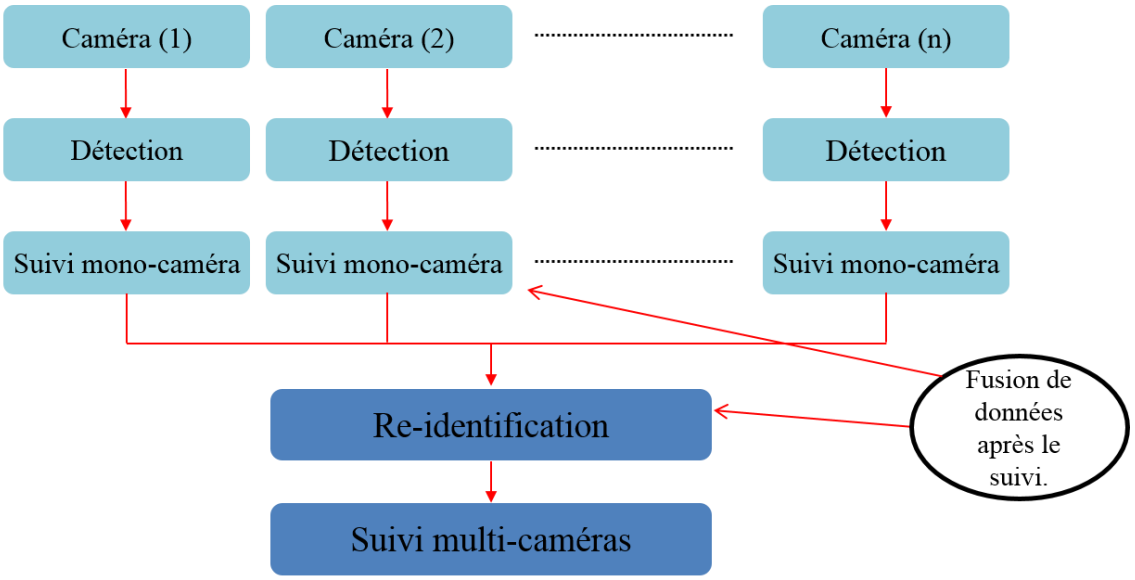


Figure 7. Le pipeline de base pour une approche distribuée.

Dans cette approche, la première étape est un suivi mono-caméra qui utilise une des méthodes classiques de suivi mono-caméra comme un filtre de Kalman, un filtre particulaire ou simplement un suivi par détection. Cette partie d'algorithme est utilisée pour le recouvrement des trajectoires des individus suivis dans le champ de vision de chacune des caméras indépendamment des autres. Ensuite, une fois les segments de trajectoire récupérés la deuxième étape consiste dans la recherche des correspondances entre les individus suivis dans l'ensemble des caméras pour trouver les correspondances entre chaque apparition d'individu dans une caméra donnée avec ces apparitions dans les autres caméras. Cette étape est communément connue comme la ré-identification de personnes. La troisième et dernière étape est celle ou les segments de trajectoires obtenus dans la première étape sont combinés à l'aide des correspondances obtenues dans la deuxième étape pour le recouvrement des trajectoires globales.

Toutes ces méthodes ont le même pipeline de base avec de petites différences. Dans ce pipeline, l'étape qui fait le plus de différence entre les méthodes proposées est l'étape de ré-identification dans laquelle une variété d'algorithmes a été proposée. Une synthèse des méthodes utilisant cette approche est présentée dans le Tableau 2.

Référence	Vecteur de caractéristique	Méthode d'association	Calibration
Back et al Kuo et al	Haar-like, Histogramme de couleur, Matrix de covariance, HOG.	Adaboost.	-
Avraham et al Kenk et al	HOG, Histogramme de couleur.	SVM.	-
Nakajama et al	HOG.	SVM multi-classes	-
Prosser et al	Couleur et texture.	SVM de classement.	-
Loy et al	Couleur.	Graphe laplacien.	-
Bazzani et al	histogramme plus modèle.	K-NN	-
Bak et al	Descripteur de covariance.	SVM	-
Javed et al	Couleur.	Géométrie épipolaire.	Oui.
Aziz et al	SIFT, SURF, image spin.	Modèle 2D, k-NN.	Oui.
Cheng et al Wang et al	Couleur, texture.	-	-
Cong et al	Couleur.	Modèle 2D, k-NN.	-
Chen et al	Spectre de couleurs.	Association de Tracklets.	-
Hamdoun et al	SURF	Algorithme de vote, K-NN.	-
Farenzena et al	Une variété de descripteurs d'apparence.	Les axes symétriques et asymétriques d'un modèle de corps.	-

Mazzon and Cavallaro	Positions et vitesse.	Modèle de force sociale.	-
Bouma et al	Couleur et informations spatiale.	Intersection d'histogrammes.	-
Nie et al	HSV, HOG, informations spatiale.	Appariement de graphes.	-
Brendel et al	Positions, taille, histogramme de couleur, intensité de gradient et mouvement.	Algorithme de MWIS.	-
Li et al	Couleur.	-	-
Colombo et al	Propriétés spatiotemporelle, Couleur.	Mesure de distance.	Oui

Tableau 2.Synthèse des méthodes utilisant une approche distribuée.

1.2.2.1. Suivi mono-caméra de personnes

Dans les approches distribuées, le suivi de personnes est fait pour chacune des caméras indépendamment des autres caméras du réseau. Cette étape de suivi communément nommé le suivi mono-caméra de personnes. Elle consiste dans la reconstruction des trajectoires des cibles lorsqu'elles se déplacent dans le champ de vision d'une seule camera. Dans le cadre de suivi multi-caméras avec une approche distribuée, cette étape est généralement réalisée à l'aide d'un algorithme classique de suivi comme un filtre de Kalman, un filtre particulaire ou un TLD (*Tracking by Learning Detection*). Cette étape est utilisée pour récupérer des fragments de trajectoires (tracklets) pour chaque individu. Ce suivi mono-caméra peut être considéré comme une étape préalable pour la tâche principale qui est la ré-identification qui permet d'associer les individus dans les différentes caméras pour récupérer la trajectoire globale de la personne suivie sur toute la scène.

1.2.2.2. Ré-identification

Comme nous l'avons dit précédemment, l'étape principale pour toutes les méthodes de suivi multi-caméras qui utilisent une approche distribuée est l'étape de la ré-

identification. La ré-identification de personnes dans un réseau de caméras peut être définie comme la capacité à décider si une personne qui est actuellement détectée dans le champ de vision d'une caméra donnée correspond à une personne précédemment détectée dans le champ de vision d'une autre caméra du réseau. Ce principe est illustré dans la *Figure 8*.

Figure 8. Illustration du principe de re-identification.

Une variété de méthodes a été proposée dans ce contexte. Ces méthodes de ré-identification peuvent être classées en deux grandes catégories les méthodes directes et celles par apprentissage.

1.2.2.3. Méthodes directes

Les méthodes directes de ré-identification sont basées sur l'extraction d'une signature discriminante pour chaque individu suivi. Ces signatures sont directement utilisées pour comparer et apparier l'ensemble des individus suivis dans toutes les caméras. Ces méthodes n'ont pas besoin d'une base d'apprentissage. Plusieurs méthodes ont été proposées en se basant sur cette approche. Parmi ces techniques de ré-identification, on peut distinguer les méthodes qui sont dites single-shot et celles qui sont dites multiple-shot. Les single-shot sont basées sur l'utilisation d'une seule image pour chaque individu suivi dans le champ de vision d'une caméra tandis que les multiple-shot sont basées sur l'utilisation de plusieurs images pour le même individu dans le champ de vision d'une caméra. L'efficacité de ce genre de techniques est liée directement à la signature utilisée pour la description des personnes. De ce fait, une variété de descripteurs globaux et

locaux ont été utilisés pour construire ces signatures. Parmi ces descripteurs utilisés on peut citer la texture [36], la forme [37, 38] et la combinaison d'un ensemble de descripteurs locaux [39].

Bak et al [40] ont proposé un modèle d'apparence basé sur l'utilisation de régions de covariance spatiale obtenue à partir des différentes parties du corps humain. Dans leur méthode, ils ont utilisé le HOG pour détecter les différentes parties du corps et la covariance comme descripteur pour obtenir les informations colorimétriques pour chacune de ces parties. Ces informations colorimétriques sont utilisées en combinaison avec les informations structurelles pour obtenir une signature discriminante qui est utilisée pour la ré-identification. Dans le même contexte, Chen et al [41] ont utilisé le spectre de l'histogramme de la couleur principale avec deux mesures de similarités ; la similarité positive (PS) pour les segments connectés et la similarité négative (NS) pour les segments non connectés afin d'améliorer leur suivi inter-caméras. De la même manière, Aziz et al [42] ont développé une méthode basée sur la classification des apparences en deux catégories les apparences de front et celles de dos avant la segmentation des silhouettes en trois parties (la tête, le torse et les jambes). Une fois les individus détectés, classifiés et segmentés, une signature discriminante basée sur l'utilisation du SIFT, SURF et spin image est extraite pour chaque partie du corps. Ces signatures sont utilisées pour l'appariement des différentes parties afin de ré-identifier les individus suivis. Cheng et al, [43] pour le même principe, ont utilisé une structure picturale personnalisée afin de localiser les parties du corps et d'extraire leurs descripteurs ; alors qu'Alahi et al [44] ont proposé une méthode basée sur la création d'un descripteur d'objet construit à partir de la cascade de grilles de descripteurs de régions.

Toujours dans les méthodes directes de ré-identification, Wang et al [37] ont introduit le concept de contexte de forme et d'apparence qui est utilisé dans le calcul d'une matrice d'apparence qui représente le descripteur de l'objet (personne) en question. Ce descripteur est construit par la capture des relations spatiales entre les différents labels d'apparences. Pour calculer ces matrices d'apparence, ils ont créé un framework dans lequel le calcul se fait en temps réel. Dans la même idée, Cong et al [45] ont proposé une nouvelle signature qu'ils ont appelé histogramme de couleur-position. Cet histogramme est construit par la division de la silhouette de l'individu à ré-identifier en plusieurs régions horizontales et l'utilisation des valeurs RGB normalisées de chaque région afin d'obtenir une signature discriminante à utiliser dans la ré-identification. Bouma et al [46] ont utilisé une méthode basée sur les intersections entre histogrammes pour associer les

descripteurs colorimétrique et spatiaux, tandis que Nie et al [47] ont utilisé une méthode d'appariement de graphes.

Dans le même contexte, Hamdoun et al [38] ont proposé un schéma d'identification basé sur l'appariement des points d'intérêt qui sont collectées sur un ensemble d'images dans des vidéos à courte durée. Ils ont utilisé le SURF pour extraire ces points d'intérêts et leurs descripteurs afin de construire une signature discriminante pour chaque individu suivi. Jungling et Arens [48] ont de la même manière utilisé le SIFT comme descripteur associé au modèle (ISM) proposé dans [49] pour la construction de leur système de ré-identification.

Parmi les méthodes qui utilisent une variété de descripteurs locaux, on peut également citer celle proposée par Farenzena et al [39]. Cette méthode est basée sur l'accumulation de plusieurs descripteurs locaux pour obtenir une signature discriminante qui peut être utilisée dans la différentiation entre les individus. L'algorithme est construit en trois étapes principales et il est utilisable en single-shot comme en multiple-shot. La première étape consiste dans la localisation des différentes parties du corps à l'aide des axes de symétrie et d'asymétrie. Une fois détectées, ces parties sont utilisées pour l'extraction de plusieurs descripteurs d'apparence. Ces descripteurs sont ensuite utilisés pour l'appariement des individus en accumulant les distances entre l'ensemble des descripteurs. De la même manière, Mazzon et al [50] ont combiné les descripteurs d'apparence extrait du haut du corps des individus suivis avec leurs positions ; tandis que Brendel et al [51] ont proposé une méthode basée sur l'utilisation d'un descripteur qui regroupe la position, la taille, l'histogramme de couleur, le gradient d'intensité ainsi que le flow optique pour chaque individu.

Toujours dans les méthodes directes, Li et al [52] ont combiné les points forts des algorithmes basés sur l'appariement et ceux basés sur la classification. Ils ont extrait les descripteurs utilisés dans l'appariement d'un espace métrique appris au lieu de les extraire directement de l'espace de données ; tandis que Colombo et al [53] ont proposé une technique pour estimer la transformée entre l'espace de couleur d'une caméra et l'espace de couleur d'une autre caméra pour obtenir une bonne constance de couleur entre les valeurs de couleurs associées à chaque caméra. L'utilisation d'une telle transformation améliore la performance de l'utilisation d'un descripteur de couleur pour ré-identifier les individus à travers les différentes caméras. Pour suivre les individus dans un réseau de caméras, Mazzon et Cavallaro [54] ont proposé une méthode qui exploite l'environnement. Leur méthode est basée sur l'utilisation d'un modèle de force sociale qui

prend en compte les points d'intérêts dans la scène étudiée (les entrées, les sorties, les points de ralliements). Ce modèle facilite la prédiction des trajectoires en se basant sur les buts et les barrières qui peuvent influencer les trajectoires des personnes suivies.

1.2.2.4. Méthodes par apprentissage

Les méthodes de ré-identification par apprentissage sont basées sur l'utilisation d'une base d'apprentissage pour analyser les vecteurs de caractéristiques extraits de ces échantillons spécifiques et puis généraliser le concept pour ré-identifier les nouveaux individus qui ne figurent pas dans la base d'apprentissage. Ces méthodes diffèrent dans les algorithmes de classification qui sont utilisés. De ce fait, plusieurs méthodes ont été proposées avec les différents algorithmes de classification. Parmi ces méthodes on peut citer celle utilisant le Adaboost [55], le SVM [56], les méthodes d'apprentissage de descripteurs [57], les méthodes d'apprentissage de distances [58], le plus proche voisin [59] et la réduction des moindres carrés [60].

Bak et al [55] ont proposé un algorithme pour la ré-identification de personnes basé sur l'utilisation de Adaboost avec les descripteurs de Haar. Dans cette méthode la détection et le suivi de personnes est fait par l'utilisation d'un histogramme de HOG. Une fois le suivi effectué, un ensemble de détections pour chacune des personnes suivies est accumulées sur une petite période de temps. Cet ensemble de détections est utilisé par la suite pour extraire les descripteurs de Haar de cette personne. À la fin ces descripteurs sont utilisés avec l'Adaboost pour extraire une signature visuelle qui sera utilisée pour la ré-identification.

Les méthodes utilisant un classificateur SVM sont généralement basées sur le même principe. Cela consiste à apprendre un SVM en se basant sur un ensemble d'apprentissage qui regroupe plusieurs images des individus détectés d'une caméra donnée et ensuite utiliser ce classificateur pour associer les individus détectés dans les autres caméras. Dans ce contexte, Avraham et al [61] ont proposé un algorithme qui est appris pour trouver les correspondance entre deux caméras qui ne se chevauchent pas. Cette méthode est basée sur l'utilisation d'un classificateur qui peut différencier les paires positive et les paires négative de détections en se basant sur l'utilisation d'un descripteur couleur. Ils considèrent comme paire positive le cas où les deux images représentent la même personne et paire négative le cas où les deux images représentent deux personnes différentes et à la fin ce classifieur est utilisé pour décider si une nouvelle paire de détections représente le même individu. Toujours dans la même catégorie, Nakajama et

al [62] ont proposé un SVM multi-classes qui repose sur la combinaison de plusieurs classifieurs binaires. Ce classifieur est utilisé pour la ré-identification de personnes à l'aide de descripteurs de couleur et de forme. De la même manière, Kenk et al [63] ont utilisé un histogramme de couleur pour créer un classifieur multi-classes basé sur des classifieurs binaires pour résoudre le problème de ré-identification dans un environnement distribué de caméras. Tandis que Prosser et al [56] ont reformulé le problème de ré-identification comme un problème d'apprendre à classer, pour cela ils ont utilisé un SVM (RankSVM) sur des descripteurs de couleur et de texture pour pouvoir ordonner les différents appariements possible au lieu de faire un appariement direct. Loy et al [64] ont adopté le même principe en proposant un cadre de classification multiple pour la ré-identification, à cet effet, ils ont étudié deux modèles de classification multiples : le graphe Laplacien normalisé et le graphe itéré Laplacien non normalisé.

Une autre catégorie d'approches est celle qui regroupe les méthodes basées sur l'apprentissage des descripteurs. Cette catégorie de méthodes repose sur l'utilisation d'une combinaison pondérée de plusieurs descripteurs qui décrivent l'apparence et la forme de l'individu suivi. Dans ce cas, la phase d'apprentissage est utilisée pour pondérer les différents descripteurs qui composent la signature qui décrit les individus. Parmi ces méthodes, on peut citer celle proposée par Figueira et al [57], dans laquelle, l'algorithme extrait un ensemble de descripteurs d'apparence des différentes parties du corps pour chaque individu. Cet ensemble de descripteurs est utilisé pour estimer les paramètres du modèle en utilisant le framework d'apprentissage multi-descripteurs de [65]. Toujours dans le même ordre d'idées, Kuo et al [66] ont proposé une méthode qui repose sur l'apprentissage online d'apparence discriminante. Ils ont utilisé l'Adaboost sur un ensemble de descripteurs (histogramme de couleur, matrice de covariance, HOG) calculés dans différentes positions sur l'individu détecté pour pouvoir combiner les descripteurs les plus discriminants dans un modèle d'apparence qui sera par la suite utilisé dans la ré-identification. Bazzani et al [67] ont proposé un nouveau descripteur qui condense les informations de plusieurs images de la même personne dans une seule signature appelé «épitomé» (représentant une accumulation de descripteurs locaux sur un ensemble de détections). Cette signature englobe les descripteurs d'apparences globaux et locaux pour chaque individu détecté.

Les méthodes qui se basent sur l'apprentissage de distances visent à apprendre une mesure de similarité entre les paires d'images ou les paires de descripteurs qui représentent les individus détectés. En général, l'algorithme essaye d'apprendre les différences entre les paires positives qui représentent le même individu et les paires

négatives qui représentent des individus différents au lieu d'apprendre les descriptions visuelles de ces individus. Pour cette catégorie de méthodes, on peut mentionner les travaux de Zheng et al [68] dans lesquelles le problème de ré-identification est reformulé comme un problème d'apprentissage de distances relatives. Dans cette reformulation, il est considéré que la distance entre les paires de vrais appariements est inférieure à la distance entre les paires de faux appariements. Ce modèle de comparaison de distances relatives est appliqué sur des descripteurs d'apparence (texture et histogrammes de couleurs). Dans le même contexte, Bak et al [36] ont utilisé le descripteur de covariance proposé dans [69] qui code les informations de covariance entre les différentes régions d'une image afin de caractériser les individus. Mais au lieu d'utiliser qu'un seul descripteur, ils ont utilisé les descripteurs les plus discriminants pour chacune des parties de l'image. On peut aussi citer la méthode de Javed et al [70] qui est basée sur l'apprentissage des relations spatio-temporels et d'apparence entre deux caméras données pour pouvoir calculer la probabilité que deux observations appartiennent à la même personne. Cet algorithme permet d'apprendre la fonction de transfert de luminosité (*brightness transfer function*) entre les deux caméras ainsi que le positionnement des entrées, sorties, directions de mouvement et le temps dont a besoin un individu qui quitte le champ de vision d'une caméra pour atteindre le champ de vision d'une autre caméra.

1.2.3. Utilisation des différentes approches

Dans cette section, le Tableau 3 représente une comparaison des deux approches. Il présente les avantages et les inconvénients pour chacune des deux approches ainsi que certains domaines d'application où le suivi peut être utilisé.

Comme décrit dans le tableau, il n'y a pas de critère discriminant qui peut permettre de dire que l'une des deux approches est meilleure que l'autre. Les deux approches ont certains avantages et certains inconvénients. De ce fait, le choix de l'approche à utiliser dépends plus des spécifications de l'application qu'on souhaite implémenter.

	Approche centralisée	**Approche distribuée**
Utilisation	• Gestion des petits espaces. • Caméras avec zones de recouvrement. • Espaces encombrés (foules).	• Gestion de grandes surfaces. • Cameras sans zones de recouvrement.

Bénéfices	• Exploitation de la vision multi-vue de la scène. • Gestion des occultations. • Réduction de des effets de bruit et de panne qui peut affecter certains capteurs.	• La synchronisation entre les différents capteurs n'est pas nécessaire. • La transformation géométrique entre les cameras n'est pas nécessaire pour le suivi. • Nécessite un petit nombre de cameras. • Plus adapter pour les applications en temps réel.
Inconvénients	• La synchronisation des caméras est nécessaire. • La transformation géométrique entre les cameras est nécessaire pour le suivi. • Nécessite un grand nombre de cameras pour couvrir une petite surfaces. • Moins adaptée pour la réalisation d'applications en temps réel.	• Une seule caméra est utilisée même dans le cas où plusieurs caméras sont disponibles. • Problème dans la gestion des occultations. • La présence de bruit ou de panne affecte tout le système de suivi.
Domaines d'application	• Systèmes intelligents de surveillance. • Contrôle de Traffic urbain. • Gestion des flux d'individus dans les espaces publiques. • Comptage de personnes. • Analyse de comportement des individus. • Etudes statistiques.	

Tableau 3. Résumé de l'utilisation, les bénéfices et les inconvénients des deux approches.

1.3. Description de la méthodologie proposée

Nous terminons ce chapitre par la présentation de l'approche que nous avons mis au point dans le contexte de l'analyse du comportement des consommateurs. En effet, le développement de nouvelles stratégies de marketing pour les industries de la grande distribution est basé sur l'étude du comportement des clients. Ces études sont généralement basées sur l'exploitation des vidéos de surveillances. Ainsi, le besoin d'outils automatiques qui peuvent être utilisés pour analyser ces vidéos sans avoir besoin d'opérateurs humains devient nécessaire. La première étape dans toute application

conçue pour analyser le comportement des clients est le suivi de personnes dans un réseau multi-caméras. Dans ce contexte, notre travail consiste dans le développement d'un outil de suivi qui permet de reconstruire les trajectoires de clients dans une surface de vente. Pour ce faire, nous utilisons les images des systèmes de vidéo-surveillance qui sont déjà sur place.

1.3.1. Description générale

Dans ce travail, notre objectif est le développement d'un système qui va permettre de suivre plusieurs personnes dans un réseau multi-caméras. Pour cela, nous avons proposé un système de suivi qui est composé de trois modules principaux et d'un module de gestion. Le premier module est un module de comptage basé sur l'utilisation d'une Kinect pour mesurer les entrées. Le deuxième module est un système de suivi mono-caméra qui est basée sur l'utilisation du filtrage à particules, il est utilisé pour suivre les individus indépendamment dans chacune des caméras. Le troisième module est un module de ré-identification qui est basé sur l'utilisation de la sélection des régions saillantes, il permet d'associer les individus détectés dans les différentes caméras. Et en fin, le module de gestion basé sur l'utilisation d'ontologies pour l'interprétation des trajectoires. Un exemple de ce système est illustré dans la Figure 9.

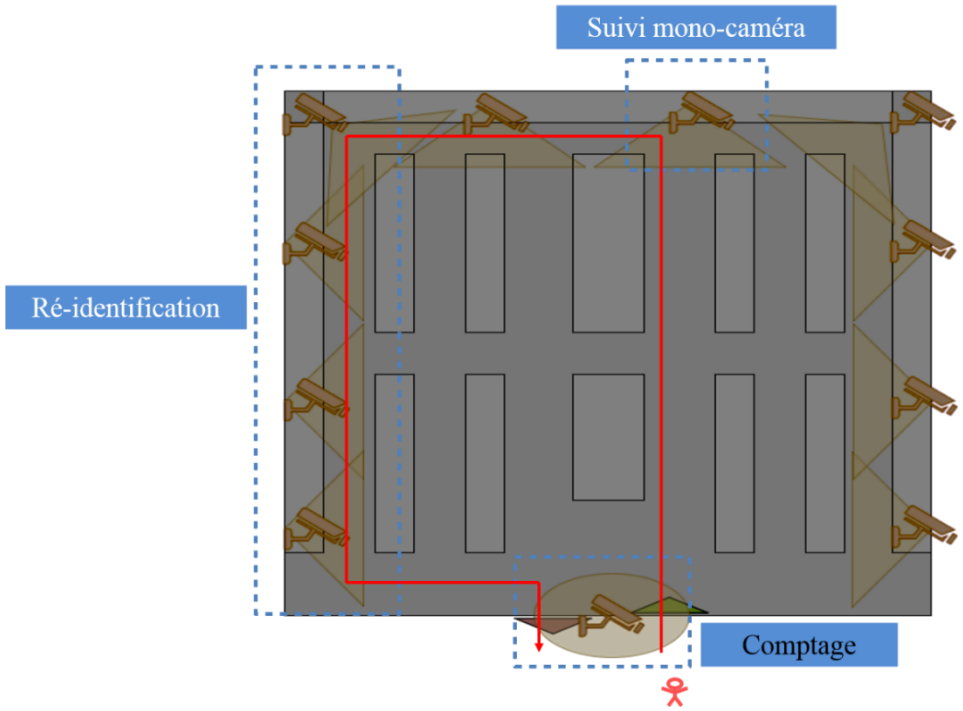


Figure 9. Système proposé.

1.3.2. Défis à relever

Nous avons divisé le problème principal de suivi de personne dans un réseau de cameras en trois sous problèmes qui sont le comptage, le suivi mono-caméra, et la ré-identification de personnes. Nous avons proposé une méthode pour résoudre chaque sous problème séparément. Les solutions proposées sont décrites dans la section suivante. Les résultats obtenus par les différents modules seront combinés pour avoir un système complet de suivi de personnes dans un système multi-caméras.

La réalisation de ce système est confrontée à un ensemble de défis :

- L'implémentation d'un système de comptage pour mesurer les entrées ; ce système doit être simple, facile à implémenter et suffisamment efficace.
- L'implémentation d'un système de suivi mono-caméra : ce système doit être capable de détecter les nouveaux individus, gérer les occultations et doit être capable de suivre plusieurs individus simultanément.
- L'implémentation d'un algorithme de ré-identification : cet algorithme doit être capable de s'adapter aux variations d'éclairage, aux détections partiales, aux apparences similaires ainsi que aux changements de poses.
- La réalisation d'un système de gestion : ce système doit être capable de donner une sémantique pour les trajectoires obtenues précédemment.

1.3.3. Description générale des différents modules proposés

1.3.3.1. Module de comptage

Le premier module de notre système est le module de comptage. Ce module est conçu pour mesurer le flux de personnes qui rentre et qui sort du magasin. Ce module est basé sur l'exploitation de données de profondeur obtenues à l'aide d'une Kinect installée d'une façon verticale à l'entrée du magasin. L'idée principale derrière ce module réside dans l'utilisation d'un seuillage qui permet la détection de personnes qui passent sous la Kinect. Une fois ces individus détectés, ils seront suivis à l'aide d'un modèle cinétique associé à un MeanShift pour éviter leur tant qu'ils sont dans le champ de vision de la Kinect. Cette procédure de comptage est illustrée dans la Figure 10 et elle est composée de plusieurs étapes :

- Initialisation de la distance qui sépare la Kinect du sol pour chaque point dans le champ de vision de la Kinect.

- Capturer les données de profondeur qui représentent la distance entre la Kinect et la scène à chaque instant t.
- Utilisation des données capturées à l'instant t avec la profondeur initiale pour détecter les changements de profondeur qui seront segmentées afin de détecter les personnes qui sont présentes dans le champ de vision de la Kinect à l'instant t.
- Chaque nouvelle personne détectée est associée à un trackeur qui est utilisé pour suivre cette personne dans les frames qui suivent (t+1, 2,….).

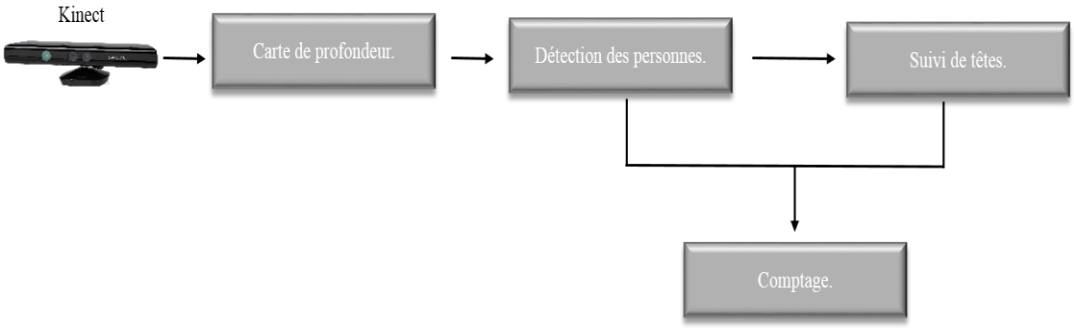


Figure 10. comptage de personne.

Les détails de cette approche ainsi que les résultats seront présentés dans le chapitre 2.

1.3.3.2. Module de suivi mono-camera

Le deuxième module de notre système est le module de suivi mono-caméra. Dans notre approche, nous avons proposé un module de suivi mono-camera qui est utilisé pour récupérer des fragments de trajectoires (tracklets) pour chaque individu.

Pour récupérer ces fragments de trajectoires, nous avons utilisé une stratégie de suivi qui est basé sur l'utilisation d'un filtrage particulaire. Nous avons donc associé un nombre de particules pour chaque individu suivi. Puis, afin d'estimer la nouvelle position de l'individu, nous avons utilisé un modèle de mouvement avec vitesse constante. La nouvelle position estimée est utilisée pour l'estimation des nouvelles positions des particules dans notre filtre particulaire. Ces particules sont pondérées par rapport à la différence entre leur histogramme couleur et l'histogramme couleur de l'individu suivi dans l'espace de couleur HSV. Les particules pondérées sont utilisées pour corriger l'estimation de la nouvelle position de l'individu suivi. Cette procédure de suivi est illustrée dans la *Figure 11*.

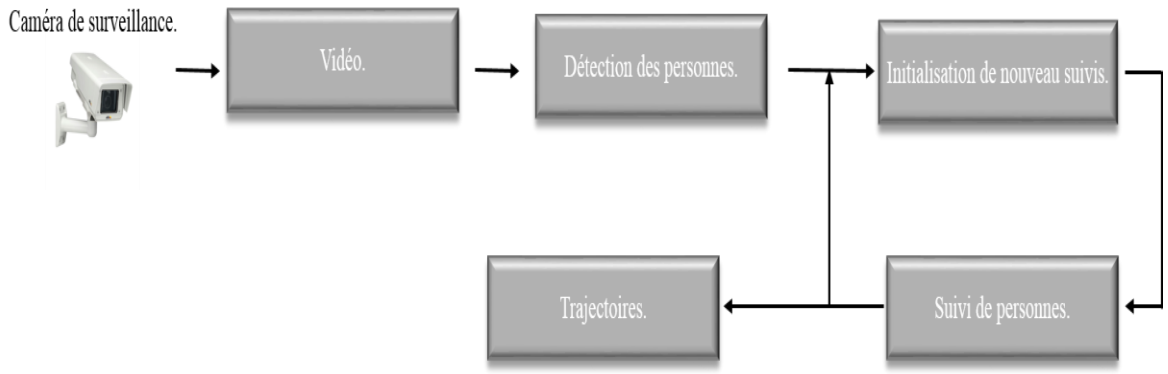


Figure 11. Suivi mono-camera.

Les détails de cette approche ainsi que les résultats seront présentés dans le chapitre 2.

1.3.3.3. Module de Ré-identification

Le troisième module dans notre système de suivi est le module de ré-identification. Ce module est utilisé pour trouver les correspondances entre les individus qui sont détectés dans les différentes cameras. Ces correspondances seront utilisées pour fusionner les fragments de trajectoires obtenues dans l'étape de suivi mono-caméra pour avoir une vue globale de la trajectoire d'une personne sur l'ensemble de la scène. Le principe de ce module est illustré dans la Figure 12.

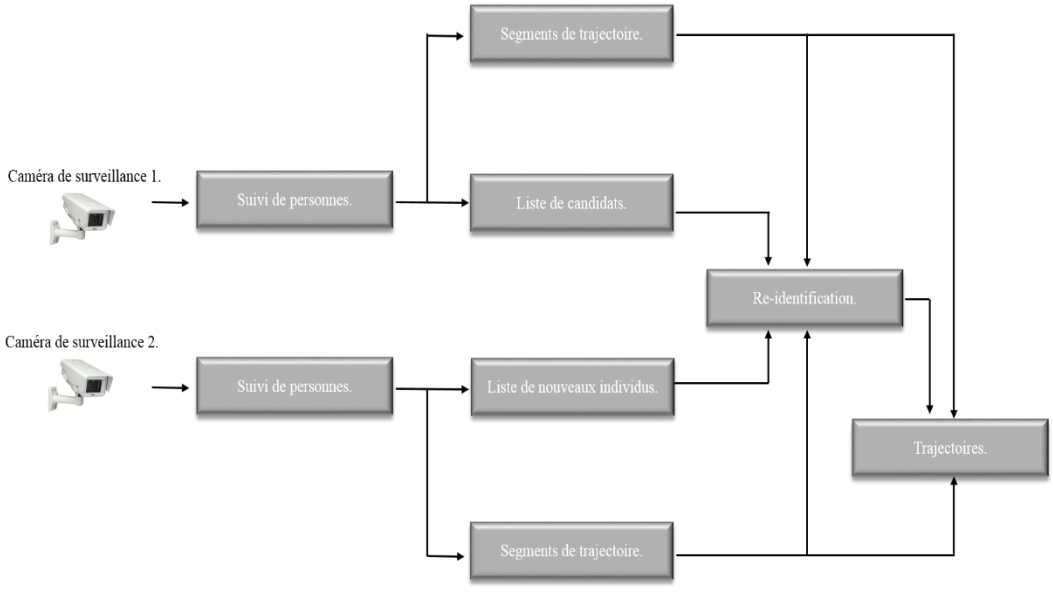


Figure 12. Utilisation du module de ré-identification.

Pour cette partie, nous avons proposé une méthode originale qui est basée sur la sélection des régions saillantes pour chaque individu.

Nous avons proposé un algorithme qui composé d'une étape d'apprentissage et d'une étape d'appariement. L'étape d'apprentissage est utilisée pour la sélection des régions saillantes qui différencient le plus possible les différents individus et l'étape d'appariement pour associer les individus à partir des régions saillantes sélectionnées précédemment.

Les détails de cette approche ainsi que les résultats seront présentés dans le chapitre 3.

1.3.3.4. Module de gestion

Le quatrième et dernier module est le module de gestion. Ce module est utilisé pour rajouter de la connaissance aux trajectoires brutes afin de faciliter leurs interprétations par les opérateurs humains. Une façon adéquate pour ajouter ces connaissances consiste dans la construction d'une ontologie pour la construction des trajectoires sémantiques. Pour ce faire, nous avons proposé un module pour la construction des trajectoires sémantiques. Ce module est basé sur la construction de trois ontologies :

- Une ontologie de modélisation de l'environnement pour modéliser les parties statiques de l'espace surveillé.
- Une ontologie de modélisation des trajectoires qui est utilisée dans la construction des trajectoires sémantiques.
- Une ontologie d'observation dans l'environnement qui permet de relier les trajectoires à l'environnement.

Cette approche sera détaillée dans le chapitre 4.

1.4. Conclusion

Dans ce chapitre, nous avons commencé par l'étude de l'état de l'art qui nous a permis d'étudier les différentes approches proposées dans le cadre de suivi multi-caméras de personnes avant de pouvoir proposer notre approche.

Dans la première partie, nous avons présenté un résumé des méthodes proposées dans la cadre de suivi multi-caméras de personnes. Nous avons classé ces méthodes en deux catégories, celles basées sur l'utilisation d'une approche centralisée et celles basées sur l'utilisation d'une approche distribuée. Nous avons fait cette classification en nous référant à la fusion de données entre les différentes caméras du réseau. Pour chacune de ces approches, une variété de méthodes a été présentée et discutée. Cet état de l'art a donné lieu à une publication dans un journal.

Dans la deuxième partie, nous avons présenté un aperçu global de notre système de suivi. Nous avons proposé une méthode basée sur l'utilisation d'une approche distribuée. Dans notre système, nous avons subdivisé le problème principal de suivi de personnes dans un réseau de caméras en quatre sous-problèmes qui sont le comptage, le suivi mono-caméra, la ré-identification et l'interprétation de trajectoires.

Chapitre 2. Comptage et suivi mono-caméra de personnes

2.1. Introduction

Dans ce chapitre, nous nous focalisons sur le comptage et le suivi mono-caméra des personnes. L'un des aspects les plus étudié est la conception d'outils automatiques de comptage qui permettent de dénombrer les personnes traversant une ligne d'intérêt ou les personnes se retrouvant dans une zone d'intérêt. Ces systèmes de comptage de personnes sont utilisés pour l'obtention d'informations sur le flux des individus dans des zones piétonniers (supermarchés, métros, gares…), à des fins de statistiques pour la gestion des flux d'individus ou à des fins de sécurité.

Ces dernières années, plusieurs méthodes de comptage qui intègrent la vision par ordinateur ont été proposées. Ces méthodes utilisent, de plus en plus, des caméras RGB ou des capteurs de profondeur pour compter soit le nombre de personnes qui traversent une ligne d'intérêt (LOI) ou le nombre de personnes qui se retrouvent dans une zone d'intérêt (ROI).

Un autre aspect très étudié dans le domaine de l'analyse du comportement est le suivi mono-caméra de personnes. Le suivi est défini comme étant le fait d'avoir la capacité de localiser les individus à chaque instant durant leurs passages dans le champ d'une caméra donnée. Dans la pratique, nous retrouvons deux types de suivi. Le premier est le suivi d'un seul individu et le deuxième est le suivi de plusieurs individus simultanément. Le suivi d'un seul individu se focalise sur la résolution des problèmes de variations

d'éclairage et d'échelle, tandis que pour le suivi multi-personnes se rajoute le problème de la différentiation entre les individus suivis. Donc, le défi majeur dans le suivi multi-personnes consiste dans la gestion du problème de confusion entre les individus suivis. Cette confusion peut être causée par les occultations et par les interactions entre les individus. Dans ce sens, plusieurs travaux de recherches ont été menés pour y répondre.

Dans ce chapitre, nous avons développé deux des quatre modules principaux qui forment notre système de suivi multi-caméras décrit dans le chapitre précédent. Nous avons commencé par la conception d'un nouveau système de comptage qui peut être utilisé pour compter le nombre d'individus qui rentrent et sortent d'une zone. Pour ce faire, nous avons proposé un système basé sur l'exploitation des données de profondeur obtenues par l'utilisation d'une Kinect. Ensuite, nous avons développé un deuxième module pour le suivi mono-caméra de personnes. Ce module est utilisé pour la reconstruction des fragments de trajectoires des personnes suivis dans le champ de vision de chacune des caméras du réseau indépendamment des autres. Pour ce faire, nous avons développé une méthode basée sur l'utilisation d'un filtre particulaire. Ces deux modules, nous permettent de premièrement connaitre le nombre de personnes qui se rendent dans l'espace surveillé et deuxièmement de récupérer les segments de trajectoires des personnes suivies durant leurs présences dans l'espace surveillé.

2.2. Comptage de personnes

Comme nous l'avons signalé précédemment, le comptage de personnes est l'un des aspects les plus étudié dans l'analyse du comportement en utilisant la vision par ordinateur. Cette problématique de base se sépare en deux sous-problématiques, le comptage de personnes dans une zone d'intérêt (ROI) et le comptage de personnes qui traversent une ligne d'intérêt (LOI). Ces deux sous-problématiques sont illustrées dans *Figure 13*.

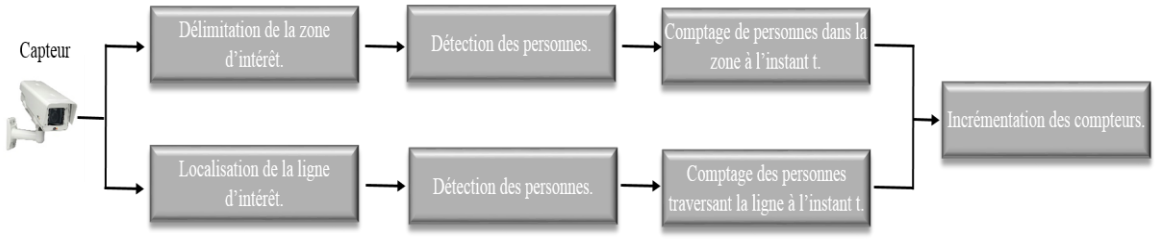


Figure 13. Problématiques de comptage.

Les comptages ROI consistent en la détermination du nombre d'individus qui se trouvent dans un espace donné à un instant donné, tandis que les comptages LOI se focalisent sur la détermination du nombre d'individus traversant une ligne (réelle ou virtuelle) à un instant donné. Dans les deux cas, les solutions adoptées sont généralement composées de deux phases. La première phase consiste dans la détection des individus dans les zones étudiées et la deuxième phase consiste dans le suivi de ces individus pour éviter leur recomptage.

Les principales difficultés liées à ces deux problématiques sont :

- Les changements d'éclairage.
- La localisation des individus.
- L'isolation des différents individus.
- Les confusions entre individus.
- Les occultations et les interactions entre individus et avec les objets de la scène.

2.2.1. État de l'art

Le comptage de personnes est considéré comme la partie principale dans la réalisation de plusieurs applications d'étude de comportement. Il peut être utilisé dans la gestion des foules, le control de présence pour des études statistiques, ainsi que pour des fins de sécurité. Ces déverse applications ont conduit à la réalisation de plusieurs recherches et au développement de plusieurs méthodes durant la dernière décennie. L'objectif principal de ces recherches consiste dans la conception d'outils automatiques pour compter le nombre de personnes qui se trouvent dans une ROI ou le nombre de personnes qui traversent une LOI. En général, toutes les méthodes de comptage sont composées de deux phases. La première phase est la détection d'individus en mouvement (ou une partie de leur corps) dans la scène. Cela est fait, en règle générale, par des algorithmes classiques de détection de mouvement, comme la soustraction de fond [71] et [72] ou la différence entre les frames [73], associés à des méthodes de segmentations adéquates comme l'analyse des composantes connexes ou le K-means qui sont utilisées pour segmenter les blobs à des individus [71] et [73] ou à leur têtes [74], [72], [75] et [76]. Ensuite, vient la deuxième phase qui est le suivi de personnes précédemment détectés afin de pouvoir déterminer la direction de leur mouvement et éviter le recomptage des mêmes individus.

Selon l'objectif du comptage, les méthodes existantes dans ce domaine peuvent être classées en deux catégories : les méthodes LOI de comptage [77], [78] et [79] et les méthodes ROI de comptage [71] et [72].

2.2.1.1. Comptage par rapport à une ligne d'intérêt (LOI counting)

La première catégorie de méthodes de comptage est celle basée sur l'analyse de mouvement à travers une ligne virtuelle (LOI) pour détecter et compter les individus. La *Figure 14* illustre le principe de cette approche.

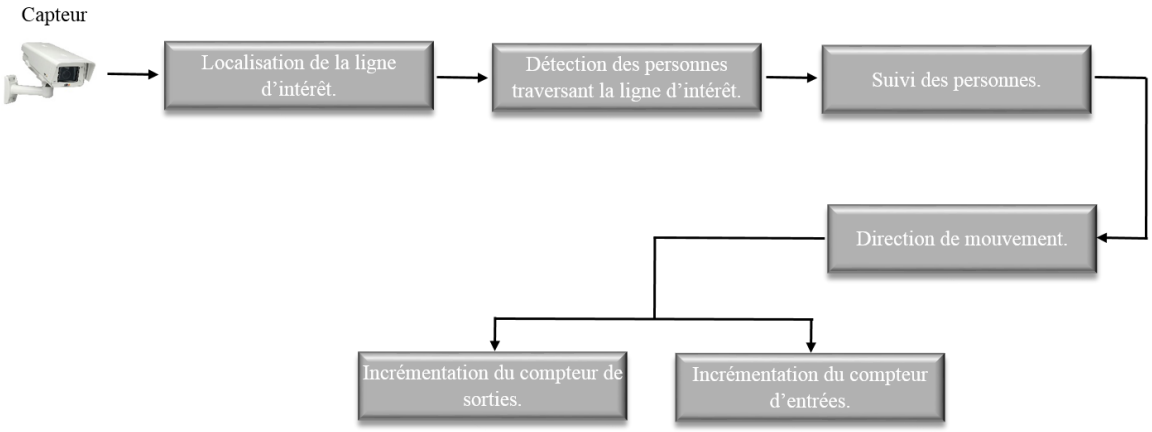


Figure 14. Les méthodes LOI.

Dans cette approche, nous pouvons cité la méthode proposée par Ma et Chan [77] pour compter le nombre d'individus qui traversent une ligne virtuelle dans une vidéo. Cette méthode est basée sur la programmation en nombre entiers. Elle est conçue pour l'estimation du nombre instantané d'individus en utilisant des caractéristiques locales et une régression. La méthode proposée est basée sur la construction d'une image de coupures à partir de coupures temporelles d'image où chaque colonne de l'image des coupures correspond à la ligne d'intérêt à un instant donné. Ensuite, l'image résultante est étudiée pour détecter les blobs qui représentent des fragments de foule traversant la ligne d'intérêt en utilisant le HOG ; enfin, une régression bayes de Poisson est utilisée pour estimer le nombre d'individus dans chaque fragment de foule. Contrairement à cela, Cong et al [79] ont proposé une méthode basée sur l'estimation du champ de vitesse d'écoulement. Dans cette méthode, la première étape consiste à détecter et à segmenter le champ de vitesse sur la ligne d'intérêt en fonction de la direction de mouvement.

Ensuite, une mosaïque dynamique est utilisée pour construire les blobs qui seront à leur tour utilisés pour estimer le nombre d'individus traversant la ligne d'intérêt. Cette estimation est faite à l'aide d'une régression sur les surfaces et les contours de ces blobs.

Toujours dans le contexte de comptage d'individus traversant une ligne d'intérêt, une autre approche a été proposée par Del Pizzo et al [78]. Cette approche est basée sur la conception d'une grille pour détecter les individus. Cette grille est associé à une méthode de détection de la direction de mouvement similaire au codeur rotatif incrémental (Ellin et Dolsak [80]) pour déterminer la direction du mouvement. La séquence d'activation de la grille associée à la direction du mouvement permet la détermination du nombre d'individus traversant la ligne d'intérêt. Dans le même contexte, Barandiaran et al [81] ont proposé une méthode basée sur l'utilisation d'un nombre fixe de lignes virtuelles uniformes et orthogonales par rapport à la direction du mouvement dans une zone d'intérêt ; ensuite le comptage de personnes sur chacune des lignes est fait par la détection des intervalles non-nuls dans l'image de différence entre la frame actuelle et la frame précédente de la vidéo sur la ligne d'intérêt.

2.2.1.2. Comptage dans une région d'intérêt (ROI counting)

La deuxième catégorie des méthodes de comptage est celle des méthodes qui se focalisent sur la détermination du nombre d'individus présents dans une zone d'intérêt (ROI). La *Figure 15* illustre le principe de cette approche.

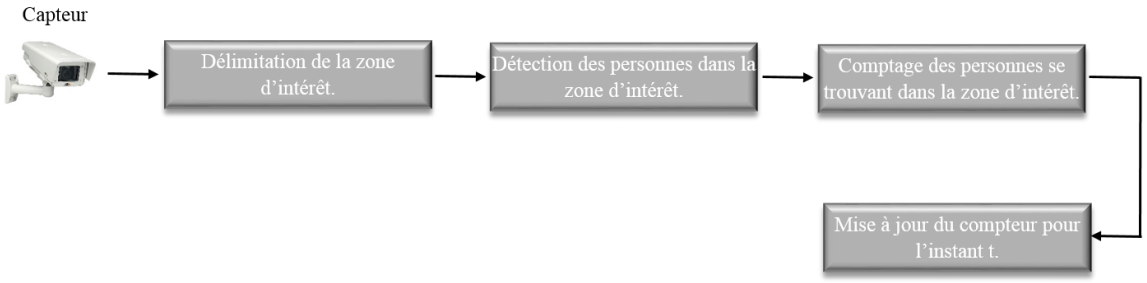


Figure 15. Les méthodes ROI.

Antic et al [71] ont proposé une méthode pour compter les individus dans une vidéo. Cette méthode est composée de trois étapes. La première consiste dans la détection des objets du premier-plan de la vidéo en utilisant un algorithme classique de soustraction de fond. La deuxième étape procède à la segmentation de ce premier-plan par un K-

means en essayant d'avoir le nombre maximal possible de clusters ; ce nombre est considéré comme étant le nombre d'individus présent dans la zone d'intérêt. La troisième et dernière étape consiste à suivre les individus détectés dans les étapes précédentes en se basant sur une association dynamique entre les clusters des frames consécutives, les points de fin des trajectoires obtenues sont utilisés dans l'incrémentation des compteurs d'entrées et de sorties. Dans le même sens, Chen et al [73] ont proposé une méthode dans laquelle la détection de mouvement est faite par la différence entre les frames et la segmentation des blobs est faite par l'analyse des composantes connexes et le suivi est assuré par les intersections entre les boites englobantes en supposant que les individus se déplacent lentement dans une foule.

Une autre approche pour cette catégorie de méthodes consiste dans l'exploitation du fait que la tête est la partie du corps la plus proche d'une caméra installée au plafond dans une position zénithale par rapport à la scène étudiée. Pour cette approche, Zhang et al [76] ont proposé une méthode basée sur un algorithme de remplissage d'eau (*water filling algorithm*). L'idée principale consiste dans la détection des minimums locaux de gouttes d'eau nécessaire pour remplir une carte de profondeur obtenue par une Kinect. Ces minimums locaux sont considérés comme étant les têtes des individus présents dans la zone d'intérêt. Dans le même contexte, Bondi et al [72] ont utilisé la carte de profondeur obtenue par stéréo pour repérer les têtes des individus présents dans la scène en localisant les minimums locaux de chacun des blobs représentant les objets de l'avant-plan. Ces blobs sont obtenus par l'utilisation d'un algorithme classique de soustraction de fond associé à une détection de contours. Cette détection de têtes associée à une stratégie simple de suivi permet de déterminer le nombre de personnes présentes dans la zone d'intérêt.

2.2.2. Description de la méthodologie propose

Dans la section précédente, nous avons présenté un état de l'art dans le domaine de comptage de personnes, ce qui nous a permis d'une part d'avoir une idée générale des différentes approches utilisées dans ce domaine et d'autre part de comprendre les différents défis rencontrés dans la réalisation de ce genre d'applications. Fort de ces éléments, nous nous sommes donc orientés vers une approche par LOI qui nous semble la plus adaptée à notre problème, à savoir le comptage du nombre d'individus qui rentrent et qui sortent d'un espace fermé.

Notre méthode est basée sur l'exploitation de données de profondeur obtenue d'une Kinect. Comme nous avons dit précédemment, la majeure partie des méthodes proposées

dans le domaine essaye de résoudre les mêmes problèmes de base qui sont : les occultations, la gestion de foules et les problèmes liés à la segmentation.

Afin de traiter ces défis, nous avons choisi d'utiliser une Kinect installée au plafond avec une position zénithale par rapport à la direction de mouvement. Ce positionnement est illustré dans la *Figure 16*.

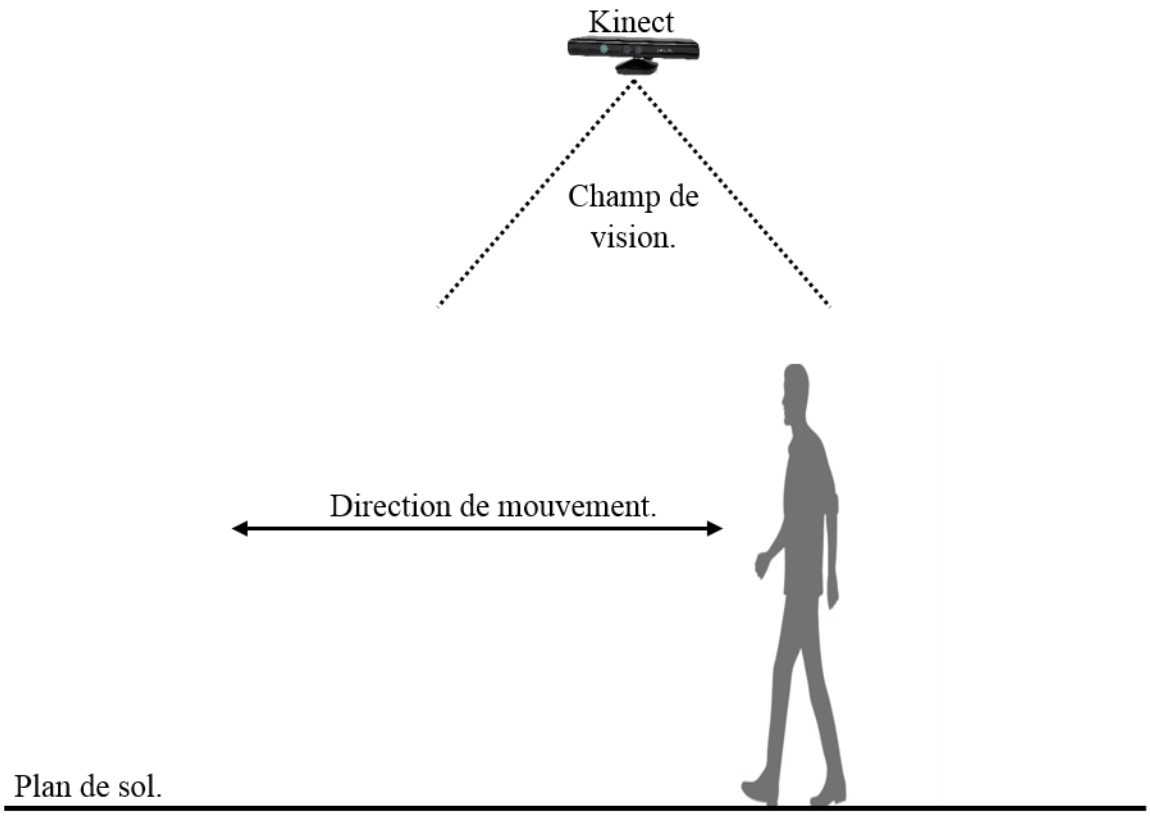


Figure 16. Positionnement de la Kinect.

La position zénithale de la Kinect permet d'éviter l'apparition des occultations et l'utilisation des données de profondeur permet d'améliorer la segmentation des blobs à des individus. Notre méthode de comptage est réalisée en deux phases : la détection et le suivi des personnes.

La première étape consiste à exploiter les cartes de profondeur obtenue de la kinect pour détecter les individus qui traversent une ligne virtuelle que nous traçons dans le champ de vision de la kinect. Il vient ensuite la deuxième étape qui consiste à suivre les

individus détectés précédemment pour déterminer la direction de leur mouvement (rentrant ou sortant) et éviter ainsi le recomptage des mêmes individus pour chaque frame de données. Cet algorithme est détaillé dans la *Figure 17*.

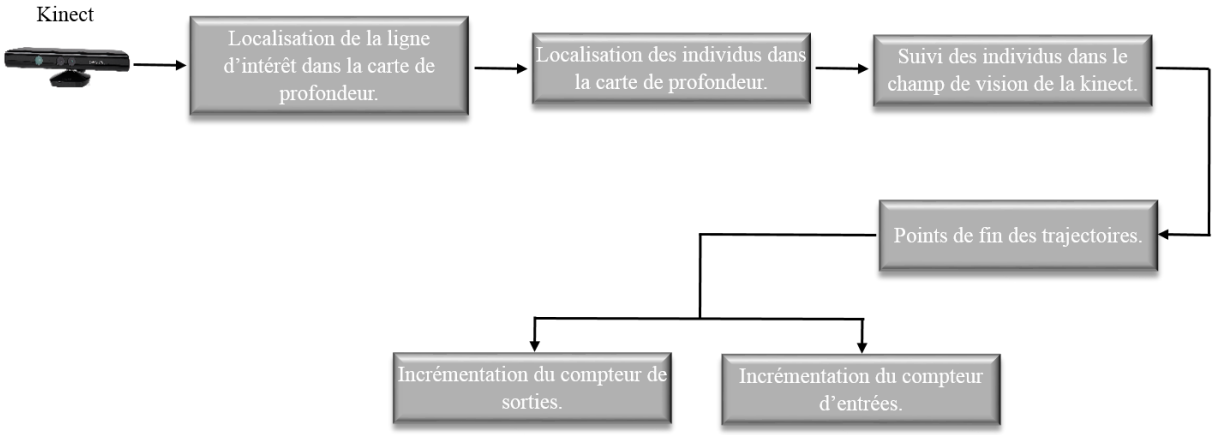


Figure 17. Notre méthode de comptage.

2.2.2.1. Détection de personnes

La première phase de notre stratégie de comptage est la détection de personnes dans les cartes de profondeur obtenues de la kinect. Les différentes étapes de cette phase sont illustrées dans le diagramme de la *Figure 18*.

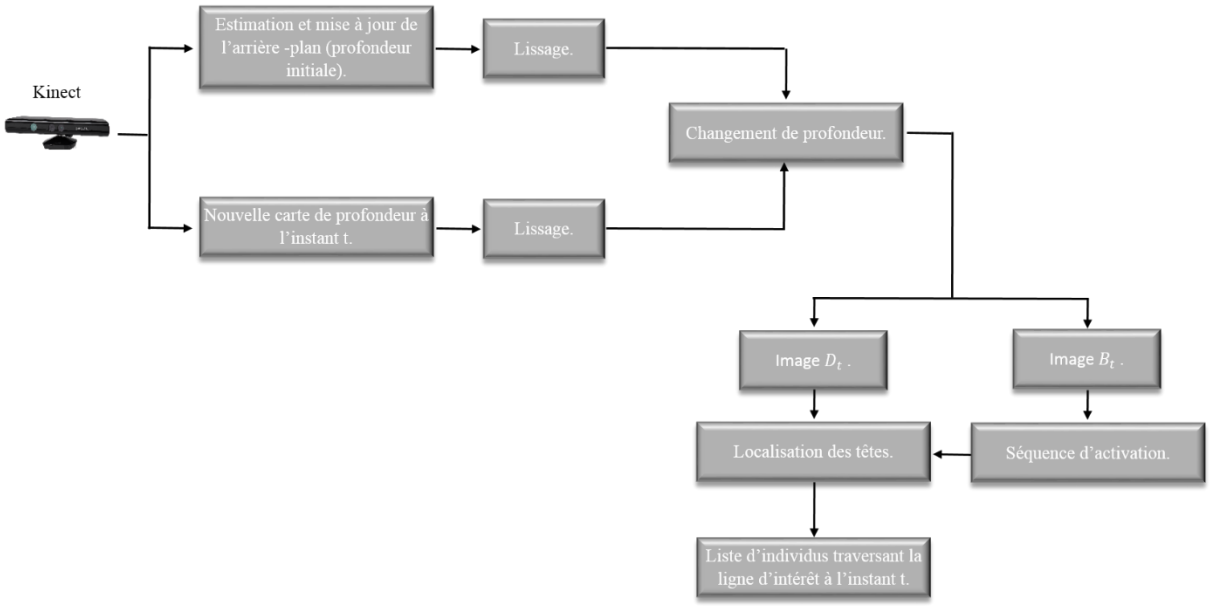


Figure 18. Diagramme de la détection des individus.

Pour détecter les individus dans les cartes de profondeur, nous commençons par la détection des changements de profondeur dans le champ de vision de la kinect. Pour cela, nous utilisons une méthode de soustraction de fond sur les cartes de profondeur. Cette méthode est basée sur l'obtention de la différence lissée entre le frame actuelle et le frame initial de profondeur. Le frame initial de profondeur est obtenu par l'estimation des distances entre le capteur et les objets statiques dans la scène étudiée. L'estimation de ces distances est réalisée par l'utilisation de la moyenne glissante sur un ensemble de N frames obtenus de la kinect. Cette carte de distances est considérée comme l'arrière-plan de la scène étudiée. Elle est initialisée et mise à jour à l'aide de l'équation (1).

$$arrière_plan = \frac{1}{N}\sum_{i=1}^{N} carte_prof_i \qquad (1)$$

Avec : $carte_prof_i$ la carte de profondeur i obtenue de la kinect et N le nombre de cartes utilisées dans l'estimation.

Une fois l'arrière-plan estimé, nous utilisons un filtre Gaussien sur la carte actuelle de profondeur et l'arrière-plan pour lisser et réduire le bruit qui peut être présent dans ces cartes. L'utilisation de ce filtre permet de garder la forme générale de profondeur tout

en réduisant le bruit. Ensuite nous utilisons ces cartes résultantes pour la création de deux images : une image en niveaux de gris D_t et une image binaire B_t.

L'image D_t représente les changements de profondeur à l'instant t, tandis que l'image B_t représente l'avant-plan de la scène au même instant t. Ces deux images sont estimées de l'image de différence $diff_t$ qui représente la différence entre la carte actuelle de profondeur (frame actuelle) et l'arrière-plan. Elle est calculée à l'aide de l'équation (2).

$$diff_t = |frame_t - arrière_plan| \qquad (2)$$

Cette image de différence $diff_t$ est utilisée dans l'estimation des deux images D_t et B_t. Nous commençons tout d'abord par la création de l'image D_t en considérant qu'un centimètre de changement de profondeur correspond à un niveau de gris dans l'image (nous avons fait cette supposition en se basant sur le fait que les individus ne peuvent pas dépasser les 255 cm en hauteur) ; ensuite nous passons à la création de l'image B_t qui représente l'avant-plan de la scène. Cette image est créée en supposant que les pixels de l'image $diff_t$ présents dans l'intervalle (Th_{min} et Th_{max}) font partie de l'avant-plan. Les images D_t et B_t sont créées à l'aide des équations (3) et (4).

$$D_t(x,y) = \begin{cases} diff_t(x,y) & si: diff_t(x,y) < 256 \\ 0 & autrement \end{cases} \qquad (3)$$

$$B_t(x,y) = \begin{cases} 1 & si: Th_{min} < diff_t(x,y) < Th_{max} \\ 0 & autrement \end{cases} \qquad (4)$$

L'étape précédente nous a donc permis de créer les images D_t et B_t qui seront utilisées dans les étapes suivantes pour la détection et le suivi des individus. Une fois ces images créées, nous avons besoin de concevoir une grille qui sera capable de détecter et segmenter le premier plan de la scène afin de pouvoir localiser les têtes des individus traversant la ligne d'intérêt. Nous avons donc conçu une grille rectangulaire autour de la ligne d'intérêt comme montré dans la *Figure 19*.

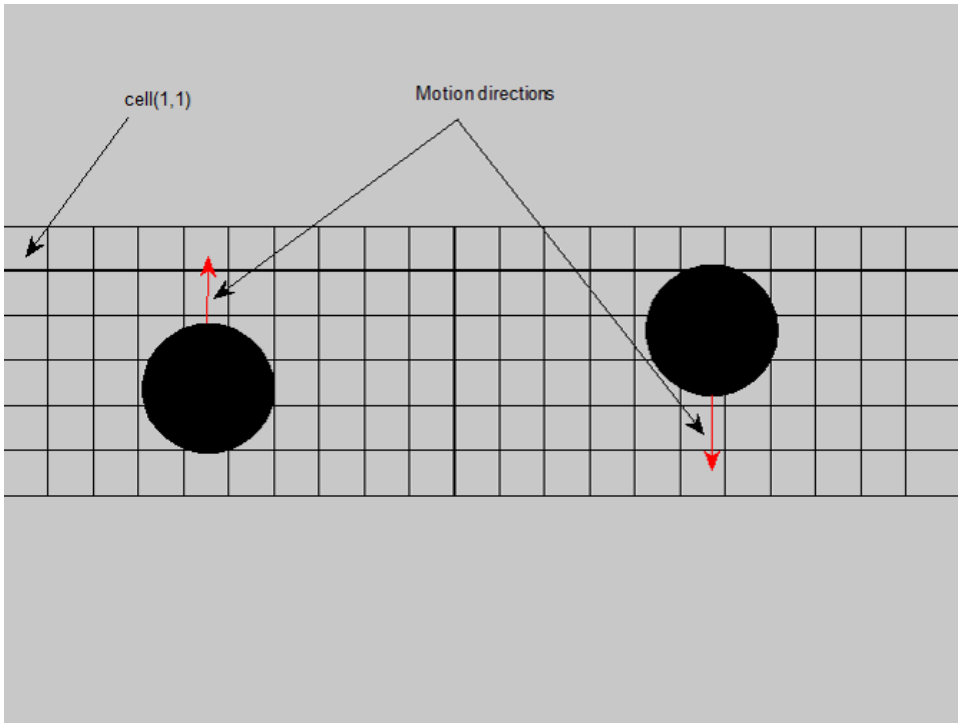


Figure 19. Conception de la grille.

Cette grille a une forme rectangulaire avec une largeur qui correspond au double de la largeur moyenne de la surface occupée par un seul individu détecté dans la scène et une longueur qui correspond à la largeur de la carte de profondeur obtenue par la kinect. Ce rectangle représentant la grille est divisé en plusieurs cellules carrées d'une largeur correspondant au tiers de la largeur moyenne de la surface occupée par un seul individu dans la scène.

Maintenant que nous avons conçu cette grille, la prochaine étape consiste dans la détermination de la séquence d'activation qui sera utilisée dans la détection et la localisation des individus dans la scène. Cela est fait en deux étapes. Dans la première étape, nous utilisons l'image B_t dans la détection des cellules actives de notre grille pour chacune des frames étudiées. Pour cela, nous nous basons sur le nombre N de pixels de premier-plan présents dans chacune des cellules, qui caractérise les cellules actives lorsque celui-ci est supérieur à un seuil Th_{active}. Cette détection est faite à l'aide de l'équation (5).

$$Cellule_{ij} = \begin{cases} 1\ (activée) & si: N > Th_{active} \\ 0\ \ (none\ activée) & autrement \end{cases} \quad (5)$$

Un exemple de la détection de la séquence d'activation est montré dans la *Figure 20*.

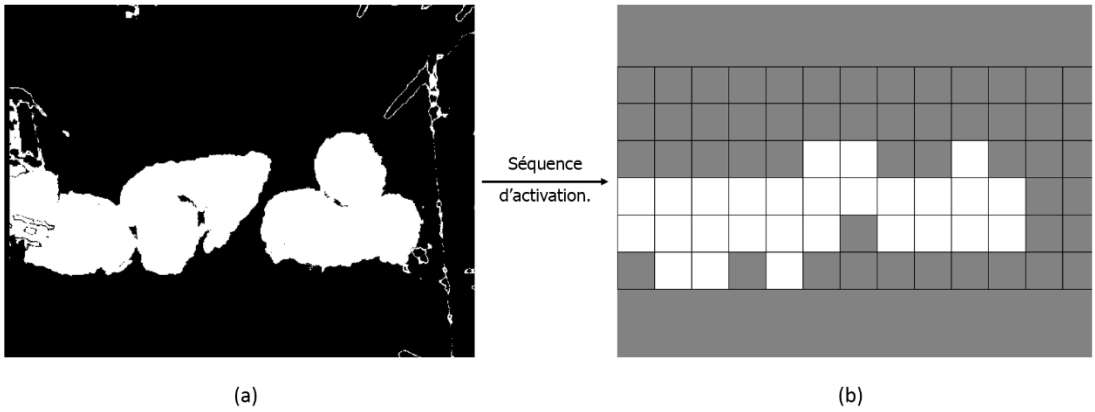


Figure 20. Exemple pour la détection de séquence d'activation : (a) image B_t et (b) séquence d'activation.

Cette étape nous permet de détecter les segments significatifs du premier-plan qui peuvent représentés un seul individu ou un groupe d'individus. Une fois les individus et les groupes d'individus détectés, l'étape suivante consiste alors dans l'utilisation de l'image D_t pour la localisation des individus et pour la segmentation des groupes en individus distincts. Cela est fait par la détection des maximums locaux dans les cellules actives de l'image D_t. Pour cela, nous avons utilisé l'équation (**6**) pour la sommation des pixels formants chacun des cellules actives.

$$cell(i,j) = \begin{cases} \sum_{cellule_{ij}} D_t & si: cellule_{ij}\ est\ activée \\ 0 & autrement \end{cases} \quad (6)$$

Ensuite, une cellule est considérée comme un maximum local si la valeur $cell(i,j)$ est supérieure à toutes les valeurs des cellules voisines $(cell(i-1,j-1);\ cell(i-$

$1, j); \ldots .;\ cell(i+1, j+1))$. Les maximums locaux détectés dans cette étape sont considérés comme étant les têtes des individus traversant la ligne d'intérêt à l'instant t correspondant au frame $frame_t$. Un exemple de cette localisation de têtes est représenté sur la *Figure 21*.

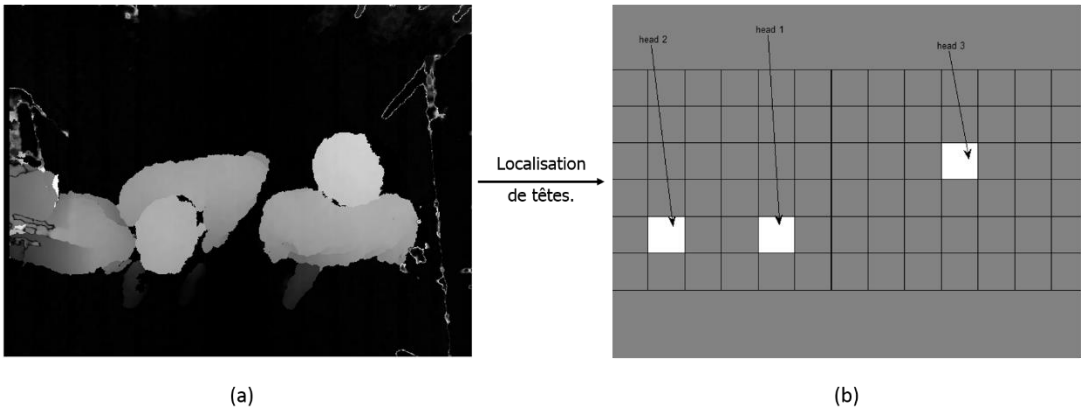


Figure 21. Exemple de détection des têtes dans l'image D_t : (a) image D_t et (b) localisation des têtes.

2.2.2.2. Suivi des personnes

La deuxième phase de notre algorithme de comptage consiste en l'intégration d'un système qui permet le suivi des individus détectés dans la première phase. Ce système de suivi est utilisé pour déterminer la direction du mouvement et pour éviter le recomptage des mêmes individus dans chacun des frames. Pour cela, nous avons utilisé une méthode de suivi composée de plusieurs étapes. Le schéma général des différentes étapes de suivi est décrit dans la *Figure 22*.

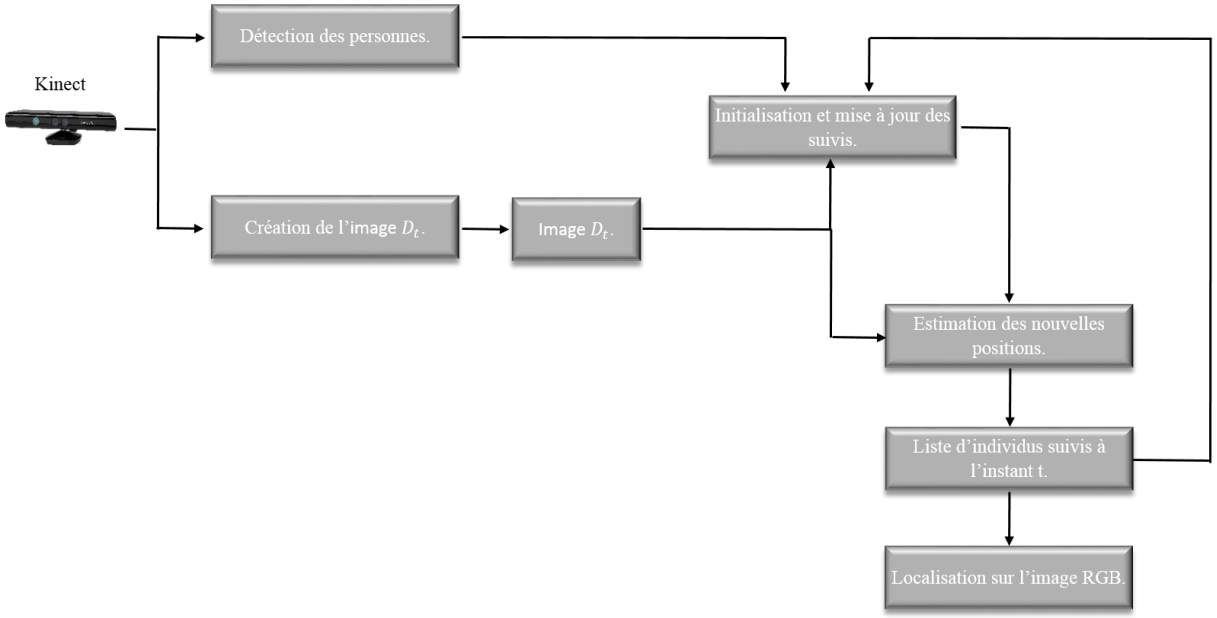


Figure 22. Déférentes étapes de suivi des individus détectés dans le champ de vision de la kinect.

La première étape de notre stratégie de suivi consiste en l'initialisation d'une liste des individus suivis. Cette liste regroupe les modèles des individus suivis obtenus de l'image D_t ainsi que leur histogramme obtenu de cette même image. Ensuite, par l'utilisation d'un modèle cinétique à vitesse constante nous estimons la position de chaque individu à l'instant $t+1$ en se basant sur sa position à l'instant t. Pour cela, les individus détectés sont représentés par des carrés centrés sur le point (x, y) avec une largeur w qui correspond à la largeur moyenne de la surface occupée par un seul individu dans la scène. La nouvelle position de cet individu est estimée par le modèle cinétique représenté par les équations (7) et (8).

$$(x,y)_t = (x,y)_{t-1} + (u,v)_{t-1} * \Delta t \qquad (7)$$

$$(u,v)_t = (u,v)_{t-1} \qquad (8)$$

Avec : (x, y) et (u, v) les positions et les vitesses de l'individu suivi dans les axes des x et des y respectivement.

Cette position approximative est ensuite utilisée comme valeur d'initialisation d'un MeanShift afin d'affiner ce positionnement. A la fin de cette étape de suivi, le point de fin de trajectoire est utilisé pour incrémenter le compteur des entrées ou celui des sorties. Un exemple de suivi est montré dans la *Figure 23*.

Figure 23. Exemple pour le suivi des individus.

2.3. Suivi mono-caméra de personnes

Le suivi de personnes dans une vidéo est défini comme étant la capacité de localiser les individus suivis dans chacun des frames de la vidéo afin d'enregistrer leurs trajectoires. Une illustration des différentes phases d'un système de suivi mono-caméra est montrée dans la *Figure 24*.

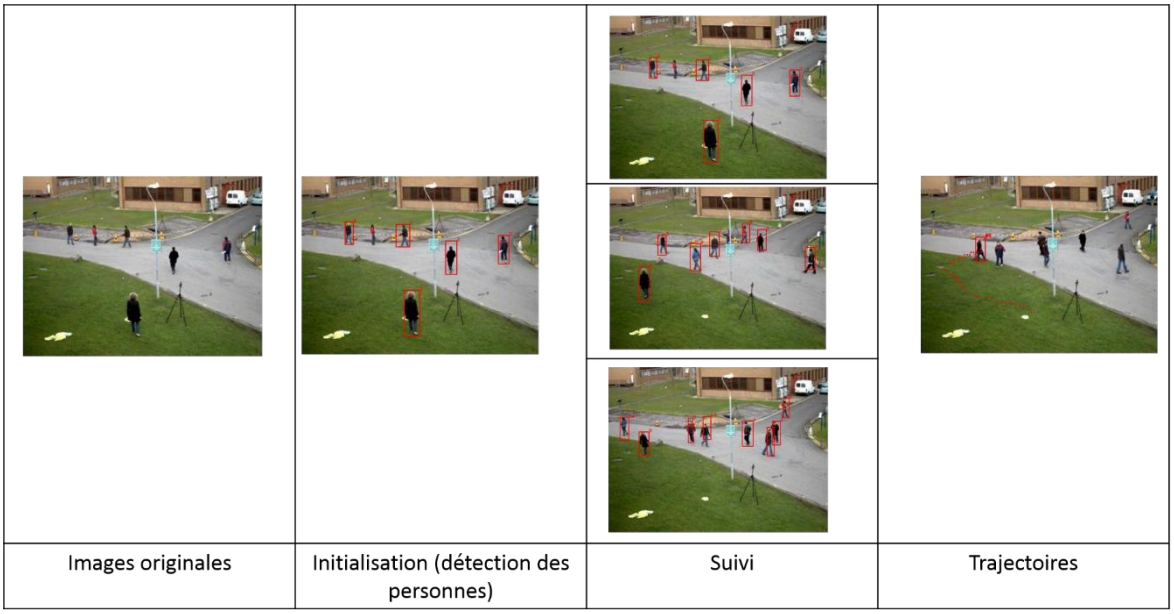


Figure 24. Illustration d'un système de suivi mono-caméra.

Dans un système de suivi mono-caméra, l'objectif principal est l'estimation des différentes positions occupées par l'individu (ou les individus) suivi durant son passage

dans le champ de vision d'une seule caméra. Pour cela, une variété de méthodes a été proposée et toutes rencontrent toutes les mêmes défis de base que sont :

- Les variations d'éclairage.
- Les changements d'échelle et de pose des individus suivis.
- La gestion des occultations.
- La détection de mouvement.
- Les problèmes de segmentation pour isoler les différents individus.
- Les confusions dans l'attribution des identités dans le contexte de suivi de plusieurs individus simultanément.

2.3.1. État de l'art

Durant les dernières années, le suivi mono-caméra était l'un des sujets de recherche les plus actif dans le domaine de la vision par ordinateur. Ces méthodes de suivi peuvent être classées en fonction du nombre d'objets à suivre [82]. Dans cette classification, nous pouvons distinguer les méthodes de suivi par association (ABT) dans lesquelles le nombre d'objets suivis varie avec le temps [83], [84], [85] et [86] et les méthodes CFT (*Category Free Tracking*) dans lesquelles le nombre d'objets suivis est préalablement connu pour l'ensemble des frames de la vidéo [87], [88], [89] et [90].

Les méthodes à base d'association sont conçues pour suivre un type particulier d'objets comme par exemple les piétons. De manière générale, ces méthodes sont basées sur l'utilisation d'un détecteur d'objets capable de reconnaitre une catégorie d'objets. Ce détecteur est utilisé pour chacun des frames de la vidéo afin de localiser les objets suivis ; ensuite, une procédure d'association est utilisée pour trouver les correspondances entre les différents individus dans les différents frames. Ce principe de fonctionnement est illustré dans la *Figure 25*.

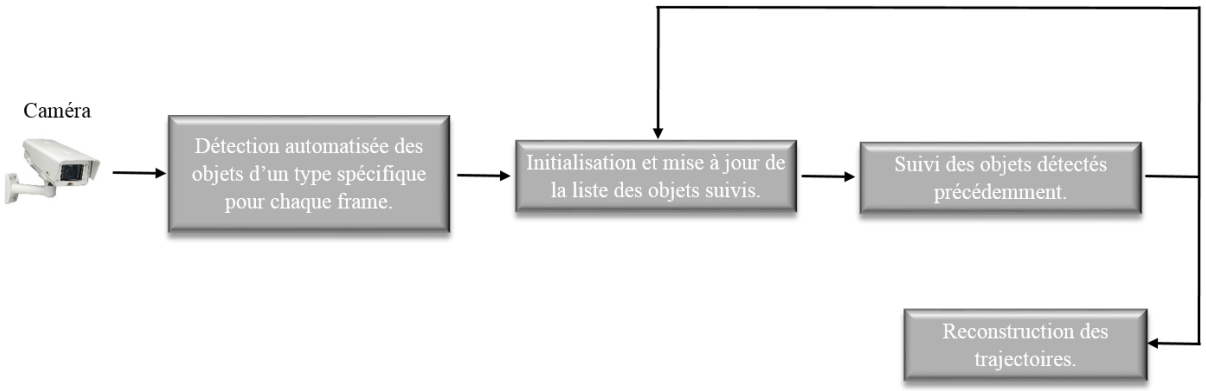


Figure 25. Principe de suivi par association.

Parmi ces méthodes, nous pouvons citer celles basées sur l'optimisation, et notamment celle proposée par Berclaz et al [85]. Leur détecteur d'objets qui produit un plan d'occupation regroupant l'ensemble des positions occupées par tous les objets suivis dans tous les frames de la vidéo. Ensuite ils introduisent un framework dans lequel les fausses détections sont supprimées et les vides causés par la non-détection de certains objets sont remplis. A la fin de cette étape, l'ensemble des détections corrigées sont utilisées avec une optimisation des K-plus courts chemins afin de reconstruire les trajectoires des différents objets. Andriyenko et al [83] et [91] ont formulé le suivi comme une minimisation d'une fonction d'énergie. Ils ont regroupé l'association de données et l'estimation de trajectoires dans la minimisation d'une énergie. Cet algorithme est capable de suivre plusieurs individus et il est basé sur l'utilisation du HOG pour la détection des individus.

Une autre approche utilisée dans cette catégorie de méthodes est la formulation du suivi comme une minimisation dans un champ conditionnel aléatoire (CRF). Dans ce contexte, Yang and Nevatia [92] ont proposé une approche d'apprentissage en ligne pour le suivi multi-objets qui est formulée comme une minimisation dans un modèle de champ conditionnel aléatoire. De la même manière, Milan et al [84] ont combiné un champ conditionnel aléatoire discret avec un autre continu pour exploiter les contraintes d'exclusion dans le suivi. Ils ont exploité le fait que deux détections simultanées ne peuvent pas appartenir à la même trajectoire, et le fait qu'une détection ne peut pas appartenir à deux trajectoires. Leal-Taixe et al [86] ont proposé un modèle de suivi qui prend en compte les interactions entre les individus. Pour cela, ils ont introduit une chaîne qui code l'interaction entre les individus à partir des caractéristiques de l'image. Ce modèle permet de coder l'effet des individus non détectés dans la scène.

La deuxième grande catégorie des méthodes de suivi est celle des méthodes CFT (*Category Free Tracking*). Ces méthodes ne sont pas conçues pour suivre un type particulier d'objets. De ce fait, elles sont généralement basées sur l'initialisation de l'objet à suivre dès sa première apparition dans la scène ; ensuite un algorithme de suivi est utilisé pour déterminer les positions occupées par cet objet dans les frames qui suivent de la vidéo. Ce principe de fonctionnement est illustré dans la *Figure 26*.

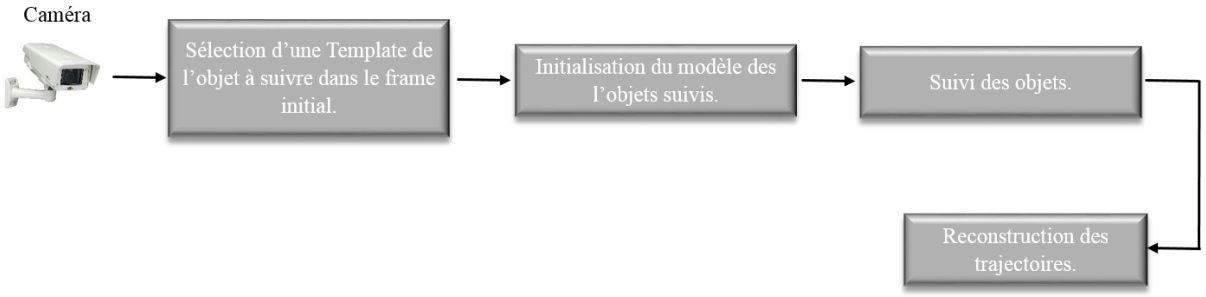


Figure 26. Principe de suivi libre de catégorie.

Zhang et van der Maaten [88, 89] ont exploité les contraintes spatiales pour apprendre un classificateur sur les objets suivis, ainsi que les contraintes structurelles en utilisant un SVM. Ils ont aussi proposé une méthode pour améliorer les détecteurs génériques d'objets qui sont déjà utilisé dans la détection d'un type spécifique d'objets. Yang et al [90] ont proposé une méthode dans laquelle le suivi est formulé comme un jeu de compétition entre les différents trackeurs, et où chaque trackeur essaie de maximiser son évidence visuelle, le problème de suivi étant résolu par la recherche de l'équilibre du jeu. Shi et Karl [93] ont implémenté un modèle basé sur les régions en exploitant les résultats de suivi dans les frames précédentes pour initialiser une courbe, l'utilisation d'une fonction de minimisation d'énergie permettant la localisation des objets suivis dans le frame actuelle de la vidéo. Toujours dans la même catégorie, Hu et al [87] ont proposé un algorithme de suivi avec un raisonnement d'occultations. Dans cette méthode, les objets suivis sont divisés en plusieurs blocs pour lesquels des modèles de sous-espaces riemanniens log-euclidiens de faible dimension sont appris. Ces blocs permettent de mettre à jour l'apparence des objets suivis, de sorte que seuls les blocs apparents sont mis à jour. Cette stratégie permet de mettre à jour les apparences pour les objets partiellement occultés.

L'un des problèmes majeurs dans le suivi mono-caméra des personnes est la récupération des identités des individus suivis après une occultation de longue durée. Pour cela, plusieurs travaux ont été effectués dans ce cadre pour gérer les occultations dans un système de suivi multi-objets. Parmi ces méthodes, nous pouvons citer [94], [95], [96] et [97]. Bae et Yoon [94] ont proposé une méthode permettant sur la gestion des occultations et la gestion des apparences similaires qui peuvent induire à des associations erronées individus. Ils ont ainsi développé une méthode basée sur la certitude des tracklets qui est associée à un modèle d'apparence appris en ligne pour distinguer les apparences similaires. Cette méthode est utilisée pour l'association des tracklets appartenant au même individu. Chen et al [95] ont également proposé un système de suivi mono-caméra capable de gérer les occultations partielles et les occultations de longue durée. Leur système est basé sur la création d'une séquence annotée de contraintes qui est utilisé avec les vecteurs de caractéristiques des différents individus suivis, ceci afin de pouvoir éviter les confusions entre les identités des différentes personnes. Toujours dans le cadre de gestion des occultations, Izadinia et al [97] ont implémenté un algorithme dans lequel le suivi de personnes est effectué en trois étapes. Dans la première étape, les différentes détections d'individus sont associées afin d'avoir des tracklets. Dans la seconde, une autre association est réalisée mais cette fois entre les différentes parties des corps des individus pour corriger les associations semblant erronées parmi celles effectuées dans la première étape. Dans la troisième étape, une association simultanée entre l'ensemble des tracklets est effectuée pour obtenir les trajectoires. Tang et al [96, 98] ont exploité le fait que la plupart des occultations dans un système de suivi apparaissent entre les individus suivis. Par conséquent, ils ont créé un détecteur de paires de personnes qui est capable de détecter deux personnes s'occultant mutuellement. Ils ont intégré ce système de détection de paires de personnes aux stratégies de suivi proposées par Andriyenko and Schindler [83] and Pirsiavash et al [99] afin de gérer les occultations.

2.3.2. Stratégie adoptée

Le deuxième module que nous avons implémenté et qui est présenté dans ce chapitre et le module de suivi mono-caméra de personnes. Ce dernier est conçu pour la récupération des segments des trajectoires parcourues par les individus suivis durant leurs passages dans le champ de vision de chacune des caméras indépendamment des autres. Nous avons implémenté une méthode de suivi par filtre particulaire, similaire à celui proposé par Nummiaro et al [100]. Notre système de suivi mono-caméra est conçu en trois phases principales. La première phase consiste dans la détection des individus pour l'initialisation des trackeurs. Cette phase est composée d'une étape de détection de

mouvement pour la localisation des objets en mouvement dans la scène et d'une étape de segmentation des blobs obtenus précédemment pour isoler les différents individus. Ensuite vient la deuxième phase qui consiste à suivre les individus détectés pour la récupération des segments des trajectoires. Cette phase nous permet de récupérer des petits segments de trajectoire avec une grande certitude d'appartenir au même individu. Une fois ces petits segments de trajectoires récupérés, la troisième phase consiste dans l'utilisation d'une méthode de ré-identification afin d'associer ces segments et d'obtenir des tracklets qui représentent les trajectoires parcourues par chacun des individus. Le schéma de base de l'implémentation adoptée est illustré par la *Figure 27*.

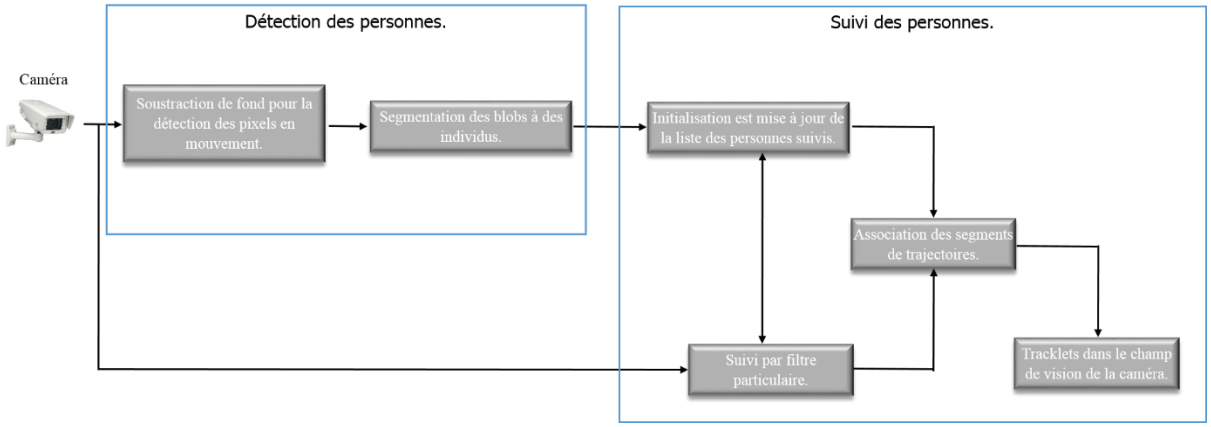


Figure 27. Implémentation du module de suivi mono-caméra.

2.3.2.1. Détection des personnes

Comme nous l'avons mentionné précédemment, la première phase de notre implémentation consiste dans la détection des personnes et l'initialisation des trackeurs.

Pour la détection des mouvements, nous avons commencé par utiliser un algorithme de soustraction de fond pour récupérer l'avant-plan de la scène. Cet avant plan représente les parties en mouvement dans le frame actuel de notre vidéo. Pour cela, nous avons utilisé un algorithme de soustraction de fond similaire à celui proposé par Shimada et al [101] et basé sur l'utilisation de mélange de gaussiennes. L'utilisation d'une méthode basée sur un mélange de gaussiennes permet de réduire les effets des changements brusques d'éclairage. Cette méthode de soustraction de fond repose sur l'utilisation de K distributions gaussiennes pour chaque pixel de l'image. Ensuite, chaque pixel du nouveau

frame est comparé à l'ensemble des distributions jusqu'à ce qu'une correspondance soit trouvée. Si cette distribution correspondante est l'une des distributions des modèles de l'arrière-plan, alors le pixel est classé comme un pixel d'arrière-plan, sinon il est classé comme un pixel de l'avant-plan. Cette procédure est répétée pour l'ensemble des pixels de l'image. A la fin de cette étape, nous obtenons une image binaire représentant l'avant plan de la frame actuelle de la vidéo. Un exemple des résultats de cette procédure de soustraction de fond est montré dans la *Figure 28*.

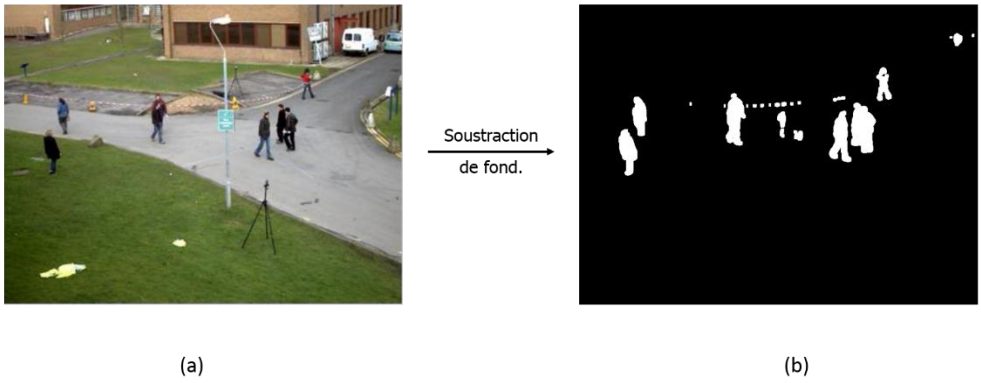


Figure 28. Exemple de résultats de soustraction de fond.

Une fois l'avant-plan de la scène récupéré, la prochaine étape consiste dans la segmentation de cet avant-plan pour isoler les individus. Pour cela, nous avons utilisé une étude par composante connexes pour délimiter les blobs de premier-plan qui représentent des individus isolés ou des groupes d'individus qui sont trop proche les uns des autres. Nous avons ajouté à cette analyse par composantes connexes une contrainte sur la largeur et la hauteur des blobs correspondants à un seul individu. Pour ce faire, nous avons représentés les blobs par des rectangles centrés autour du point (x_0, y_0) avec une largeur w et une hauteur h ; les propriétés du rectangle sont calculées à l'aide des équations (9), (10) et (11).

$$(x_0, y_0) = \left(\frac{x_{min} + x_{max}}{2}, \quad \frac{y_{min} + y_{max}}{2}\right) \qquad (9)$$

$$w = x_{max} - x_{min} \qquad (10)$$

$$h = y_{max} - y_{min} \quad (11)$$

Tous les blobs obtenus par cette procédure sont représentés par des rectangles. La prochaine étape, consiste dans l'utilisation d'une restriction sur le ratio entre la largeur et la hauteur de chaque blob. Si ce ratio se trouve dans l'intervalle $[Th_{min}, Th_{max}]$, le blob est considérer comme représentant un seul individu (voir équation (12)).

$$N_{ind} = \begin{cases} 1 & si: Th_{min} < \frac{w}{h} < Th_{max} \\ round\left(\frac{w}{h * \left(\frac{Th_{min} + Th_{max}}{2}\right)}\right) & si: \frac{w}{h} > Th_{max} \\ round\left(\frac{h * \left(\frac{Th_{min} + Th_{max}}{2}\right)}{w}\right) & si: \frac{w}{h} < Th_{min} \end{cases} \quad (12)$$

Dans le cas où le ratio de largeur et hauteur ne se trouve pas dans l'intervalle, le nombre d'individus représentés par les blobs est estimé en se basant sur le même ratio. Cette estimation nous permet de segmenter les blobs en plusieurs individus. Les positions et les dimensions des boites englobantes (rectangles représentant les individus) sont calculés à l'aide des équations (13), (14), (15) et (16) pour un ratio $\frac{w}{h} > Th_{max}$ et un nombre d'individus $N_{ind} = 2$.

$$(x_{01}, y_{01}) = \left(\frac{x_{min} + x_{max}}{4}, \frac{y_{min} + y_{max}}{2}\right) \quad (13)$$

$$(x_{01}, y_{01}) = \left(\frac{3 * (x_{min} + x_{\max})}{4}, \frac{y_{min} + y_{max}}{2}\right) \quad (14)$$

$$w = \frac{x_{max} - x_{min}}{2} \quad (15)$$

$$h = y_{max} - y_{min} \quad (16)$$

Des exemples de résultats obtenus par cette segmentation sont illustrés dans les *Figure 29* et *Figure 30*.

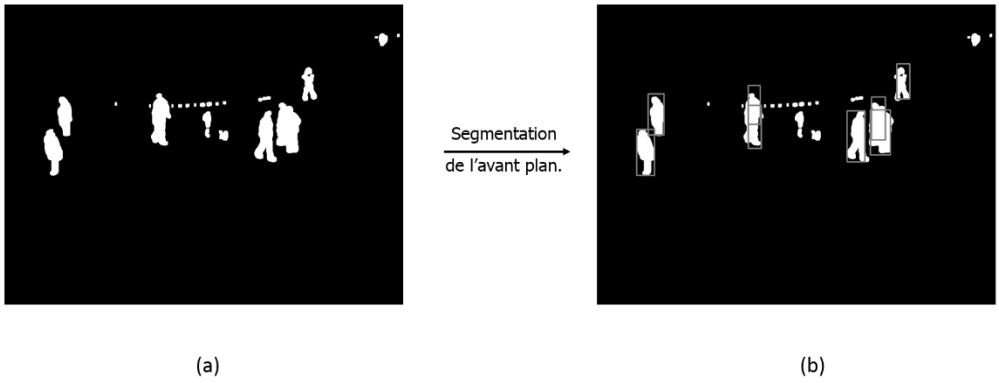


Figure 29. Exemple de résultats de soustraction de fond après la segmentation.

Figure 30. Exemple de détection de personnes.

2.3.2.2. Suivi des personnes

La phase de détection nous a permis de localiser les individus présents dans la scène à un instant donné t_0 correspondant à un frame de la vidéo. La liste des individus détectés est utilisée pour initialiser la liste des individus suivis ; ensuite ces individus sont suivis dans cette phase pour récupérer des fragments de trajectoires (tracklets) pour chaque individu.

Pour récupérer ces fragments de trajectoires, nous avons utilisé une stratégie de suivi basée sur l'utilisation d'un filtrage particulaire similaire à celui proposé par Nummiaro et al [100]. Nous avons associé un nombre N de particules pour chaque individu suivi et chaque individu est représenté par un rectangle qui est centré autour du point (x_0, y_0) avec une largeur w et une hauteur h. Pour estimer la nouvelle position de l'individu suivi à l'instant t en se basant sur sa position à l'instant $t-1$, nous avons adopté un modèle cinétique qui considère que les individus suivis se déplacent avec une vitesse constante comme montrer par les équations (17) et (18).

$$(x, y)_t = (x, y)_{t-1} + (u, v)_{t-1} * \Delta t \qquad (17)$$

$$(u, v)_t = (u, v)_{t-1} \qquad (18)$$

Avec : (x, y) et (u, v) la position et la vitesse de l'individu suivi sur le plan x-y.

Ensuite, ces particules sont pondérées par rapport à la différence entre leurs histogrammes de couleur et celui de l'individu suivi dans l'espace de couleur HSV à l'aide de la fonction de vraisemblance donnée par l'équation (19).

$$P_{(n)} = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{d^2}{2\sigma^2}} \qquad (19)$$

Avec : d la distance de Bhattacharyya entre les histogrammes et $\boldsymbol{\sigma}$ la variance de la distribution gaussienne.

Les résultats de cette fonction de vraisemblance sont utilisés dans le calcul des nouveaux poids de pondération pour l'ensemble des particules. Ces poids sont tout d'abord utilisés dans le ré-échantillonnage des particules qui se base sur la substitution des particules de plus faibles poids par d'autres particules. Ensuite ils sont utilisés dans l'estimation de la nouvelle position de la personne suivi à l'instant t en se basant sur les positions et les poids des différentes particules. Cette estimation est réalisée à l'aide de l'équation (20).

$$\begin{bmatrix} x \\ y \end{bmatrix} = \sum_{n=0}^{N} w_t^{(n)} \begin{bmatrix} x \\ y \end{bmatrix}^{(n)} \qquad (20)$$

Avec (x, y) la position estimée pour l'individu suivi à l'instant t, $w_t^{(n)}$ le poids de la particule située à la position $\begin{bmatrix} x \\ y \end{bmatrix}^{(n)}$, et N le nombre de particules utilisées dans notre filtre particulaire.

2.3.2.3. Association des segments de trajectoires

Les segments de trajectoires calculés précédemment ne représentent pas les trajectoires complètes parcourues par les individus suivis. Cela est dû à la perte de suivi et la création de nouvelles identités pour les individus perdus. Cette perte est causée soit par les occultations de longue durée soit par la none convergences du filtre particulaire lors de suivis de longues durées (nombre insuffisant de particules).

Pour résoudre ce problème, nous avons intégré un système de ré-identification dont l'objectif est d'associer les segments de trajectoires afin d'obtenir les tracklets qui représentent l'ensemble des positions des individus suivis. Pour cette tâche, nous avons utilisé le même module de ré-identification que nous avons proposé dans le cadre de la ré-identification multi-caméras (ce module de ré-identification est détaillé dans le chapitre 3).

Un exemple de résultat de suivi est illustré dans la *Figure 31*.

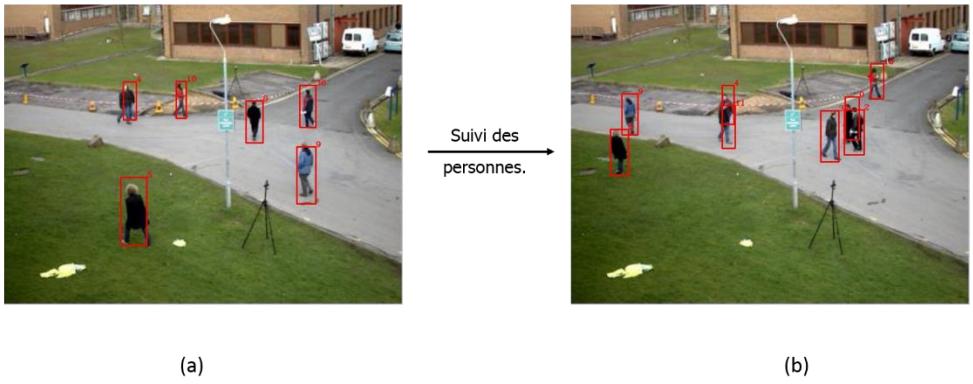


Figure 31. Exemple de résultats de suivi mono-caméra.

2.4. Evaluation

2.4.1. Comptage de personnes

Pour l'évaluation de notre algorithme de comptage, nous avons utilisé une kinect V1. Nous avons installé la kinect au plafond dans une position zénithale par rapport à la direction de mouvement. Les expériences ont été réalisées dans le laboratoire. Une illustration de cette vue est montrée dans la *Figure 32*.

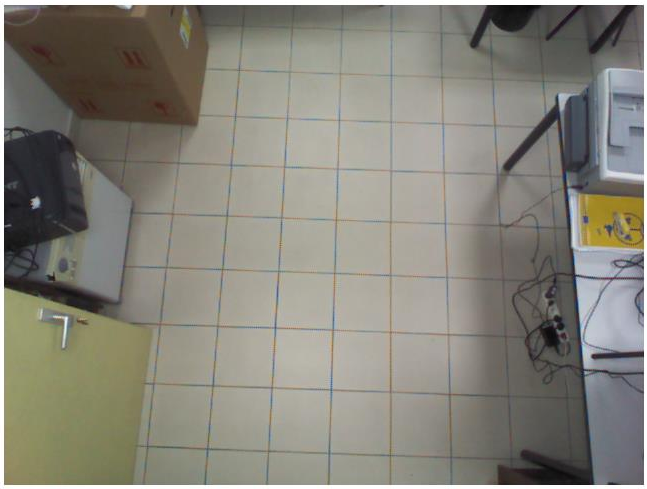


Figure 32. Illustration de la scène observée dans le champ de vision de la kinect.

Durant les expériences, cinq personnes rentraient et sortaient de la salle observée selon des scénarios caractéristiques des différents cas de figure qui peuvent se produire dans la réalité (personnes qui se croisent et plus ou moins proches). Nous avons enregistré plusieurs vidéos qui durent entre 2 et 5 minutes chacune avec 6 à 12 frames par second.

Dans la première série de tests, nous avons évalué la performance de notre algorithme de détection des personnes traversant la ligne d'intérêt dans le champ de vision de la kinect. Nous avons utilisé 100 frames dans l'estimation et la mise à jour de notre modèle d'arrière-plan. Nous avons créé les images D_t et B_t en prenant $Th_{min} = 50cm\ et\ Th_{max} = 255\ cm$ et nous avons utilisé ces images avec une grille de largeur 240 pixels et avec des cellules de largeur 40 pixels. Des images illustrant les résultats de détection sont présentées dans les figures (*Figure 33* et *Figure 34*).

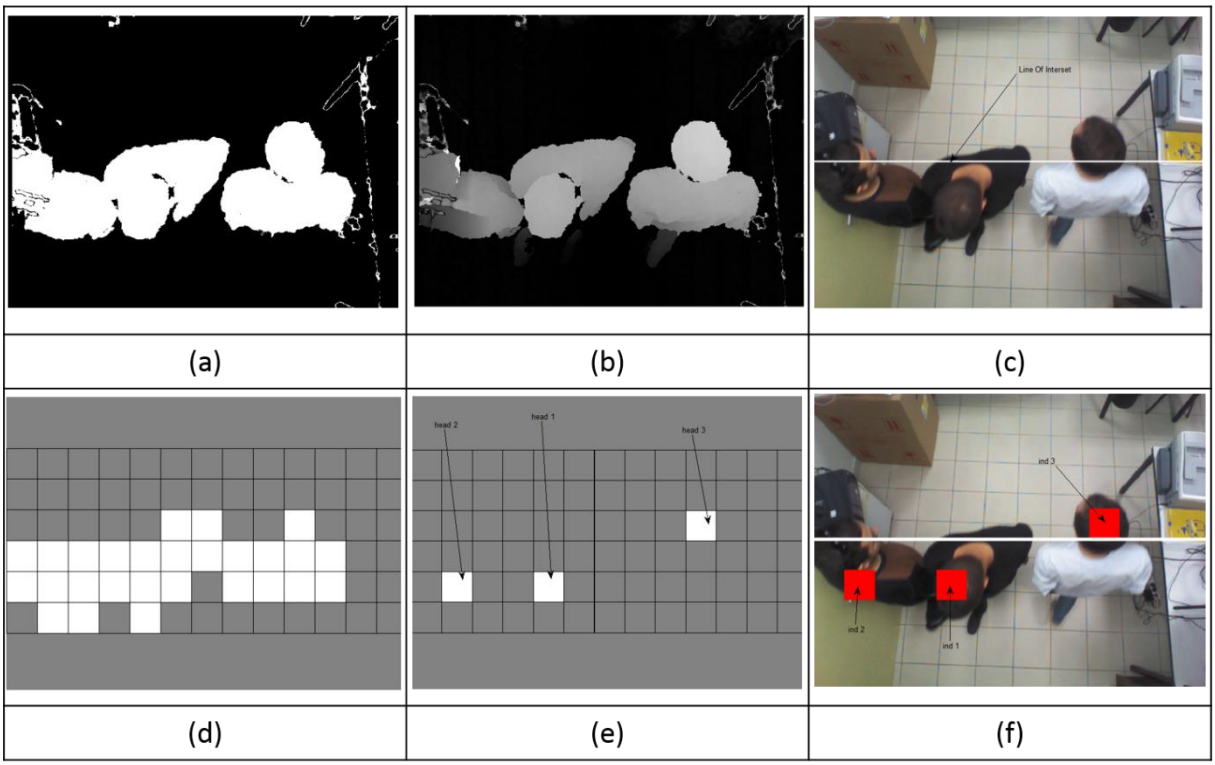


Figure 33. Résultats de détection des personnes : (a) image B_t, (b) imageD_t, (c) image RGB, (d) séquence d'activation, (e) localisation des têtes et (f) localisation des têtes dans l'image RGB.

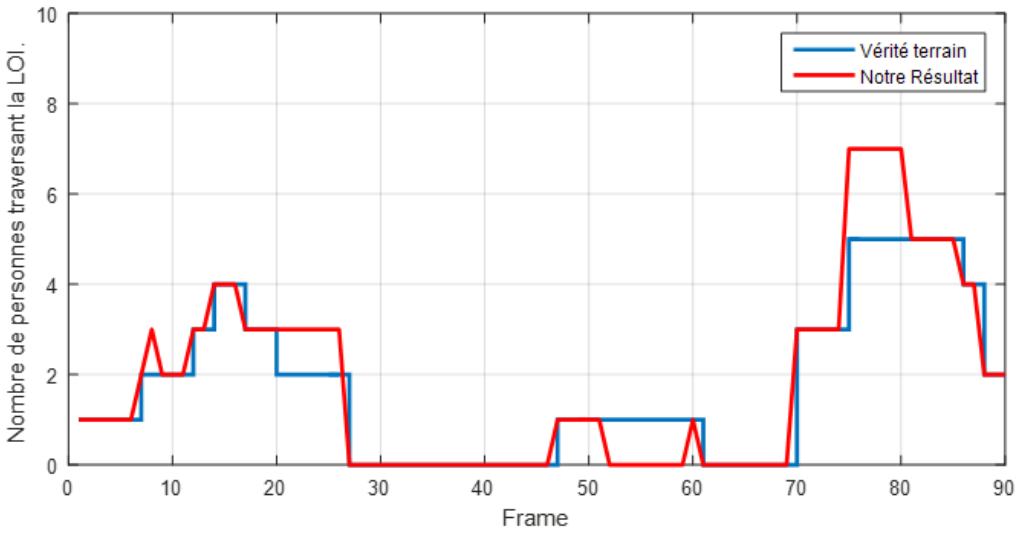


Figure 34. Résultats de détection (comparaison avec la vérité terrain).

La *Figure 33* représente un exemple des différentes étapes de détection : la première colonne illustre l'image binaire B_t ainsi que la séquence d'activation résultant de cette

image, la deuxième colonne illustre l'image D_t ainsi que la localisation de têtes correspondante à cette image ; ensuite vient la troisième colonne qui illustre la mise en correspondance de la précédente localisation dans l'image RGB correspondante dans la vidéo enregistrée.

La deuxième figure (*Figure 34*) illustre une comparaison entre les résultats de détection obtenus par notre algorithme et la vérité terrain. Cette figure montre la comparaison sur une fenêtre de donnée (entre les frames 200 et 300 de notre première vidéo préenregistrée). La comparaison est faite frame par frame. La figure montre que nous arrivons à détecter plus de 95% des individus traversant la ligne d'intérêt pour cette fenêtre de données.

Ces résultats obtenus montrent que notre algorithme de détection de personnes, basé sur l'utilisation d'une grille avec les cartes de profondeur obtenues d'une kinect, est capable de détecter les individus dans les différents scénarios testés. Nous avons fait les tests avec des personnes marchant les uns proches des autres allants dans la même direction ou se croisant. La comparaison de ces résultats avec la vérité terrain indique une détection de plus de 95% des individus traversant la ligne d'intérêt.

Dans la deuxième série de tests, nous avons évalué notre stratégie de suivis basée sur l'utilisation d'un modèle cinétique associé à un MeanShift. Cette stratégie est utilisée avec l'image D_t pour suivre les individus détectés pour déterminer la direction de leurs mouvements et éviter leur recomptage pour chaque frame de données. Les figures (*Figure 35* et *Figure 36*) montrent quelques résultats obtenus par cette stratégie de suivi avec différents scénarios.

Figure 35. Résultats de suivi des personnes : (a) frame 240, (b) frame 258.

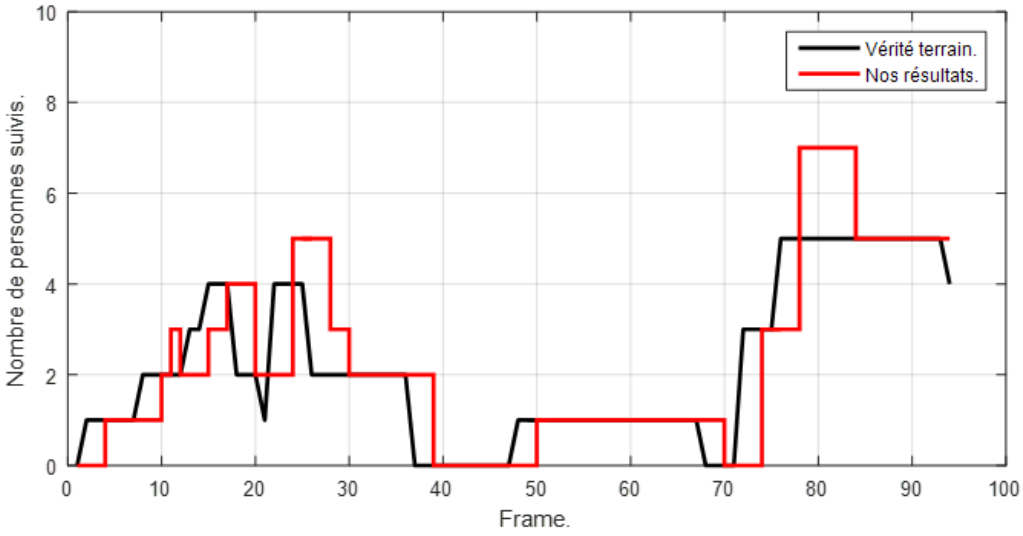


Figure 36. Résultats de suivi (comparaison avec la vérité terrain).

La *Figure 35* représente un exemple de suivi effectué avec trois individus qui se croisent en traversant la ligne d'intérêt : la première colonne illustre l'initialisation de suivi avec la détection des individus au moment où ils traversent la ligne d'intérêt ; ensuite vient la deuxième colonne qui illustre la localisation de ces mêmes individus avant qu'ils quittent le champ de vision de la kinect. Ces résultats sont illustrés sur les images RGB dans la vidéo préenregistrée.

La deuxième figure (*Figure 36*) illustre une comparaison entre les résultats de suivi obtenus par notre stratégie de suivi et la vérité terrain. Cette figure montre la comparaison sur une fenêtre de donnée (entre les frames 200 et 300 de notre première vidéo préenregistrée). La comparaison est faite frame par frame. La figure montre que nous arrivons à suivre plus de 90% des individus traversant la ligne d'intérêt pour cette fenêtre de données.

Ces résultats obtenus montrent que cette méthode de suivi est capable de suivre les individus dans le champ de vision de la kinect pour nos différents scénarios. La comparaison de nos résultats avec la vérité terrain indique un suivi de plus de 90% des individus traversant la scène.

Les résultats de détection et de suivi précédemment obtenus, nous ont permis d'incrémenter les compteurs d'entrées et de sorties. La comparaison des résultats de comptage avec la vérité terrain est montrée dans le graphe de la *Figure 37*.

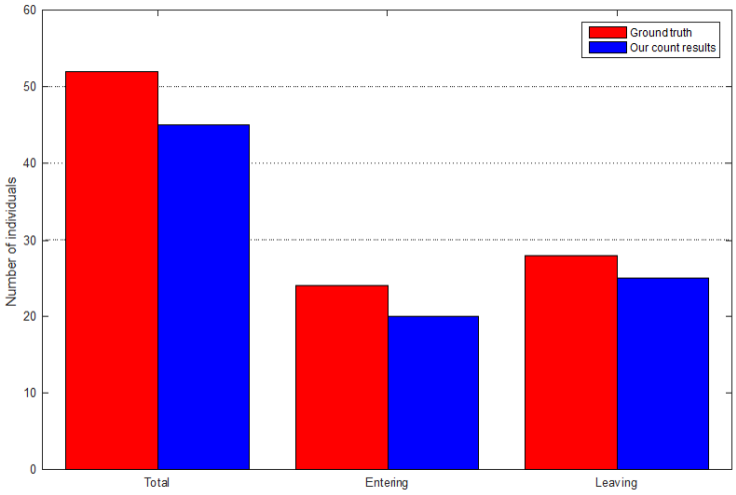


Figure 37. Résultats du comptage d'entrées et de sorties.

La *Figure 37* illustre une comparaison entre nos résultats de comptage et la vérité terrain. Elle montre le comptage des personnes qui rentrent et qui sortent de la zone surveillée, ainsi que la vérité terrain correspondante. Ces résultats sont obtenus sur l'ensemble de trois vidéos préenregistrées d'une durée d'environ 5 minutes.

Le graphe montre que notre comptage d'entrées et de sorties est très proche de la vérité terrain. Nous avons eu une erreur de moins de 5% sur l'ensemble des entrées et sorties.

Ces résultats nous ont donc permis de valider notre algorithme de comptage basé sur l'utilisation d'une grille avec les cartes de profondeur obtenues d'une kinect installée au plafond avec une position zénithal par rapport à la direction de mouvement.

2.4.2. Suivi mono-caméra de personnes

Pour évaluer notre module de suivi mono-caméra, nous avons réalisé plusieurs séries de tests. Pour cela, nous avons utilisé la base de données PETS 2009 qui est l'une des bases les plus utilisées dans l'évaluation d'algorithmes de suivi de personnes. Cette base est composée de plusieurs séquences d'images représentants des individus marchant selon plusieurs scénarios. Un exemple d'images de cette base est illustré dans la *Figure 38*.

Figure 38. Illustration de séquence d'image de la base PETS 2009 (S2.L1 walking) :(a) frame 11 et (b) frame 25.

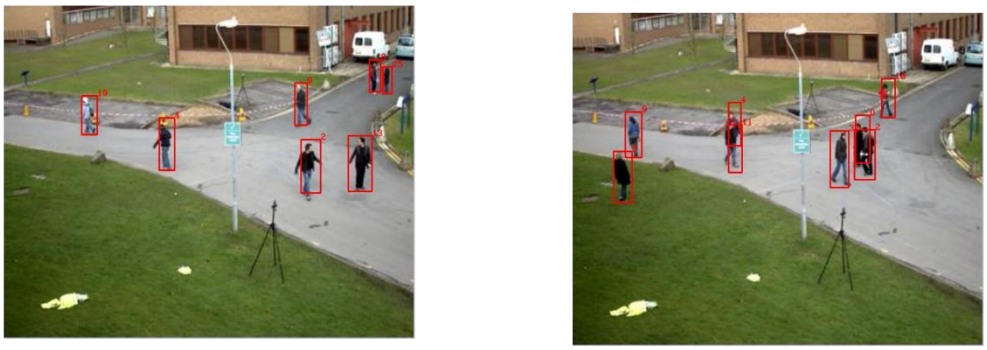
Dans la première série de tests, nous avons utilisé ces données pour évaluer notre stratégie de détection de personnes. Cette détection est utilisée pour l'initialisation et la mise à jour de la liste des personnes suivis pour chacun des frames de la vidéo. Les résultats obtenus sont présentés dans la *Figure 39*.

Figure 39. Résultats de soustraction de fond et de segmentation des personnes (frame 190 et frame 725): (a) image original, (b) soustraction de fond, (c)segmentation.

La *Figure 39* montre des images représentant différentes étapes de notre stratégie de détection de personnes : la première colonne montre les frames traités, la deuxième colonne montre les résultats de soustraction de fonds corresponds et la troisième colonne illustre les résultats de segmentation correspondants.

Ces résultats montrent l'efficacité de notre stratégie de soustraction de fonds et de segmentation que nous avons adoptée. Elle permet la bonne segmentation des blobs pour différents scénarios (individus qui se croisent et individus qui marchent ensemble).

Les résultats de segmentation sont interprétés sur les images RGB pour visualiser les résultats de la détection de personnes. Ces résultats de détection de personnes sont présentés dans la *Figure 40*.



(a) (b)

Figure 40. Résultats de la détection des personnes : (a) frame 190 et (b) frame 725.

La *Figure 40* représente l'interprétation des résultats de soustraction de fonds et de segmentation montrés précédemment sur les images RGB. On peut constater la bonne interprétation de la soustraction de fonds pour la localisation et la délimitation des boites englobantes des différents individus dans les différentes images.

Pour évaluer la performance de notre algorithme de détection de personnes, nous avons effectué une comparaison de nos résultats de détection avec la vérité terrain sur les séquences d'images obtenues de la base de données. Un exemple de résultats de cette comparaison est illustré dans la *Figure 41*.

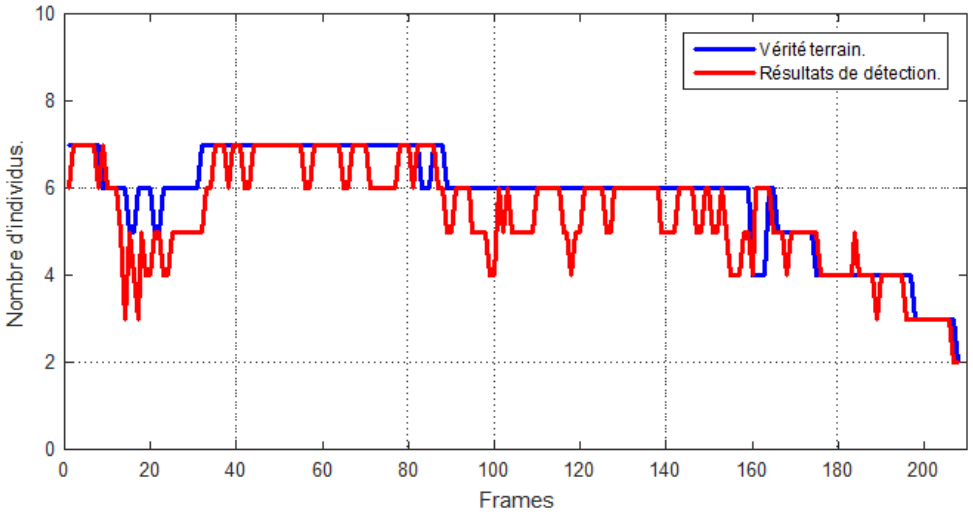


Figure 41. Résultats de détection (comparaison avec la vérité terrain).

Les graphes de la représentant une comparaison entre nos résultats de détection et la vérité terrain pour une séquences d'images de la base. Le graphe montre cette comparaison pour une fenêtre de données (entre les frames 200 et 410) sur la séquence S2 L1.

Cette comparaison entre nos résultats et la vérité terrain montre que notre stratégie de détection de personnes est capable de détecter les individus dans les différents scénarios qui peuvent survenir dans la réalité. Les résultats montrent que nous arrivons à détecter plus de 90% des individus en moyenne (93.4% pour la séquence de la *Figure 41*) ce qui un résultat très satisfaisant.

La deuxième série de tests est utilisée pour l'évaluation de la stratégie de suivi par filtre particulaire. Pour ce faire, nous avons utilisé les résultats obtenus dans la première série de tests comme initialisation pour les différents trackeurs. Ensuite, pour chaque individu suivi, nous avons associé un filtre particulaire basé sur l'utilisation d'histogrammes de couleur dans l'espace de couleur HSV.

Cette deuxième série de tests nous a permis de valider la stratégie de suivi par filtre particulaire pour l'obtention des segments de trajectoires. Un exemple de résultats de suivi est montré dans *Figure 42*.

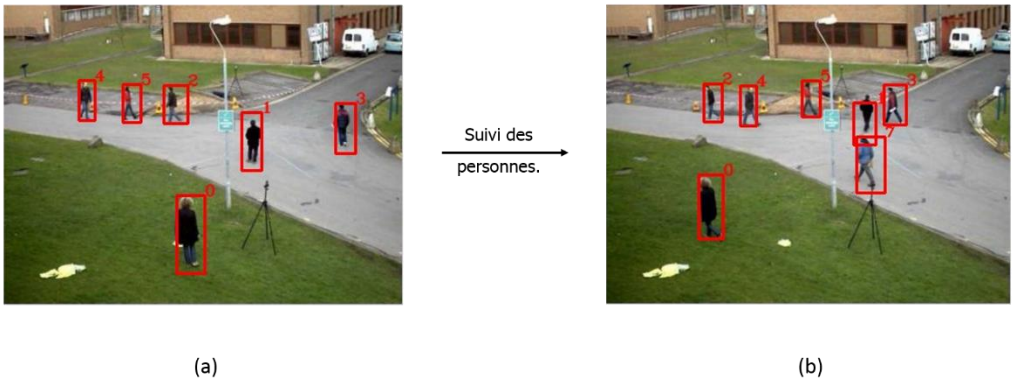


Figure 42. Exemple de suivi de personnes par filtre particulaire : (a) frame 650 et (b) frame 680.

Ces résultats montrent que l'utilisation d'un filtre particulaire permet le suivi de personnes sur une courte durée. Cela nous donne donc la possibilité de récupérer les segments de trajectoires pour chaque individu suivi.

Pour l'évaluation de suivi, nous avons comparé nos résultats de suivi avec la vérité terrain. Un exemple des résultats obtenus est illustré dans la *Figure 43*.

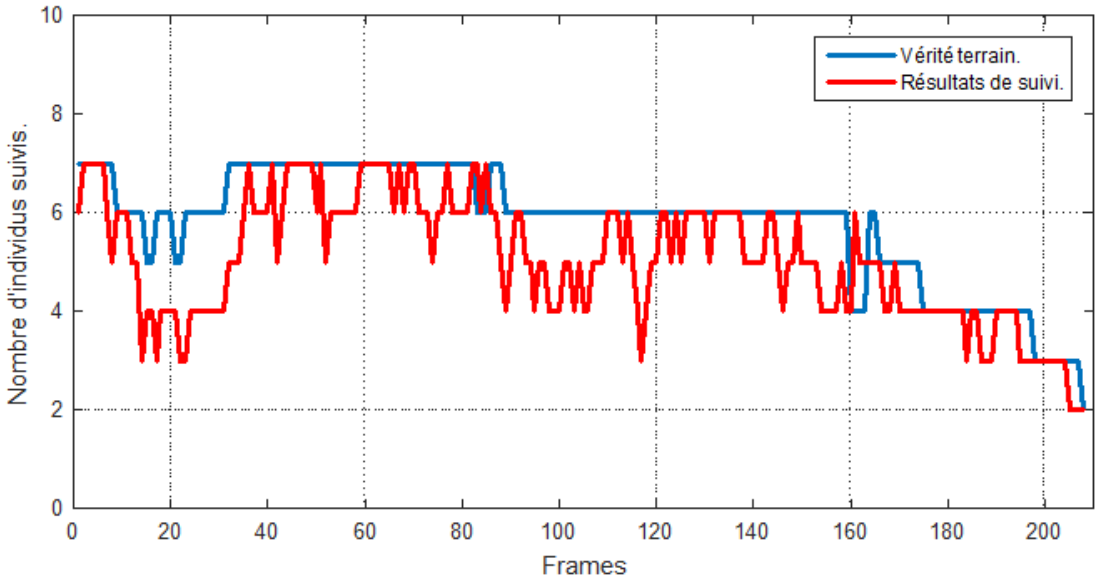


Figure 43. Résultats de suivi (comparaison avec la vérité terrain).

La *Figure 43* représente une comparaison entre nos résultats de suivi et la vérité terrain pour la séquence de données (entre les frames 200 et 410) sur la séquence S2 L1.

Cette comparaison montre que pour les courtes durées, un suivi par filtre particulaire donne des résultats satisfaisants. Le suivi par filtre particulaire, nous permet de suivre correctement environ 80% des individus en moyenne (86.6% pour les résultats illustrés par la *Figure 43*.)

La troisième série et dernière série de tests est réalisée dans le but de tester notre stratégie d'association de segments de trajectoires pour obtenir des tracklets qui représentent les parcours des différents individus. Un exemple de résultats de correction de parcours est représenté dans la *Figure 44* et plus de résultats sur l'utilisation de notre méthode de ré-identification dans la gestion des occultations dans le cadre du suivi mono-caméra sont montrés dans le chapitre 3.

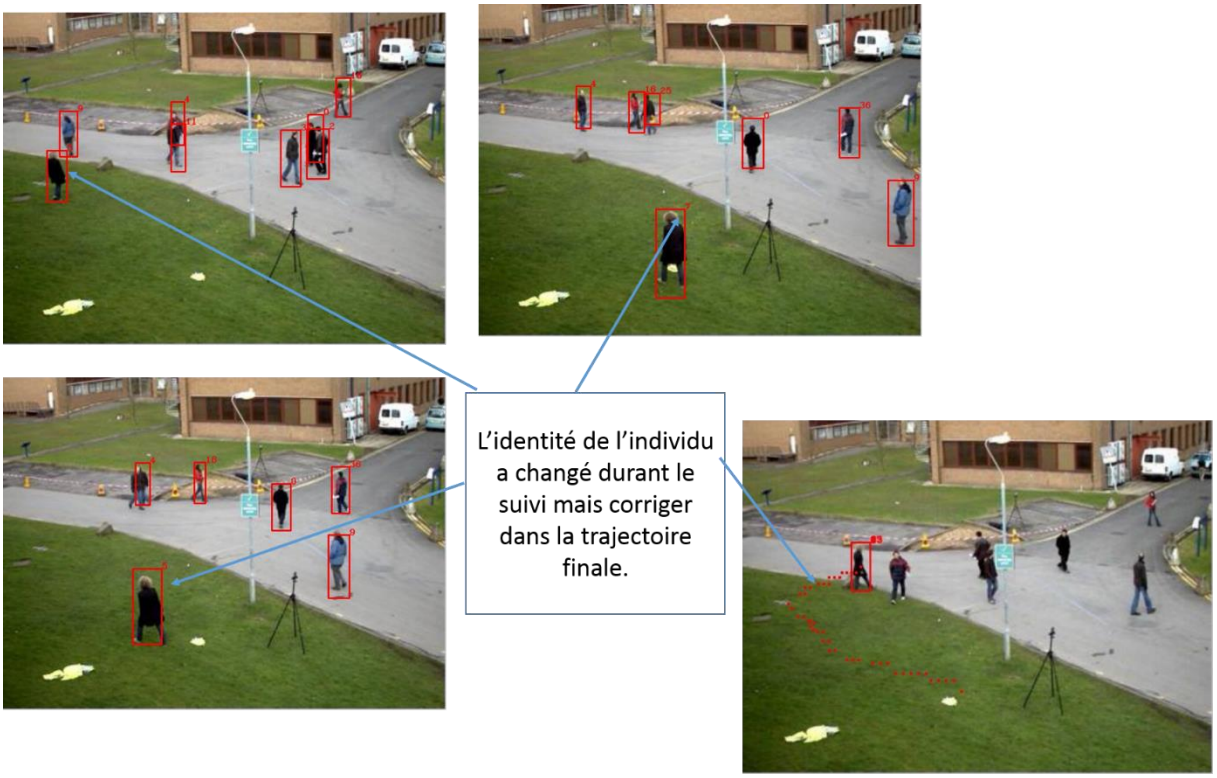


Figure 44. Exemple de résultats de correction de parcours.

Ce résultat nous a permis de valider l'apport de notre méthode de ré-identification dans la récupération des identités des individus perdus dans le suivi par filtre particulaire.

2.5. Conclusion

Dans ce chapitre, nous avons présenté deux des quatre modules principaux qui composent notre système de suivi multi-caméras de personnes. Nous avons présenté les stratégies que nous avons adoptées pour le comptage et le suivi mono-caméra.

Nous avons tout d'abord commencé par le module de comptage qui est basé sur l'exploitation des cartes de profondeur obtenues d'une kinect pour le comptage des individus traversant une ligne d'intérêt. Ce module a pour objectif le comptage des personnes qui rentrent est qui sortent de l'espace surveillé. Nous avons effectué plusieurs séries de tests pour valider les différentes parties composant ce module. Pour cela, nous avons commencé par la validation de la phase de détection de personnes. La stratégie que nous avons adoptée pour cette phase basée sur l'utilisation d'une grille a montrée des résultats très satisfaisants qui dépassent les 95% de bonnes détections. Ensuite, nous sommes passés à la validation de la stratégie de suivi qui est basée sur l'utilisation d'un modèle cinétique avec un MeanShift. Cette stratégie aussi a montrée des résultats très satisfaisants qui dépassent les 90% de bons suivis. La combinaison de ces deux stratégies nous a permis de valider notre algorithme de comptage qui lui aussi a montré des résultats très satisfaisants qui dépassent les 95%. Nous avons validé cette stratégie avec les différents cas de figure qui peuvent survenir dans la réalité par la construction d'une base dans laquelle les différents scénarios sont simulés. Cette stratégie que nous avons proposée a montré une certaine limite pour le cas d'individus avec une grande différence de taille qui marchent trop proches les uns des autres.

Ensuite, nous sommes passés au module de suivi mono-caméra qui est utilisé dans la récupération des tracklets qui représentent les parcours d'individus dans le champ de vision de chacune des caméras. Nous avons proposé une stratégie basée sur l'utilisation d'un filtre particulaire que nous avons validé par plusieurs séries de tests effectués dans la base PETS 2009. Pour cela, nous avons commencé par la réalisation de tests de validation pour notre stratégie de détection de personnes qui est basée sur la détection et la segmentation de mouvement. Les tests que nous avons effectués nous ont permis de valider cette stratégie qui nous a permis de détecter environ 90% des individus présents dans la scène. Les détections obtenues dans cette première phase nous ont permis d'initialiser les trackeurs qui sont basés sur l'utilisation d'un filtre particulaire. Ces trackeurs ont montré une grande capacité à suivre les individus sur des courts laps de temps pour la récupération des segments de trajectoires ; mais ils ont aussi montré une limite pour le suivi sur des longues durées. Cette limite nous a conduit à conception de

la troisième phase de notre système de suivi mono-caméra qui est responsable de la fusion des segments de trajectoires pour l'obtention des tracklets qui représentent les parcours des individus dans le champ de vision de chacune des caméras. Pour ce faire, nous avons proposé une stratégie de ré-identification que nous avons utilisé pour la correction des parcours.

Chapitre 3. Ré-identification de personnes

3.1. Introduction

Durand la dernière décennie, l'étude des comportements humains est l'un des domaines de recherche les plus actifs. Ces études sont effectuées à des fins de statistique ou à des fins de sécurité. De ce fait, la conception d'applications automatiques exploitant la vidéosurveillance pour étudier le comportement humain devient de plus en plus nécessaire. L'un des sujets les plus étudiés dans ce domaine est le suivi des personnes dans un réseau de caméras. Cette problématique peut être considérée comme l'étape de base dans la conception de la plupart des applications. Pour la résoudre, deux approches principales ont été adoptées : les approches centralisées et les distribuées. L'approche centralisée est basée sur l'exploitation de la relation entre différentes caméras pour effectuer un suivi soit par la reconstruction 3D de la scène, soit par la récupération de la carte d'occupation du plan de sol. L'approche distribuée est fondée sur la combinaison d'un module de suivi mono-caméra conçu pour la récupération des segments de trajectoire dans le champ de vision d'une seule caméra et d'une stratégie de ré-identification conçue pour associer les individus suivis à travers les différentes caméras. Pour cela, la conception de stratégies de ré-identification adéquates est une tâche très importante et compliquée en même temps. Cette tâche a été traitée très souvent au cours des dernières années. Les méthodes de ré-identification proposées peuvent être classées en trois catégories principales : les méthodes basées sur les signatures discriminantes, les méthodes basées sur les transformations de caractéristiques et les méthodes basées sur l'apprentissage métrique.

Dans ce chapitre, nous proposons une nouvelle méthode de ré-identification qui est utilisée pour associer les individus entre des caméras qui ne se chevauchent pas ; elle est conçue à des fins de suivi multi-caméras de personnes dans le contexte d'analyse du marketing comportementale. La méthode proposée se décompose en deux phases : la première utilisée pour la sélection des régions saillantes, et la seconde utilisée pour l'association de ces régions et la confirmation de la ré-identification.

Plus précisément, la première phase consiste dans la conception d'un nouveau framework permettant la sélection des zones les plus discriminantes (régions saillantes) pour chaque individu suivi. Tandis que la seconde consiste dans l'utilisation d'une méthode d'appariement adéquate pour associer les régions sélectionnées précédemment afin de confirmer la ré-identification. Pour ce faire, nous avons proposé et implémenté trois méthodes d'appariement, la première basée sur un appariement direct, la seconde sur un appariement par SVM, et la dernière sur un appariement par apprentissage profond.

3.2. Ré-identification de personnes

La ré-identification de personne est considérée comme la principale étape dans la réalisation de différentes applications. Elle peut être utilisée dans la restriction d'accès à certaines zones pour la surveillance comme elle peut aussi être utilisée pour l'analyse de comportements d'individus.

La ré-identification des personnes à partir de flux vidéo (images) est définie comme étant la capacité d'associer un individu détecté dans une image donnée à un autre individu détecté dans une autre image, et ainsi pouvoir affirmer qu'il s'agit du même individu. Ce principe est illustré dans la Figure 45.

Figure 45. Principe de la ré-identification.

En se basant sur la définition donnée précédemment, deux sortes de ré-identification avec différents défis peuvent être différenciées : la ré-identification avec une seule caméra et la ré-identification dans un réseau de caméras.

3.2.1. Ré-identification avec une seule caméra

La ré-identification avec une seule caméra consiste dans le cas où les deux détections de l'individu à ré-identifier sont obtenues de la même caméra. Cette ré-identification intervient plus souvent dans le cadre de la restriction d'accès ou l'étude de comportement dans une seule caméra comme la gestion d'occultations dans le cadre de suivi mono-caméra de personnes.

Pour créer un algorithme de ré-identification avec une seule caméra plusieurs défis sont à relever. Les problèmes principaux qu'il faut résoudre consistent dans :

- La gestion des visions partiales (l'individu à ré-identifier n'est pas totalement visible dans les différentes détections).
- La gestion des différentes poses que l'individu peut prendre.
- L'espacement temporel entre les différentes détections (cet espacement peut engendrer des changements d'éclairage ou de l'individu en question).

3.2.2. Ré-identification dans un réseau de caméras

Dans cette problématique de ré-identification et contrairement à la ré-identification avec une seule caméra, les différentes détections à associer peuvent être obtenues de plusieurs caméras. Cette ré-identification est souvent utilisée dans les applications d'analyse de comportement, car qu'un réseau de caméras permet d'avoir une vue plus globale sur le comportement de l'individu en question.

Cette problématique est plus compliquée comparativement à la précédente. Pour créer un algorithme de ré-identification dans un réseau de caméras, il faut résoudre, en plus de ceux rencontrés pour le cas d'une seule caméra, d'autres problèmes reliés à la multi-caméra comme :

- L'angle de vue de chacune des caméras.
- La large différence d'éclairage qui peut exister entre les différentes caméras.
- Les paramètres des caméras.

3.3. État de l'art

Au cours de la dernière décennie, de nombreux travaux ont été réalisés et plusieurs méthodes ont été proposées pour résoudre le problème de l'association des individus entre les caméras. Les travaux proposés sont généralement basés sur l'utilisation de l'une de ces trois approches : les méthodes basées sur les signatures discriminantes, les méthodes basées sur les transformations de vecteurs de caractéristiques et les méthodes basées sur l'apprentissage métrique.

3.3.1. Méthodes basées sur la signature discriminante

La première catégorie de méthodes, de loin la plus utilisée, est celle basée sur l'approche des signatures discriminantes. Ces méthodes se focalisent sur la recherche de la signature la plus discriminante pouvant être utilisée pour différencier les différents individus. Son principe de base est illustré dans la *Figure 46*.

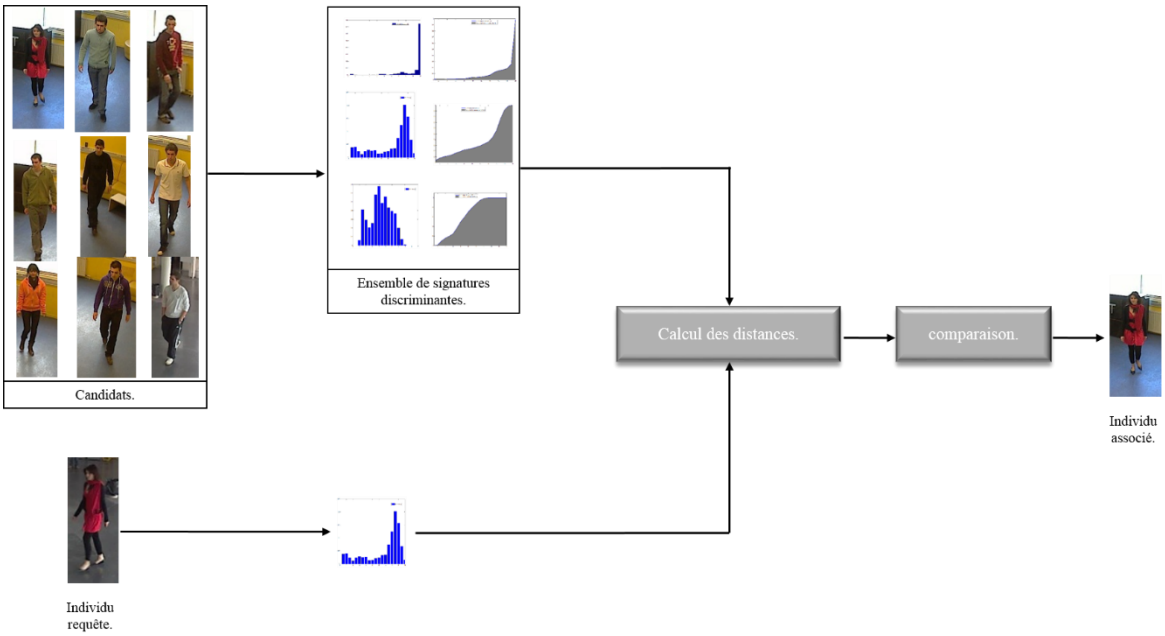


Figure 46. Principe de ré-identification avec signature discriminante.

Parmi les méthodes utilisant l'approche des signatures discriminantes nous pouvons citer des méthodes basées sur l'apprentissage d'un classifieur utilisé pour récupérer l'identité d'individus, telles que celles proposées dans [102] et [103], et des méthodes basées sur la recherche directe des signatures les plus discriminantes soit en segmentant les individus en différentes parties du corps, telles que celles proposées dans [104] et [105], en cumulant plusieurs caractéristiques locales dans la même signature comme dans [106], ou en utilisant la saillance comme dans [107] et [108]. Ces différentes approches sont illustrées dans les figures (*Figure 47*, *Figure 48* et *Figure 49*).

Bak et al [102] ont proposé un algorithme basé sur l'utilisation d'*Adaboost* avec des caractéristiques similaires à celles du *Haar* pour ré-identifier les individus. Dans cette méthode, un ensemble de détections pour chaque individu est accumulé sur une période de temps. Cet ensemble de détections est utilisé pour extraire des caractéristiques de type *Haar* qui seront ensuite utilisées avec l'*Adaboost* pour extraire une signature visuelle de l'individu à ré-identifier. Cette signature discriminante est utilisée pour ré-identifier les individus suivis dans un système de suivi multi-caméras. Dans le même contexte, Zhao et al [103] ont proposé une méthode basée sur l'apprentissage de filtres de niveau intermédiaire qui sont utilisés pour identifier les modèles visuels. L'idée principale de cette méthode est la sélection de certains patches locaux qui sont discriminants et qui sont utilisés pour apprendre des filtres. L'apprentissage de ces filtres est basé sur une

stratégie d'apprentissage *cross-view* pour obtenir des filtres invariants et discriminants. Les filtres résultants sont utilisés avec un *RankSVM* pour ré-identifier les individus.

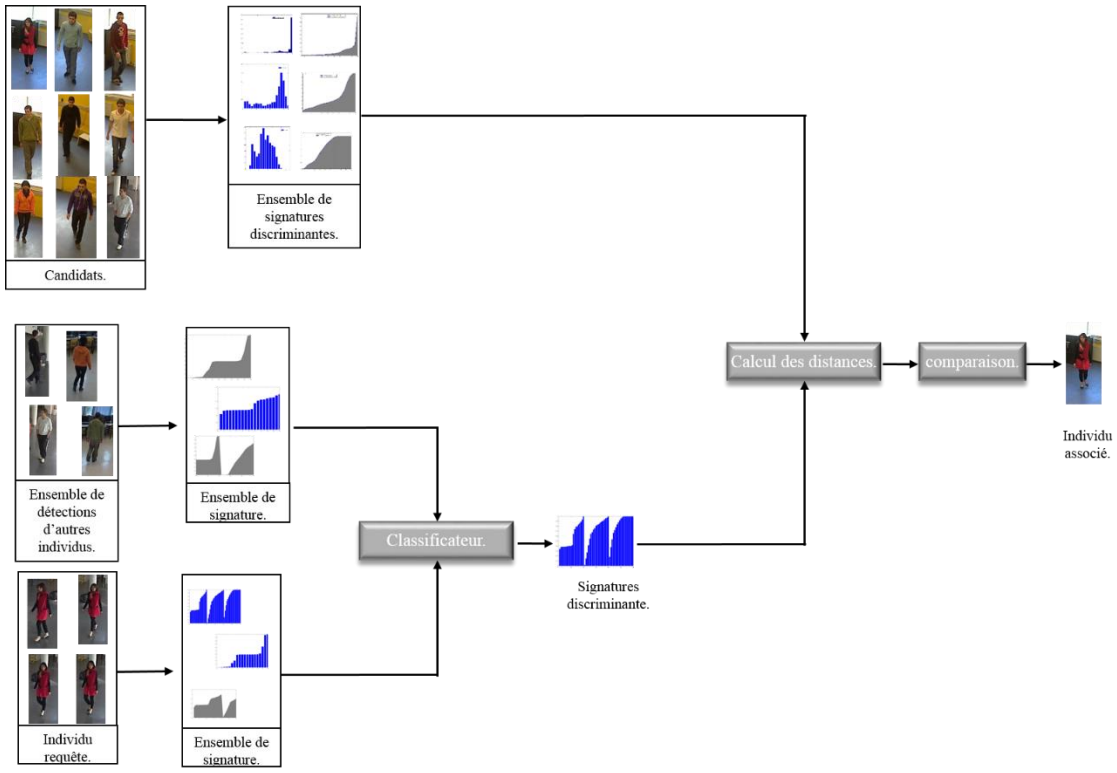


Figure 47. Signature discriminante par apprentissage d'un classificateur.

Pour extraire la signature la plus discriminante, Martinel et al [105] ont proposé une méthode basée sur trois étapes: la segmentation des individus aux différentes parties du corps, l'extraction des vecteurs de caractéristiques locales pour chaque partie (couleur *HSV*, pyramide *HOG* et *SIFT*), l'utilisation d'une mesure de distance exploitant la combinaison pondérée de la distance entre le *PHOG* et les histogrammes de couleur pour ré-identifier les individus. De la même manière, Aziz et al [104] ont proposé une méthode de ré-identification dans laquelle les apparences des individus détectés sont classées en apparence de front et apparence de dos. Ensuite les silhouettes de ces détections sont segmentées en trois parties (tête, torse et jambes). Pour chacune des parties segmentées une signature discriminante basée sur *SIFT*, *SURF* et *spin image* est extraite. Ces signatures discriminantes sont utilisées pour associer les individus détectés dans un réseau avec des caméras qui ne se chevauchent pas.

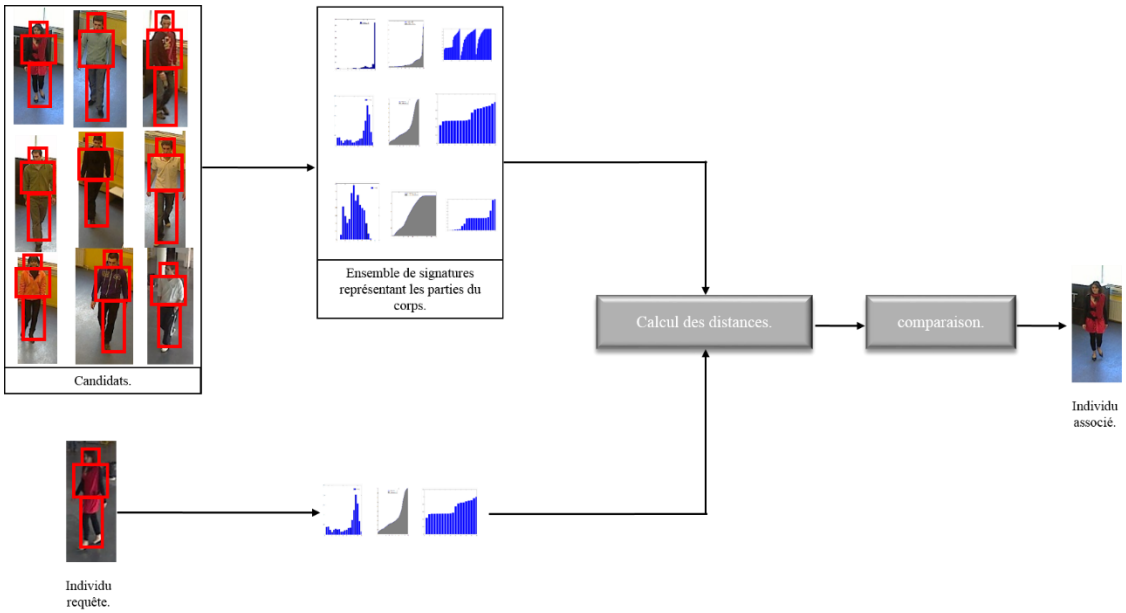


Figure 48. Signature discriminante par segmentation.

Dans le même contexte d'extraction de la signature la plus discriminante, [107] et [108] ont exploité la saillance de l'image. Martinel et al [107] ont ainsi proposé une stratégie basée sur l'identification des régions saillantes pouvant être utilisées pour la ré-identification des individus. Ils ont utilisé une technique de noyaux sur graphes pour calculer la saillance de chaque pixel en fonction de ses alentours. Cette saillance est utilisée pour pondérer les vecteurs de caractéristiques dans la ré-identification. Dans le même contexte, Zhao et al [108] ont proposé une méthode de ré-identification basée sur les correspondances des saillances. La première étape de cette approche est l'estimation de la distribution de la saillance sur l'ensemble du corps de l'individu. Ensuite, le problème de ré-identification de la personne est formulé comme un appariement de saillance où l'appariement entre individus est fait en fonction à de la distribution spatiale de la saillance et la similarité visuelle entre les taches qui ont la même saillance.

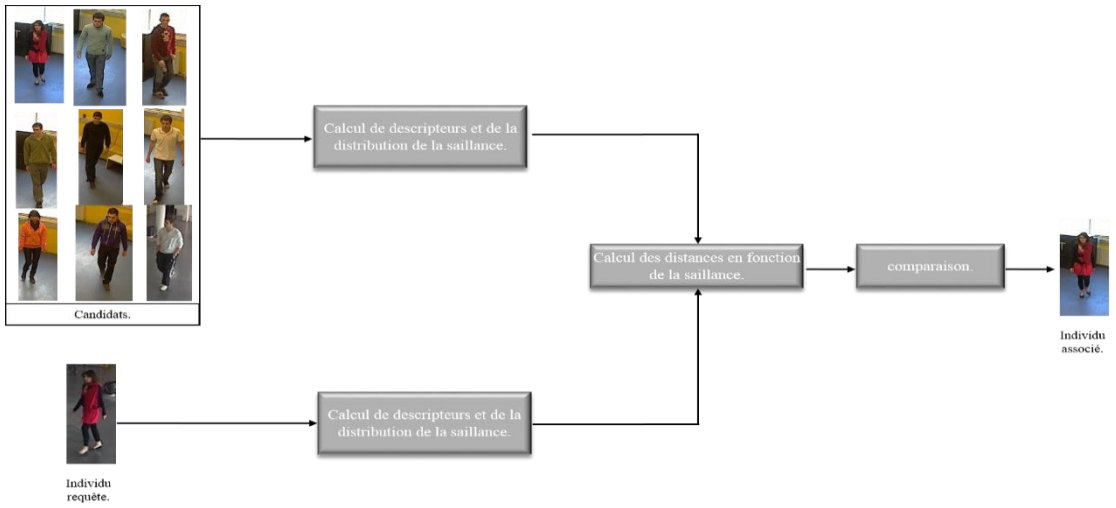


Figure 49. signature discriminante par saillance.

3.3.2. Méthodes basées sur la transformé des vecteurs caractéristiques

La deuxième catégorie de méthodes de ré-identification est celle basée sur les transformations des vecteurs de caractéristiques. Ces méthodes de ré-identification se focalisent sur la recherche de la transformée qui relie les vecteurs de caractéristiques dans la première caméra à leurs vecteurs de caractéristiques correspondants dans la deuxième caméra. Ce principe de base est illustré dans la *Figure 50*.

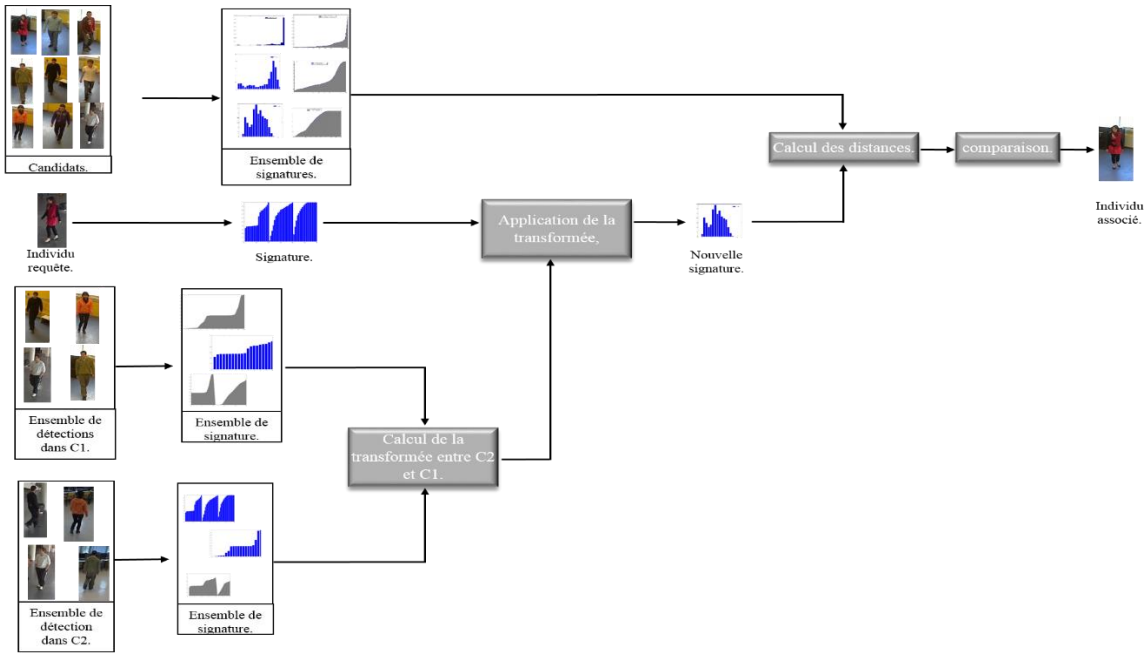


Figure 50. Principe de la ré-identification avec les transformées des vecteurs de caractéristiques.

Une variété de travaux a été proposée dans cette catégorie en se focalisant sur deux axes principaux. Premièrement les méthodes qui essayent directement de trouver la transformée entre les vecteurs de caractéristiques dans les deux caméras, pour l'utiliser dans la transformation de l'un des vecteurs de caractéristiques avant sa comparaison avec son homologue dans la deuxième caméra [109], [110] et [111]. Deuxièmement les méthodes qui essayent d'apprendre un classificateur qui peut différencier les paires d'images représentant le même individu et celles représentant des individus différents, ceci afin de l'utiliser dans la comparaison des nouvelles paires de détections [112], [113].

Pour pouvoir calculer la probabilité que deux observations correspondent au même individu, Javed et al [111] ont proposé une méthode basée sur l'apprentissage de la relation spatio-temporel ainsi que la relation d'apparence entre deux caméras. La relation d'apparence consiste dans la fonction de transfert de luminosité (*brightness transfer function*) entre les deux caméras, tandis que la relation spatio-temporel consiste dans les entrées, les sorties, les directions et le temps moyen dont on a besoin pour atteindre le champ de vision de l'une des deux caméras en quittant le champ de vision de l'autre caméra. De la même manière, Datta et al [109] ont proposé une méthode pour la récupération de la fonction de transfert de luminosité qui associe les descripteurs de couleur et de texture de haute dimension afin de l'utiliser dans la ré-identification.

Pour associer les détections d'individus dans deux caméras qui ne se chevauchent pas, Avraham et al [113] ont proposé un algorithme basé sur l'utilisation d'un classifieur entrainé pour distinguer les paires positives et les paires négatives d'images en utilisant un histogramme de couleur. Dans cette méthode, ils assument que les paires positives sont les paires d'images qui représentent le même individu, tandis que les paires négatives sont celles qui représentent deux individus différents. Ils ont utilisé un SVM entrainé sur un ensemble de paires positives et négatives pour pouvoir, ensuite, classer les nouvelles paires détectées en tant que positive ou négative, ceci afin de décider si les deux détections représentent le même individu et ainsi confirmer la ré-identification. Martinel et al [112] ont utilisé la même idée de base mais en entrainant plusieurs experts de ré-identification avec une représentation robuste des caractéristiques obtenues des paires d'images qui représentent soit le même individu soit deux individus différents. Ensuite, ces experts sont utilisés pour évaluer les représentations obtenues par de nouvelles paires d'images afin de confirmer la ré-identification en fonction des réponses de l'ensemble des experts.

3.3.3. Méthodes basées sur l'apprentissage de métriques

La dernière catégorie de méthodes de ré-identification est celle basée sur l'apprentissage métrique. Cette approche se focalise sur l'apprentissage d'une métrique optimale qui peut être utilisée pour comparer les signatures (vecteurs de caractéristiques) qui sont extraites des différentes détections. Ces méthodes améliorent la ré-identification en utilisant une métrique adaptée pour le descripteur utilisé [114], [68], [115] et [116]. Ce principe de base est illustré dans la *Figure 51*.

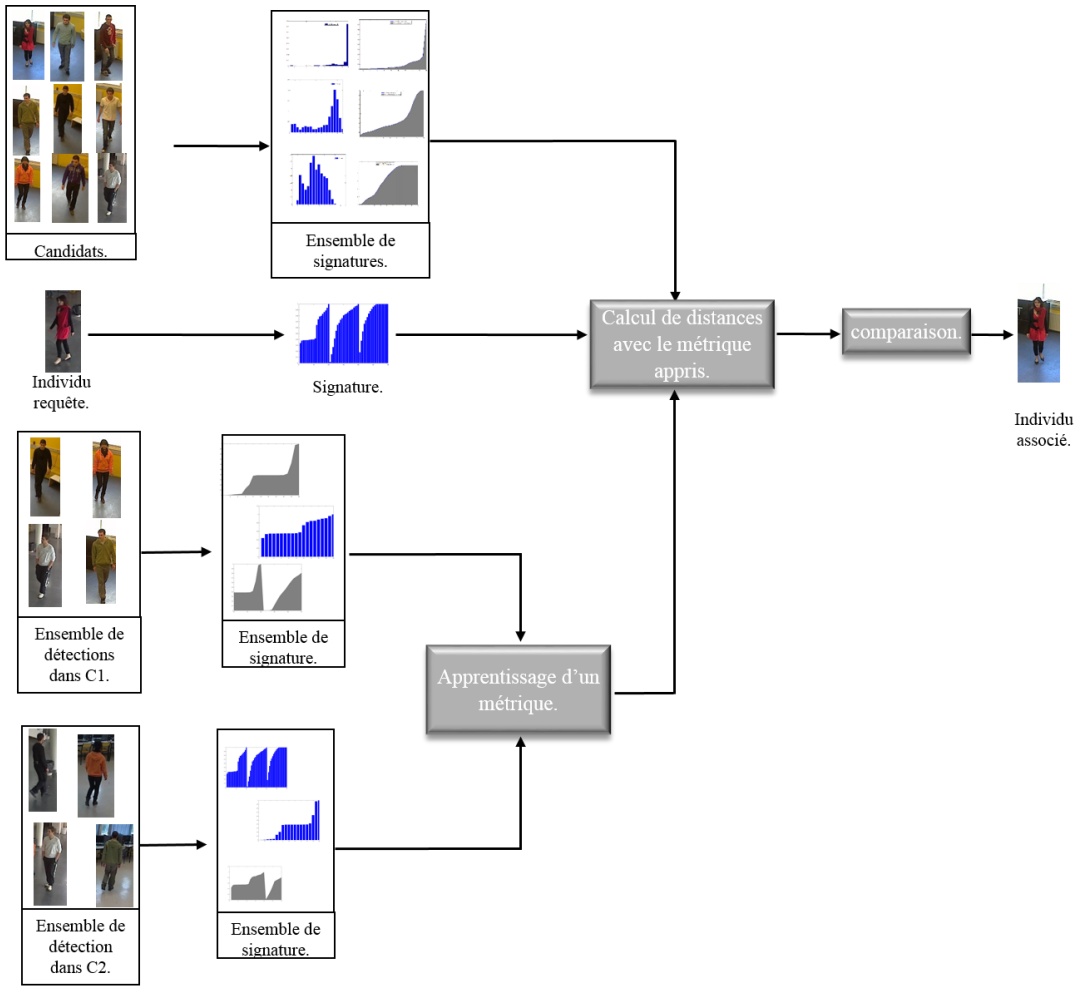


Figure 51. Principe de la ré-identification avec l'apprentissage métrique.

Pour cette catégorie d'algorithmes la métrique est utilisée pour la sélection de la signature la plus discriminante afin d'améliorer les résultats de la ré-identification. Pour se faire, Pedagadi et al [116] ont proposé une méthode basée sur l'utilisation d'une analyse locale discriminante de fisher (Local Fisher Discriminant Analysis) afin d'apprendre une métrique entre un ensemble de paires de descripteurs qui représentent des paires

d'observations d'un ensemble d'individus. Dans le même contexte, Liao et al [115] ont créé une méthode qui combine la sélection de la signature la plus discriminante et l'apprentissage de la métrique la plus adaptée pour la comparaison des différentes signatures. Ils ont ainsi proposé un descripteur nommé LOMO (occurrence maximale locale) basé sur l'analyse horizontale de l'occurrence des caractéristiques locales. Ce descripteur est utilisé avec une métrique entrainée par l'analyse quadratique discriminante.

3.4. Description de la méthodologie proposée

L'étude de l'état de l'art nous a permis d'analyser les différentes approches utilisées dans le cadre de la ré-identification des personnes dans un réseau de caméra. Cette étude nous a permis aussi de voir les différents défis rencontrés dans ce domaine.

Indépendamment de l'approche utilisée, les méthodes proposées essayent de résoudre les mêmes problèmes de base. Ces défis standards consistent dans les variations de l'éclairage, le changement de perspective et les apparences similaires. Ces problèmes sont dus à l'environnement, les positions des différentes caméras ainsi qu'à la posture de la personne à ré-identifier. Ces défis ont été étudiés dans une variété de travaux, mais sont toujours d'actualité.

Dans ce chapitre, nous proposons une nouvelle méthode de ré-identification dans le cadre de l'association d'individus suivis dans un réseau de caméras avec ou sans zones de recouvrement. Cette méthode est basée sur la conception d'un nouveau framework utilisé pour sélectionner les patches les plus discriminants pour chaque individu suivi. Ensuite, ces patches sont utilisés pour associer les individus détectés dans l'ensemble des caméras du réseau. Les étapes principales de ce nouvel algorithme sont :

- La construction d'un nouveau framework pour la sélection des régions saillantes pour chaque individu en se basant sur l'utilisation d'informations spatiales et descripteurs d'apparence obtenus des individus détectés.
- L'utilisation des patches sélectionnés afin d'associer les individus dans l'étape d'appariement.
- L'utilisation d'une méthode directe et d'une méthode basée sur l'apprentissage profond, pour l'association des patches.

Donc, nous proposons une méthode de ré-identification pour associer les individus suivis dans un réseau de caméras afin de pouvoir reconstruire leurs trajectoires dans toute

la scène surveillée. Cette méthode est utilisée pour trouver les correspondances entre les individus suivis dans les différentes caméras ce qui permet de combiner les segments de trajectoires obtenues dans chacune des caméras et ainsi avoir la trajectoire globale de l'individu suivi. Pour la conception d'une méthode efficace nous devons faire avec plusieurs défis causés par la configuration du réseau ainsi que les poses des individus en question. Ces défis sont illustrés dans la *Figure 52*.

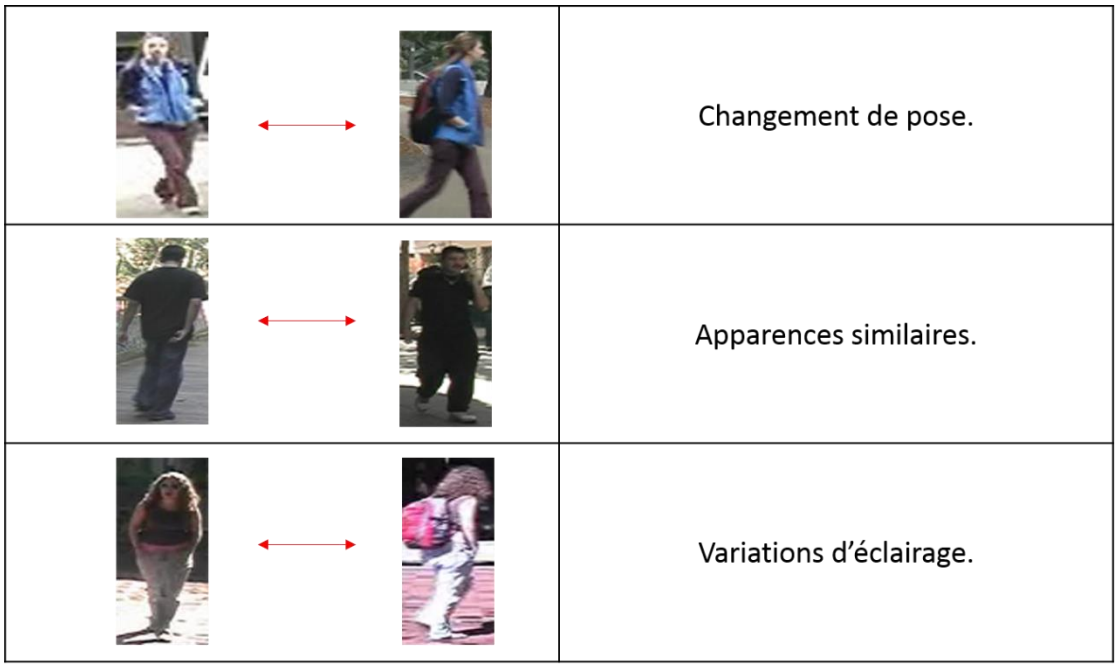


Figure 52. Illustration des principaux défis rencontre dans la re-identification.

La méthode que nous proposons exploite des différences locales entre les individus suivis. Cette exploitation est faite par la sélection des régions les plus discriminantes pour chaque individu suivi. Notre méthode est réalisée en deux phases : la phase d'apprentissage et la phase d'appariement. La première est utilisée pour la sélection des patches les plus discriminants en construisant un framework qui regroupe des descripteurs d'apparence ainsi que des informations spatiales pour un ensemble de détections de chaque individu. Une fois les patches sélectionnés, la deuxième phase consiste dans l'appariement de ces patches afin d'associer les individus en se basant en premier lieu sur une approche directe puis sur une approche par apprentissage. Le diagramme général de notre méthode est illustré dans la *Figure 53*.

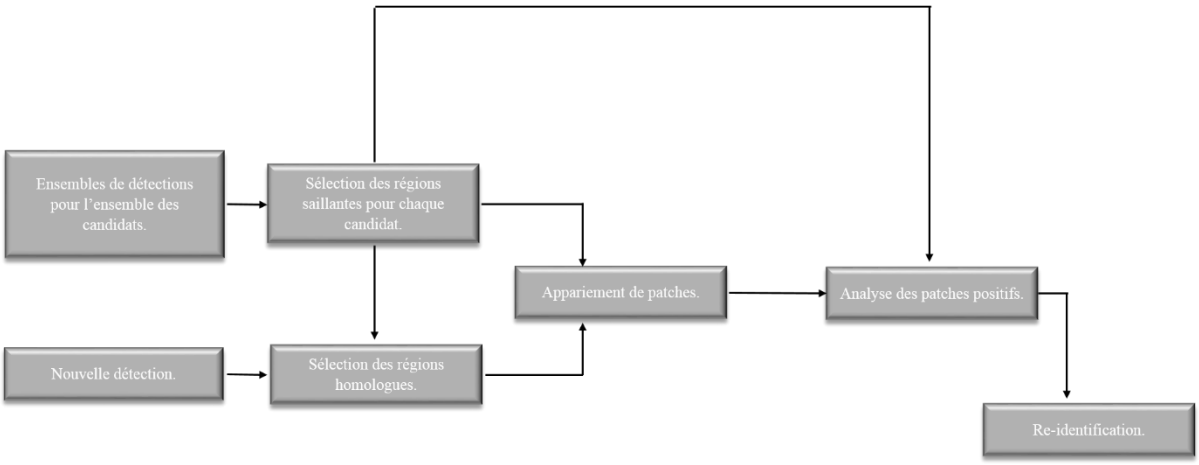


Figure 53. Diagramme général de notre algorithme de ré-identification.

3.4.1. Sélection de patches discriminants

La première phase de notre système de ré-identification consiste dans l'étape d'apprentissage qui a pour objectif de trouver les patches les plus discriminants pouvant être utilisés dans la représentation de chacun des individus dans le framework.

Cette partie est basée sur l'utilisation d'une stratégie d'apprentissage pour la sélection des régions saillantes. Pour ce faire, deux étapes sont nécessaires : la première est la construction d'un framework dans lequel chaque individu est représenté par un ensemble de graphes représentant l'apparence et le positionnement d'un ensemble de patches obtenus d'un ensemble de détections de l'individu en question, et la deuxième consiste dans l'utilisation de ce framework (représentation des différents individus) pour sélectionner les patches les plus discriminants. Cette procédure est illustrée dans la *Figure 54*.

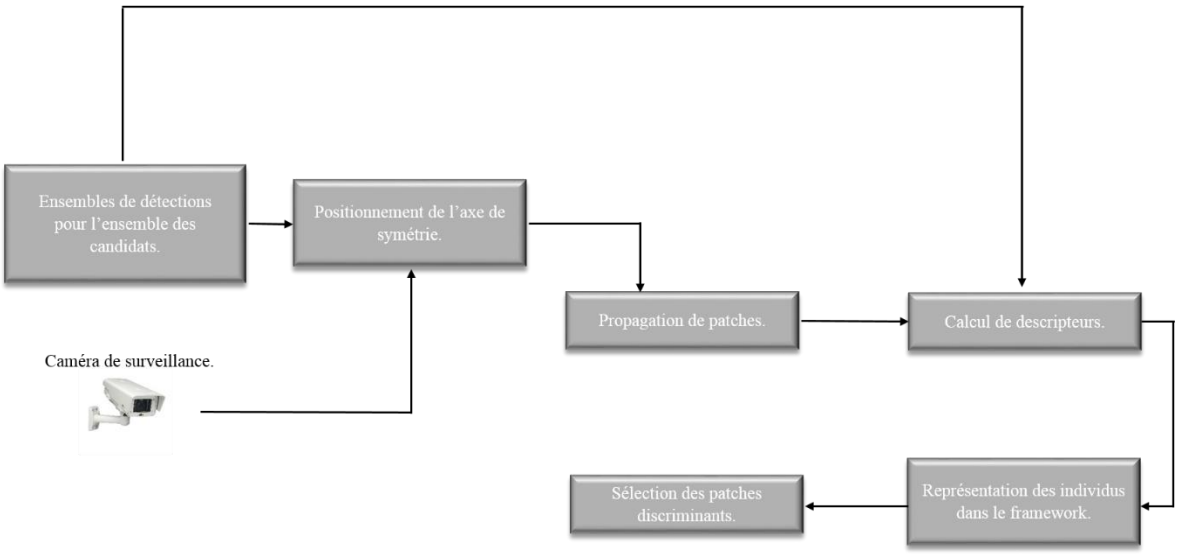


Figure 54. Sélection des patches discriminants.

3.4.1.1. Introduction du framework utilisé

Nous avons proposé cette méthode de ré-identification dans le cadre de suivi multi-caméra de personnes pour associer les individus suivis dans l'ensemble des caméras. De ce fait, nous avons proposé une stratégie utilisant plusieurs détections pour chaque individu (*multi-shot*) afin de pouvoir exploiter les résultats obtenus dans le suivi mono-caméra qui donne un ensemble de détections (tracklets) pour chaque individu dans le champ de vision de chacune des caméras. Le suivi mono-caméra résulte un ensemble de détections successives pour les différents frames de la vidéo. Ces détections représentent l'individu suivi à différents instants. Donc, pour construire notre framework, nous utilisons quelque unes de ces détections pour créer un ensemble de graphes qui représentent la description spatiale de l'apparence de l'individu suivi dans le champ de vision de l'une des caméras.

La construction du framework est faite en deux phases. La première phase consiste dans l'estimation de l'axe central vertical pour chacune des détections utilisées, tandis que la deuxième phase consiste dans le calcul des descripteurs et dans la représentation des individus dans le framework.

3.4.1.1.1. Estimation de l'axe central vertical

La première phase dans la construction du framework consiste dans l'estimation de l'axe central vertical de chacune des détections qui seront utilisées pour la construction

du framework. L'estimation de cet axe permettra la bonne propagation des patches dans les étapes qui suivront. Cette phase est importante par le fait qu'elle réduit les interférences avec l'arrière-plan pour les patches qui seront utilisés dans la ré-identification. Elle est réalisée en deux étapes : la première consiste dans l'estimation de l'arrière-plan de la scène pour une caméra donnée et la détection des individus présents dans la scène en se basant sur un algorithme de soustraction de fond et la deuxième consiste dans la localisation de l'axe.

Pour l'estimation de l'arrière-plan, nous avons commencé par l'utilisation de la moyenne glissante. Pour ce faire, nous avons utilisé l'équation (21) avec des patches de 100 frames de la vidéo afin d'obtenir une estimation correcte. Cette estimation est régulièrement mise à jour pour intégrer les possibles nouveaux objets statiques dans la scène.

$$\text{arrière_plan} = \frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}} \text{frame}_{\mathrm{i}} \qquad (21)$$

Une fois l'arrière-plan estimé, l'étape suivante consiste dans la délimitation des blobs qui représentent les individus en se basant sur les localisations obtenues dans le suivi mono-caméra et sur un algorithme de soustraction de fond. Pour cela, nous utilisons un algorithme simple basé sur la différence entre la frame courante et l'arrière-plan. La différence entre la frame courante et l'arrière-plan $diff_t$ est donnée par l'équation (22). Ensuite, cette image $diff_t$ est utilisée pour obtenir le masque de premier-plan B_t en utilisant un seuil th_{min} avec l'équation (23).

$$\text{diff}_{\mathrm{t}} = |\text{frame}_{\mathrm{t}} - \text{arrière_plan}| \qquad (22)$$

$$\mathrm{B_t(x,y)} = \begin{cases} 1 & \text{si: } \mathrm{th_{min}} < \mathrm{diff_t(x,y)} \\ 0 & \text{autrement} \end{cases} \qquad (23)$$

La combinaison de ce masque avec une analyse en composantes connexes permet de délimiter les blobs qui représentent l'individu suivi dans chacune des détections.

La deuxième étape est l'estimation de la position de l'axe de symétrie à partir des blobs obtenus dans l'étape précédente. On suppose que cet axe de symétrie vertical représente l'axe qui sépare les blobs détectés sur l'axe horizontal (axe des x) en deux parties qui ont le même nombre de pixels de premier-plan. De ce fait, la détection de cet axe peut se faire par la minimisation de la distance entre la somme des pixels situés à gauche et à droite de cet axe dans l'image binaire des blobs représentants l'individu suivi. Donc, pour détecter l'axe de symétrie vertical, nous avons minimisé l'équation (24) qui représente la distance entre le nombre de pixels de premier-plan à gauche et à droite de l'axe situé à la position x_0. La minimisation cette fonction avec une procédure itérative nous permet de localiser l'axe de symétrie vertical comme illustré dans la *Figure 55*.

$$\mathrm{Dist}_{x_0} = \sum_{x=0}^{x_0} \sum_{y=0}^{y_{max}} p(x,y) \; - \; \sum_{x=x_0}^{x_{max}} \sum_{y=0}^{y_{max}} p(x,y) \qquad (24)$$

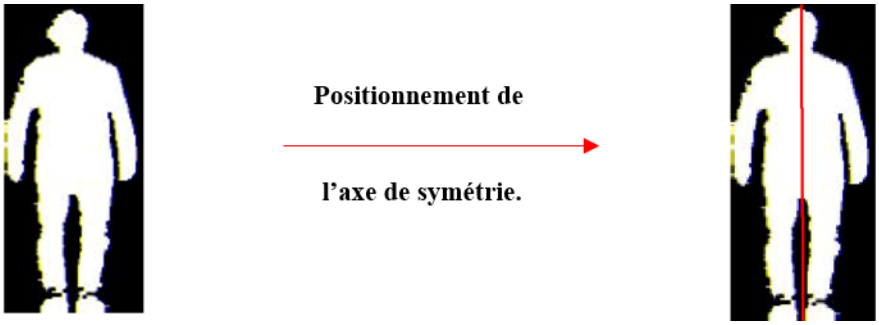


Figure 55. Positionnement de l'axe de symétrie vertical.

3.4.1.1.2. Représentation des individus dans le framework

La deuxième phase dans la construction du framework consiste dans le calcul des descripteurs locaux qui seront utilisés dans la construction des graphes qui représentent les différents individus. Pour ce faire, nous commençons par la propagation de patches autour de l'axe de symétrie vertical localisé précédemment, ensuite le calcul de descripteurs d'apparence qui représentent chacun des patches, et finalement l'utilisation

des descripteurs calculés pour représenter les différentes détections de l'individu dans le framework.

La première étape dans la représentation des individus dans le framework consiste dans l'utilisation d'un ensemble de détections pour chaque individu. Ces détections sont obtenues à partir des segments de trajectoire obtenus par le suivi mono-caméra dans le champ de vision d'une caméra donnée. Une fois que les axes centraux verticaux pour l'ensemble des détections sont positionnés comme spécifié précédemment, un ensemble de patches est propagé autour de cet axe pour chacune des détections. Pour cela, les individus détectés sont représentés par des rectangles d'une largeur w et d'une hauteur h et un ensemble de patches rectangulaires locaux $p(x_0, y,\ wp,\ hp)$ sont propagés dans l'image représentant l'individu (illustré dans la *Figure 56*). Ces patches sont centrés autour du point (x_0, y) qui appartient à l'axe de symétrie vertical et ont des largeurs wp et des hauteurs hp. Les largeurs et les hauteurs ainsi que la position verticale (y) des patches sont variées afin d'avoir des patches de différentes tailles ce qui permet par la suite l'obtention des régions saillantes les plus discriminantes.

Figure 56. Propagation des patches.

A la fin de la première étape, nous obtenons un ensemble de patches pour chaque détection de l'individu suivi. Ces patches seront utilisés dans cette deuxième étape pour calculer un ensemble de descripteurs d'apparence qui sera utilisé dans la représentation de l'individu suivi dans le framework. Donc, dans cette partie nous calculons une description globale pour les histogrammes de couleur pour chacun des patches localisés précédemment. Nous utilisons une description globale pour optimiser les calculs lors des traitements futurs. Nous avons choisi d'utiliser la surface sous l'histogramme de couleur cumulatif dans un espace de couleur HSV, où chaque canal est utilisé indépendamment des autres. Pour ce faire, nous commençons par la conversion de l'image originale vers l'espace de couleur HSV, et nous calculons l'histogramme de couleur pour chaque canal $(h_h,\ h_s,\ h_v)$. Cet histogramme est ensuite utilisé dans le calcul des histogrammes cumulatifs $(H_h,\ H_s,\ H_v)$ que nous calculons à l'aide de l'équation (25).

$$H_{i,j}^{n} = \begin{bmatrix} H_h \\ H_s \\ H_v \end{bmatrix} = \begin{bmatrix} \sum_{l=1}^{j} h_h(l) \\ \sum_{l=1}^{j} h_s(l) \\ \sum_{l=1}^{j} h_v(l) \end{bmatrix} \qquad (25)$$

Où : $H_{i,j}^{n}$ le bin j de l'histogramme cumulatif dans l'espace de couleur HSV pour le patch i du candidat n et $h_{h,s,v}(l)$ le bin l de l'histogramme de couleur dans l'espace de couleur HSV pour le même patch et le même candidat.

Une fois les histogrammes cumulatifs $(H_h,\ H_s,\ H_v)$ obtenus, nous les utilisons pour le calcul d'une description globale de chaque patch. Cette description consiste dans la surface sous les histogrammes cumulatifs $(s_h,\ s_s,\ s_v)$. Nous calculons ces surfaces en utilisant l'équation (26).

$$s_i^n = \begin{bmatrix} s_{i,h}^n \\ s_{i,s}^n \\ s_{i,v}^n \end{bmatrix} = \begin{bmatrix} \int H_{i,h}^n \\ \int H_{i,s}^n \\ \int H_{i,v}^n \end{bmatrix} \quad (26)$$

Où : H_i^n est l'histogramme cumulatif dans l'espace de couleur HSV pour le patch i du candidat n et s_i^nest la surface sous l'histogramme pour le patch i du candidat n.

Un exemple pour ce calcul de descripteur est illustré dans la *Figure 57*.

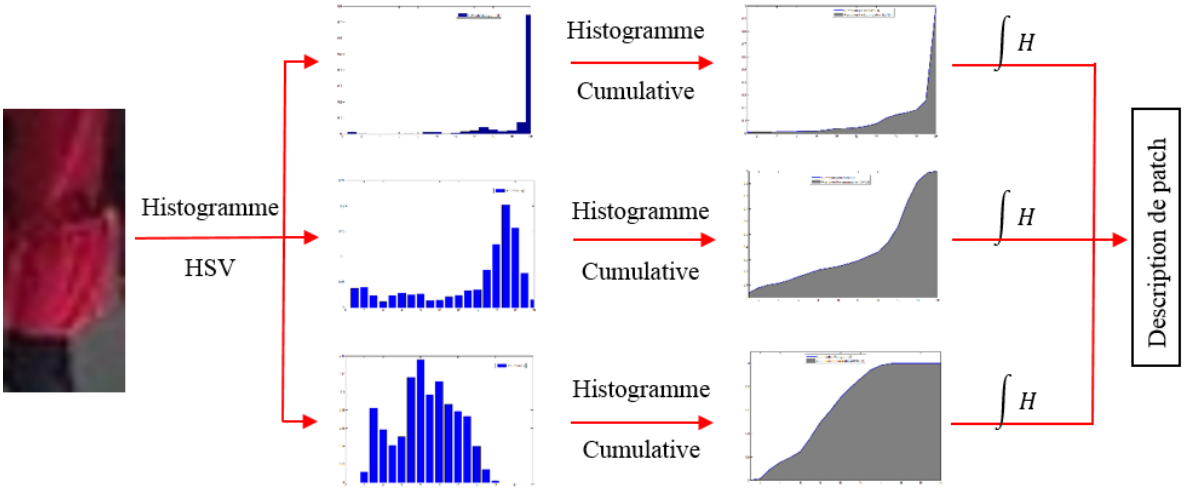


Figure 57. Calcul de descripteurs.

Une fois les descripteurs de l'ensemble des patches sont calculés, nous passons à la représentation des individus dans le framework. Dans ce framework, chaque individu est représenté par plusieurs détections et chaque détection est représentée par un ensemble de patches. La représentation consiste dans les descriptions calculées précédemment en fonction des positions des patches. Un exemple pour cette représentation est donné dans

la *Figure 58*. Dans cet exemple et pour des raisons didactiques, nous avons représenté deux individus en utilisant une seule détection pour chaque individu. Pour chaque détection, nous avons sélectionné cinq patches et nous avons utilisé uniquement le canal H de notre descripteur de couleur. Cet exemple illustre les différentes étapes, en commençant par la propagation des patches, le calcul des descripteurs et la représentation dans le framework.

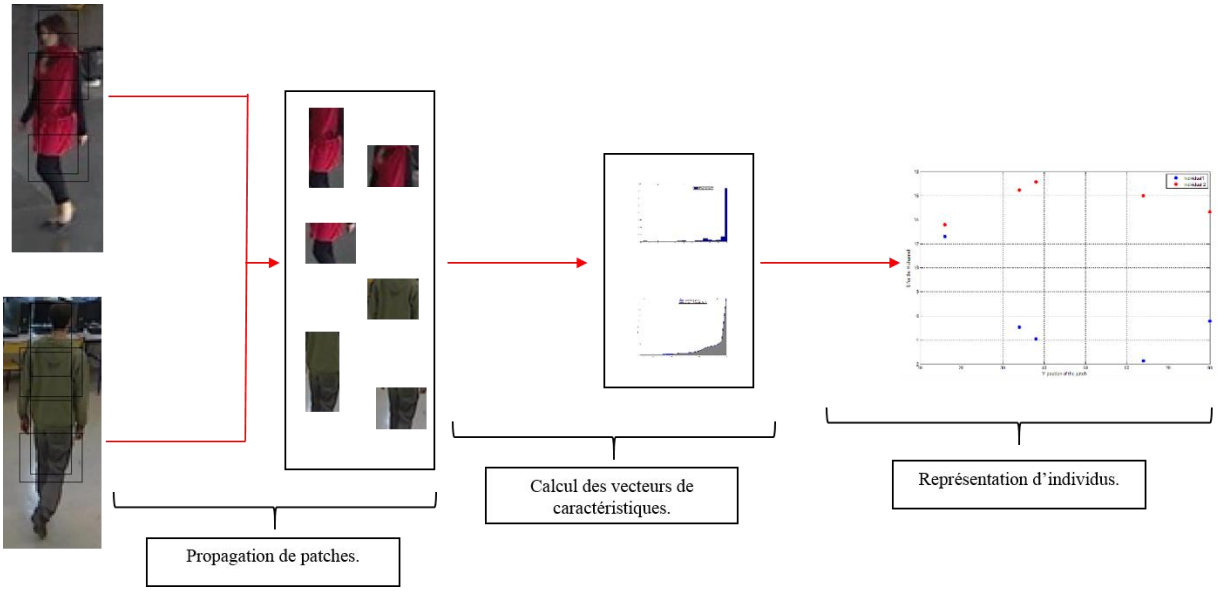


Figure 58. Représentation des individus dans le framework (Exemple de deux individus avec 5 patches dans le canal H de l'espace de couleur HSV).

3.4.1.2. Sélection de patches

Une fois tous les individus représentés dans le framework, la prochaine étape consiste dans la sélection des patches les plus discriminants pour chaque individu par rapport aux autres individus présents dans la base de données. Pour cela, nous utilisons un algorithme qui permet la sélection des patches en minimisant les distances entre les détections appartenant au même individu et en maximisant les distances entre les détections appartenant à des individus différents. Pour chaque individu nous commençons par la sélection des patches les plus stables. Pour ce faire, en se basant sur l'équation (27), nous choisissons les patches minimisant l'écart type entre les détections représentant le même individu.

$$\sigma_s = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(s_{i,m}^{n} - \frac{1}{N}\sum_{j=1}^{N} s_{j,m}^{n}\right)^2} \qquad (27)$$

Où : $s_{i,m}^{n}$ est la surface sous l'histogramme cumulatif dans l'espace de couleur HSV pour le patch m de la détection i du candidat n et N est le nombre de détections utilisées dans l'optimisation.

Cette première étape de sélection nous permet d'avoir l'ensemble de patches les plus stables pour chacun des individus. Une fois ces ensembles de patches obtenus, la prochaine phase consiste dans le choix parmi ces patches de ceux qui maximisent la distance entre les différents individus. Nous utilisons alors l'équation (28) qui permet le calcul de la distance relative entre les distances entre les détections appartenant à un individu donné et celles appartenant aux autres individus et la distance entre les détections représentant cet individu donné.

$$dist_n = \sum_{i=1}^{n} \frac{min\left\{\begin{matrix}\min\left\{\begin{matrix}|s_{i,m}^{n} - s_{i,m}^{l,l\neq n}| \\ \vdots \\ |s_{i,m}^{n} - s_{N,m}^{l,l\neq n}|\end{matrix}\right\} \\ \vdots \\ \min\left\{\begin{matrix}|s_{N,m}^{n} - s_{i,m}^{l,l\neq n}| \\ \vdots \\ |s_{N,m}^{n} - s_{N,m}^{l,l\neq n}|\end{matrix}\right\}\end{matrix}\right\} - \max\left\{\begin{matrix}\max\left\{\begin{matrix}|s_{i,m}^{n} - s_{i,m}^{n}| \\ \vdots \\ |s_{i,m}^{n} - s_{N,m}^{n}|\end{matrix}\right\} \\ \vdots \\ \max\left\{\begin{matrix}|s_{N,m}^{n} - s_{i,m}^{n}| \\ \vdots \\ |s_{N,m}^{n} - s_{N,m}^{n}|\end{matrix}\right\}\end{matrix}\right\}}{\max\left\{\begin{matrix}\max\left\{\begin{matrix}|s_{i,m}^{n} - s_{i,m}^{n}| \\ \vdots \\ |s_{i,m}^{n} - s_{N,m}^{n}|\end{matrix}\right\} \\ \vdots \\ \max\left\{\begin{matrix}|s_{N,m}^{n} - s_{i,m}^{n}| \\ \vdots \\ |s_{N,m}^{n} - s_{N,m}^{n}|\end{matrix}\right\}\end{matrix}\right\}} \qquad (28)$$

Où : $s_{i,m}^{n}$ est la surface sous l'histogramme cumulatif dans l'espace de couleur HSV pour le patch m de la détection i du candidat n et N est le nombre de détections utilisées dans l'optimisation.

A la fin de cette phase, nous obtenons une nouvelle représentation des individus dans le framework. Dans cette nouvelle représentation seuls les patches qui seront utilisés dans l'appariement sont présents. Cette procédure est détaillée dans la *Figure 59*.

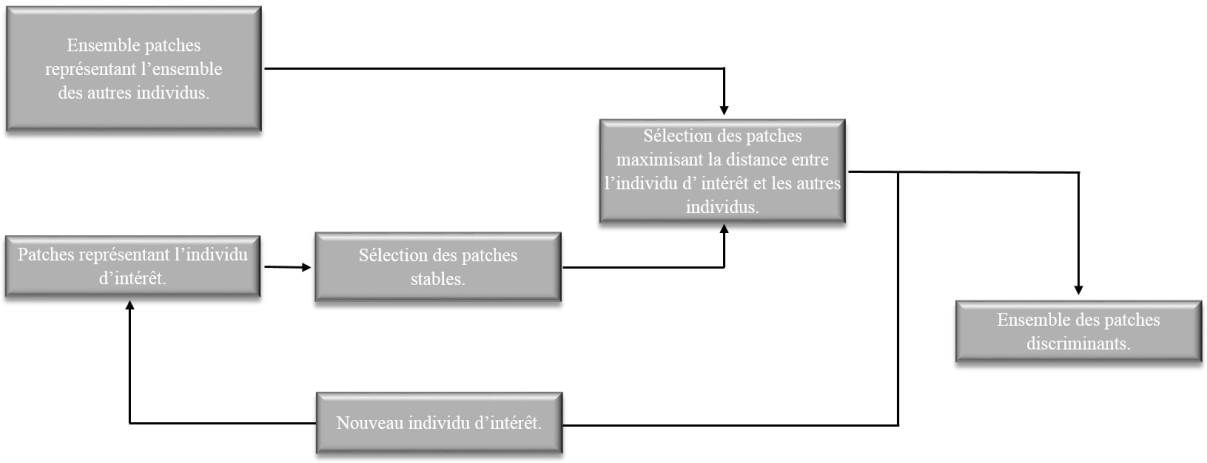


Figure 59. Sélection de patches.

3.4.2. Appariement

La deuxième partie de notre algorithme de ré-identification est l'appariement. A la fin de la première partie de notre stratégie de ré-identification, nous obtenons un ensemble d'individus détectés dans le champ de vision d'une caméra donnée. Ces individus sont les candidats auxquels nous allons comparer les individus détectés dans les autres caméras. Pour chacun de ces candidats, nous avons sélectionnés un ensemble de patches discriminants durant la partie d'apprentissage. Donc, les données qui seront utilisées dans l'appariement sont constituées d'un ensemble candidats associés à leurs patches les plus discriminants et d'un ensemble de nouvelles détections que nous devons associer à l'un des candidats. Un exemple de ces données est illustré dans la *Figure 60*.

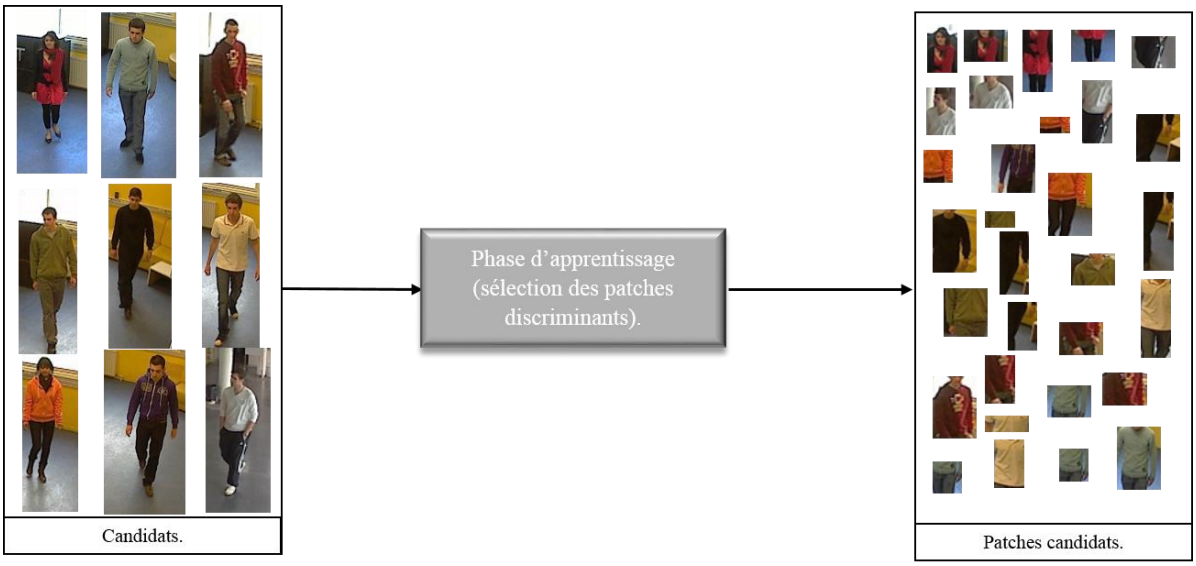


Figure 60. Données à la fin de la première phase de ré-identification.

Afin d'effectuer l'appariement entre les nouvelles détections et les candidats, nous avons utilisés deux méthodes différentes. La première repose sur un appariement direct tandis que la deuxième repose sur un appariement par apprentissage.

3.4.2.1. Appariement direct

L'appariement direct pour associer les nouvelles détections aux candidats, se déroule en trois étapes. Nous commençons par l'extraction des descripteurs de tous les patches sélectionnés précédemment pour l'ensemble des individus candidats. Ensuite, nous créons un ensemble qui rassemble tous les patches sélectionnés dans la première phase de ré-identification. Cet ensemble sera considéré comme l'ensemble des candidats. Enfin, durant l'appariement chaque patch de cet ensemble est comparé avec le patch de requête (individus à ré-identifier) correspondant à la même taille et la même position. Dans cette partie, nous avons utilisé comme descripteurs les histogrammes de couleurs dans l'espace *HSV* en raison de leur simplicité, de leur efficacité et de leur stabilité relative aux changements d'éclairage. Pour la comparaison des histogrammes, nous avons utilisé la distance *EMD* qui donne de meilleurs résultats qu'une distance simple. Cette procédure est illustrée dans la *Figure 61*.

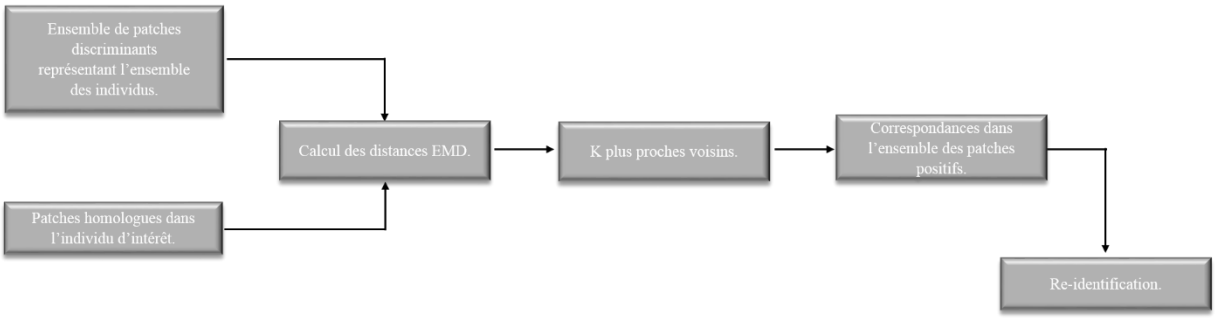


Figure 61. Appariement direct.

Notre algorithme d'association d'individus est basé sur l'utilisation des K-plus proches voisins pour sélectionner les apparences les plus proches à chaque nouvelle détection. Nous commençons pour chaque nouvelle détection (individu requête) par le calcul des distances *EMD* entre tous les patches candidats (l'ensemble candidats créé précédemment) et les patches de l'individu requête. Ces distances sont utilisées dans la sélection des k plus proche patches voisins. A la fin de cette étape, nous obtenons un ensemble de patches voisins qui sont utilisés pour confirmer la ré-identification. Pour ce faire, nous calculons le nombre de patches appartenant à chacun des individus candidats dans l'ensemble des patches voisins. Ensuite, nous confirmons la ré-identification si le nombre de patches appartenant au même candidat est supérieur à un seuil et si la distance globale entre tous les patches de ce candidat et les patches de l'individu requête est inférieur à un autre seuil. Un exemple de résultat d'appariement est montré dans la *Figure 62*.

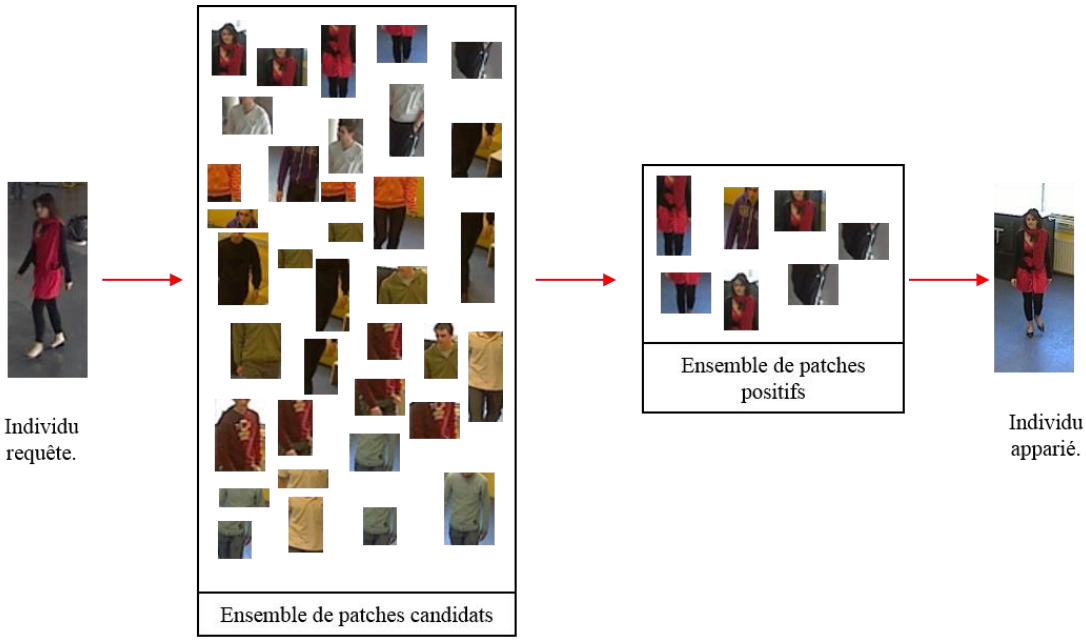


Figure 62. Exemple d'appariement.

3.4.2.2. Appariement par apprentissage

Pour l'appariement par apprentissage, nous avons alors choisi d'utiliser des classificateurs binaires pour leur performance. Nous avons commencé par ramener notre problème de classification multi-classes vers un problème de classification binaire. Dans la problématique originale, nous devons comparer chaque nouvel individu à l'ensemble des candidats ; ce qui veut dire que nous avons une problématique de classification multi-classes avec n classes si nous considérons un ensemble de n candidats. Ce type de classification est beaucoup moins performant qu'une classification binaire avec seulement deux classes. De ce fait, nous avons implémenté un algorithme de classification binaire qui classe les paires de patches en paires positives pour les patches représentant la même région d'images appartenant au même individu et paires négatives de patches pour les paires de patches représentant la même région d'images mais avec des images qui représentent des individus différents.

Cette partie est faite en deux phases, la première qui consiste à apprendre un classifieur binaire pour distinguer entre les paires positives et les paires négatives de patches, et la deuxième qui consiste dans l'utilisation de ce classificateur pour associer les patches discriminants sélectionnés précédemment avec les patches de nouveaux individus. Cette association nous permet de confirmer la ré-identification en fonction du nombre de patches positifs appartenant à chaque candidat. Pour cela, nous avons utilisé

deux algorithmes d'apprentissage, le premier est basé sur l'utilisation d'un SVM et le deuxième basé sur l'apprentissage profond. Ce principe d'appariement est détaillé dans la *Figure 63*.

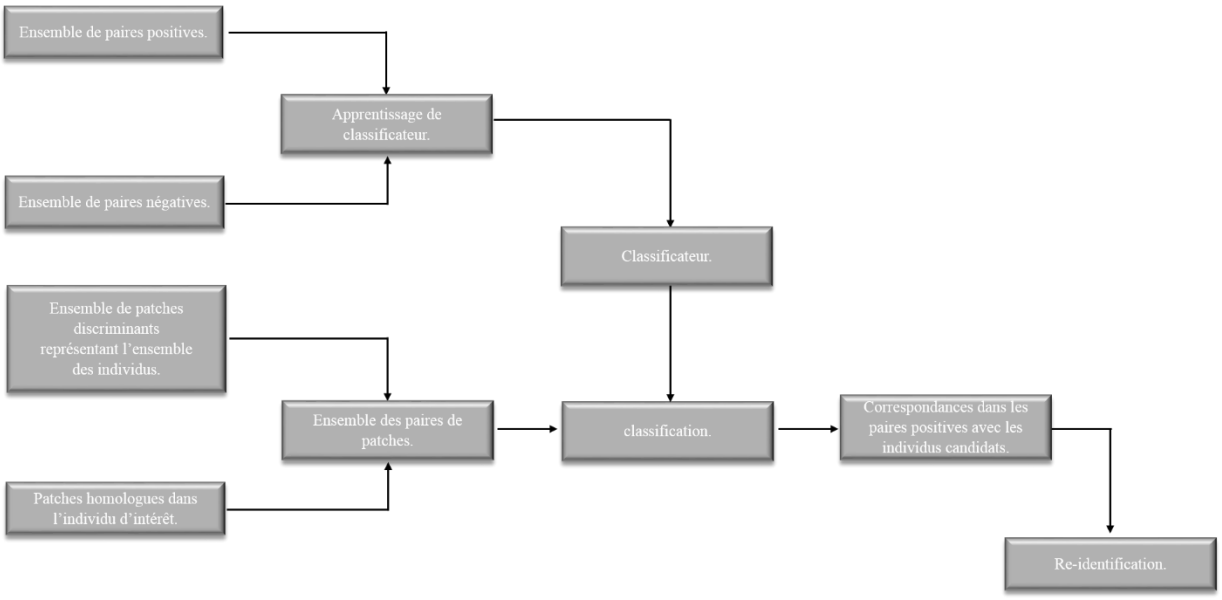


Figure 63. Appariement par apprentissage.

3.4.2.2.1. Appariement avec SVM

La première méthode d'apprentissage que nous avons utilisé pour l'association des individus entre les différentes caméras est basée sur l'utilisation d'un SVM. Cet appariement est réalisé en deux étapes. La première étape se résume dans l'apprentissage d'un SVM binaire sur un ensemble annoté de paires de patches. Ce SVM est ensuite utilisé dans la deuxième étape sur des nouvelles paires de détections pour associer les nouveaux individus détectés dans le champ de vision d'une caméra donnée à ceux détectés dans le champ de vision d'une autre caméra. Le schéma général de cette procédure d'appariement est illustré dans la *Figure 64*.

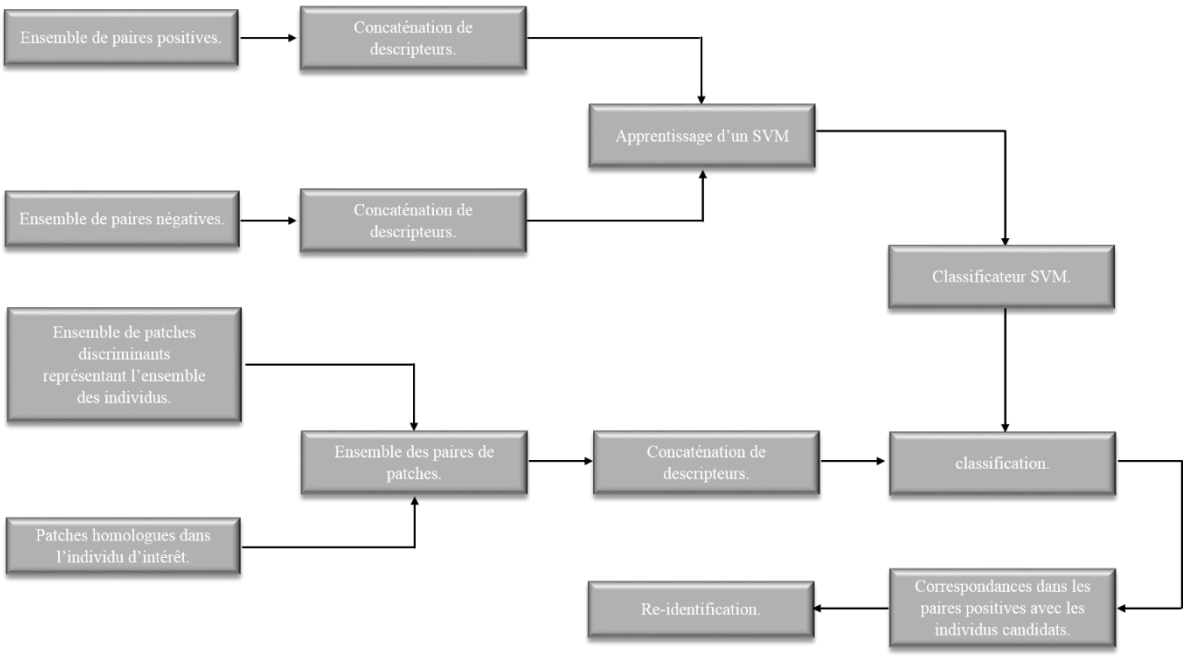


Figure 64. Appariement par SVM.

Pour l'entrainement de notre SVM, nous avons commencé par la transformation de nos images dans l'espace de couleur HSV pour avoir plus de stabilité par rapport à l'éclairage. Nous avons pris les paires positives de patches qui représentent le même individu et les paires négatives de patches qui représentent des individus différents. Pour chaque paire de patches nous avons calculé un descripteur qui consiste dans la concaténation des histogrammes de couleurs des deux patches formant la paire. Nous avons utilisé ces descripteurs pour former deux ensembles, le premier contient les paires positives et le deuxième contient les paires négatives. Ces deux ensembles de descripteurs sont utilisés dans l'apprentissage de notre SVM. A la fin de cette phase, nous obtenons un classificateur capable de distinguer les paires positives et les paires négatives de patches.

Après l'apprentissage, nous obtenons un SVM entrainé à distinguer les paires positives et négatives de patches. Dans cette étape, nous utilisons alors ce SVM pour associer les individus détectés dans les différentes caméras. Nous commençons par la construction d'un ensemble qui regroupe tous les patches discriminants sélectionnés précédemment pour l'ensemble des individus. Cet ensemble est considéré comme l'ensemble des patches candidats. Ensuite, pour l'appariement, tous les patches candidats sont comparés à leurs homologues qui ont la même position et la même taille dans l'individu requête (individu à ré-identifier). Pour chaque nouvelle détection, nous créons

des paires qui regroupent chaque patch candidat avec son homologue de la nouvelle détection et nous calculons les descripteurs caractérisant ces paires de patches. Ensuite, nous utilisons le SVM que nous avons entrainé précédemment pour classer les paires de patches en paires positives ou paires négatives, et nous nous focalisons sur les paires positives. Ce processus se termine par l'obtention d'un ensemble de paires positives de patches qui sont considérées par le classifieur comme représentant le même individu. Cet ensemble de paires positives nous permet de confirmer la ré-identification. Cela est fait par le comptage des nombres de patches appartenant à chaque individu candidat dans l'ensemble des paires positives de patches. A la fin, nous confirmons une ré-identification si le nombre de patches appartenant à un individu donné est supérieur à un seuil. Un exemple d'appariement par SVM est illustré dans la *Figure 65*.

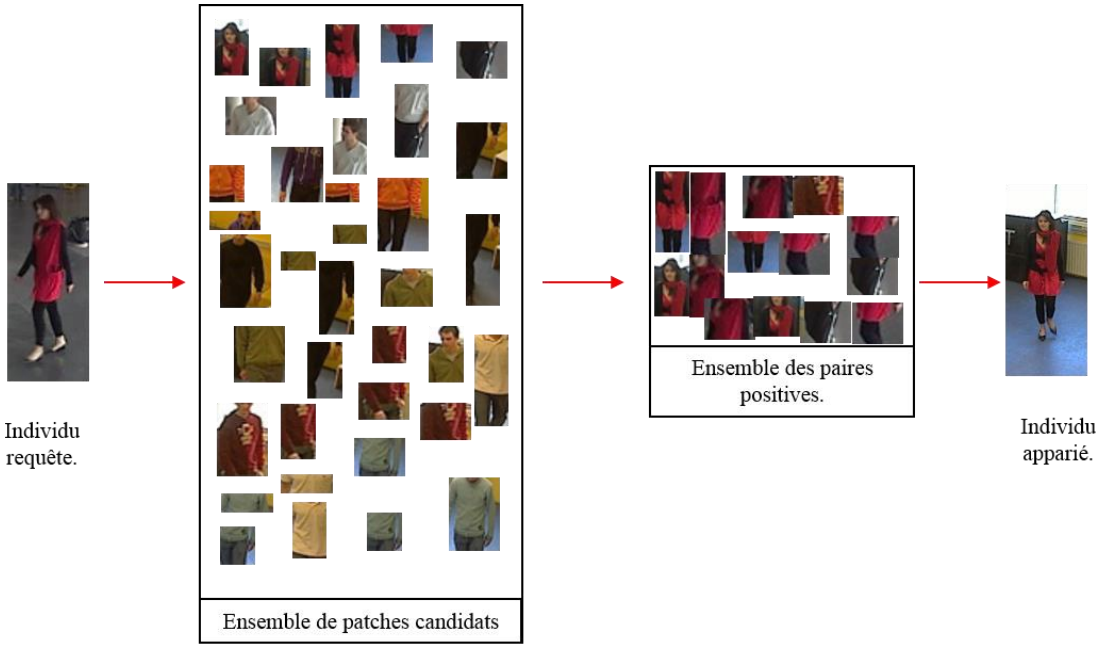


Figure 65. Exemple d'appariement par SVM.

3.4.2.2.2. Appariement avec apprentissage profond

Cette phase consiste dans l'appariement des individus détectés dans les différentes caméras à l'aide d'une technique basée sur l'utilisation d'apprentissage profond. Cette stratégie d'appariement est utilisée pour associer les patches discriminants sélectionnés auparavant pour confirmer la ré-identification. Nous avons implémenté une méthode basée sur l'utilisation d'un classifieur CNN binaire pour classer les paires positives et les paires négatives de patches. Les paires positives représentent le cas où les deux patches représentent la même région du même individu dans les deux caméras tandis que les paires négatives représentent le cas où les deux patches sont issus d'individus différents.

Cet appariement est réalisé en deux étapes. La première est l'apprentissage d'un CNN sur un ensemble annoté de paires de détections pour distinguer entre les paires positives et les paires négatives. Ensuite, la deuxième étape consiste dans l'utilisation de ce classifieur sur des nouvelles paires de détections pour associer les individus entre les différentes caméras. Ce schéma de base est illustré dans *Figure 66*.

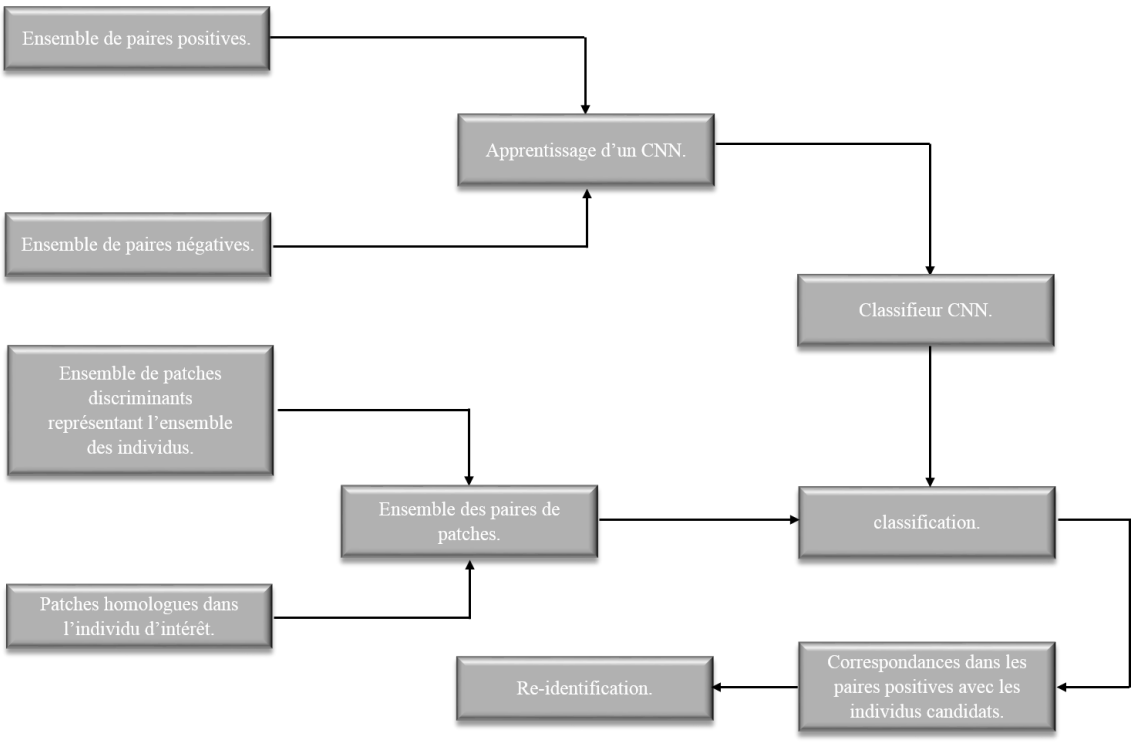


Figure 66. Appariement par apprentissage profond.

Pour ce faire, nous avons implémenté un CNN similaire à celui proposé dans [117]. Nous avons créé un CNN composé de 4 couches parmi lesquelles deux sont convolutives et deux autres entièrement connectées et une couche pour le calcul de différence (comme montré dans la *Figure 67*).

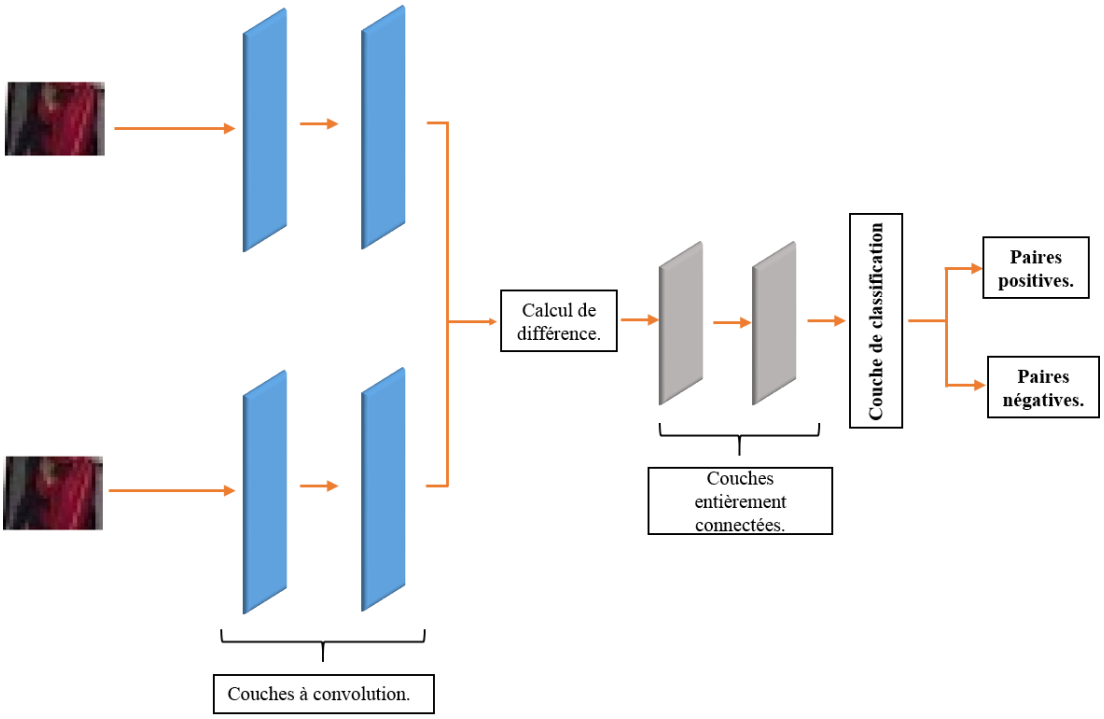


Figure 67. Architecture du CNN utilisé.

Pour entrainer notre classifieur, nous avons commencé par la création des ensembles d'images de paires positives $diff_p$ et de paires négatives $diff_n$. L'ensemble d'images $diff_p$ est créé à partir des paires positives de patches obtenus du même individu dans les deux caméras, tandis que l'ensemble $diff_n$ est créé à partir des paires négatives de patches obtenus d'individus différents dans les deux caméras. Ces deux ensembles de données sont utilisé pour entrainer le CNN qui sera utilisé ultérieurement pour distinguer entre les paires positives et les paires négatives de patches.

Les images des ensembles $diff_p$ et $diff_n$ sont des images RGB de dimension 32∗32∗3. Ensuite, la première couche de convolution filtre ces images d'entrées $32 * 32 * 3$ par 16 noyaux de taille $5 * 5 * 3$. La sortie de cette première couche est sous-échantillonnée par un facteur de 2 à l'aide d'une couche de pooling. Ensuite, la deuxième couche de convolution prend et filtre la sortie précédente par 12 noyaux de taille $5 * 5 * 3$. La sortie de cette couche est à nouveau sous-échantillonnée par une couche de pooling avant d'être utilisée dans le calcul de différence entre les deux patches (Cette procédure est illustrée dans la *Figure 68*). Le résultat de ce calcul de différence est utilisé comme entrée pour la première couche entièrement-connectée qui est constitué de 200 neurones. La sortie de cette couche est à nouveau utilisée comme entrée pour la deuxième couche entièrement-

connectée, elle-même connecté à la couche de classification qui génère deux classes, une classe positive pour les paires positives et une classe négative pour les paires négatives.

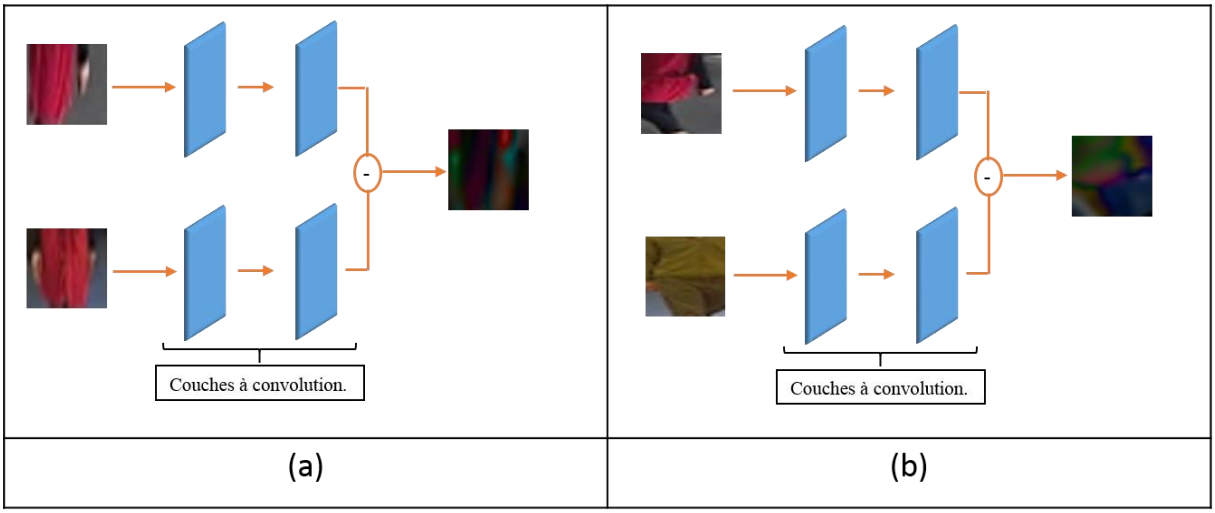


Figure 68. Images de différence : (a) paire positive, (b) paire négative.

Dans les étapes précédentes, nous avons décrit les données utilisées par le CNN ainsi que son architecture. Une fois le classifieur implémenté et les données d'entrée préparées, la prochaine étape consiste dans l'entrainement de ce classifieur sur un ensemble de paires positives et négatives de patches pour obtenir un classifieur capable de les distinguer. Pour cette tâche, nous avons rassemblé deux ensembles de paires, le premier contenant les paires positives et le deuxième contenant les paires négatives. Ces deux ensembles sont utilisés pour entrainer le classifieur. A la fin de cette phase, nous obtenons un classifieur capable de distinguer entre les paires positives et les paires négatives de patches.

La phase précédente nous permet d'obtenir un classifieur entrainé pour distinguer entre les paires positives et négatives de patches. Dans cette étape, nous utilisons ce classifieur pour associer les individus détectés dans les différentes caméras afin de pouvoir reconstruire leurs trajectoires globales. Nous commençons par la construction d'un ensemble qui regroupe tous les patches discriminants sélectionnés précédemment pour l'ensemble des individus. Cet ensemble est considéré comme l'ensemble des patches candidats. Ensuite, pour l'appariement tous les patches candidats sont comparés à leurs homologues qui ont la même position et la même taille dans l'individu requête (individu à ré-identifier). Pour chaque nouvelle détection, nous comparons l'ensemble des patches candidats à leurs homologues à l'aide du classifieur CNN défini précédemment. Ce

processus se termine par l'obtention d'un ensemble de paires positives de patches considérées par le classifieur comme représentant le même individu. Cet ensemble de paires positives nous permet de confirmer la ré-identification. Cela est fait par le comptage des nombres de patches appartenant à chaque individu candidat dans l'ensemble des paires positives de patches. A la fin, nous confirmons une ré-identification si le nombre de patches appartenant à un individu donné est supérieur à un seuil. Un exemple d'appariement par apprentissage profond est illustré dans la *Figure 69*.

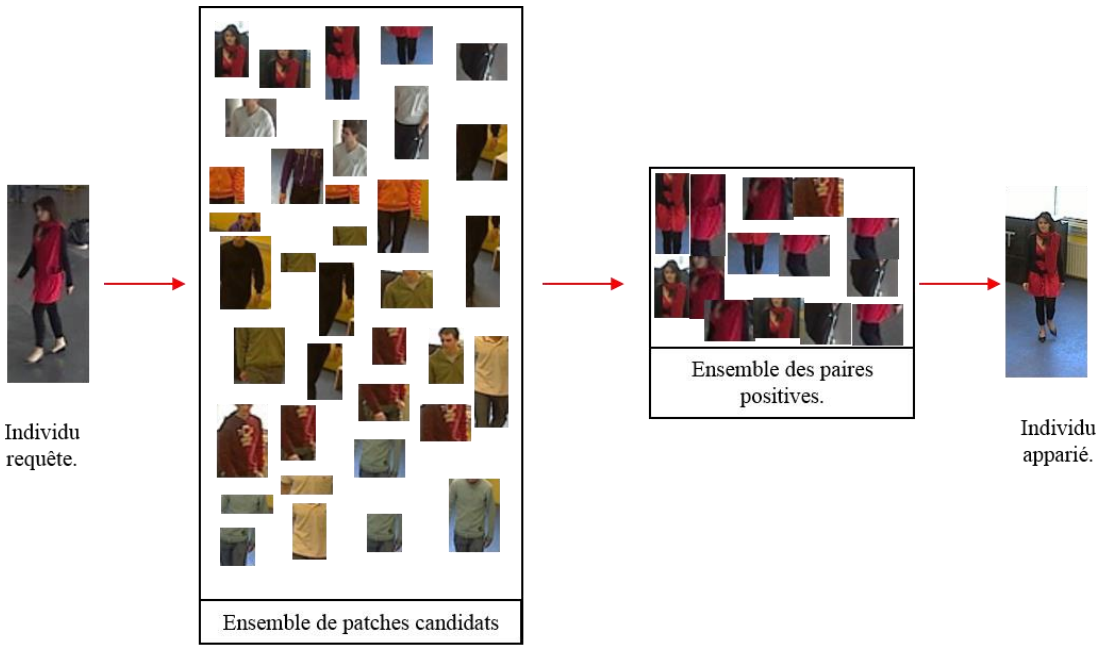


Figure 69. Exemple d'appariement par apprentissage profond.

3.5. Evaluation

Durant la dernière décennie, plusieurs travaux ont été effectués et plusieurs méthodes ont été proposées dans le cadre de la ré-identification de personnes dans un réseau de caméras. De ce fait, le besoin de bases de données qui peuvent être utilisées dans l'évaluation de ces méthodes est devenu indispensable. Cela a conduit à la construction de plusieurs bases de données pour la ré-identification de personnes dans un système multi-caméras [118]. Parmi ces bases de données nous pouvons mentionner quelques-unes parmi les plus utilisées comme VIPeR [119] pour les méthodes basées sur l'utilisation d'une seule détection pour chaque individu (single-shot methodes), CAVIAR pour le suivi de personnes, 3DPeS [120] pour la ré-identification en 3D, et Aziz et al [104] pour les

méthodes basées sur l'utilisation de plusieurs détections pour chaque individus (*multiple shot methodes*).

Cependant, à l'exception de la base Aziz et al [104], les autres bases sont fournies avec seulement des images brutes d'individu et avec un petit nombre de détections pour chaque individu. Ces restrictions limiteraient la validation de notre approche, notamment pour la sélection des patches les plus discriminants pour les individus à ré-identifier. C'est pourquoi nous avons choisi d'utiliser la base de données utilisée dans Aziz et al [104] pour la validation de notre approche, et la base VIPeR [119] pour la comparaison avec les approches existantes, même en partant avec un petit handicap dans la sélection des régions saillantes causé par le nombre limité de détections fournies pour chaque individu dans la base VIPeR.

3.5.1. Bases de données utilisées

Pour l'évaluation de notre algorithme, nous avons utilisé la base déjà utilisée dans Aziz et al [104] et Merad et al [6]. Cette base est enregistrée dans 3 caméras qui filment la scène étudiée de trois angles différents (montré dans la *Figure 70*).

Figure 70. Positionnement des caméras.

Dans les différentes scènes, 9 individus (montrés dans la *Figure 71*) ont été enregistrés lors de leur passage dans le champ de vision de chacune des caméras. Les vidéos de cette base ont été prises avec une résolution de 320×240 pixels. La fréquence d'images varie entre 20 et 25 *fps* (images par seconde). Dans chaque scène, tous les individus ont environ 200 apparences pour différents scénarios. Les séquences vidéo ont une durée variant de 15 à 25 secondes. Les vidéos issues des trois caméras ont les caractéristiques suivantes :

- Les images sont de bonne qualité ;
- L'activité humaine est multiple, incluant des variations d'échelle, des occultations et des interactions entre les personnes.

Figure 71. Individus présents dans la base de données.

Pour la comparaison de notre approche avec ceux déjà existants, nous avons utilisé la base VIPeR [119]. Cette base est l'une des plus utilisée dans le contexte de ré-identification de personnes. Elle contient 632 paires d'individus filmés dans deux caméras avec une variation de pose et d'éclairage. Ces variations rendent cette base parmi les plus compliquées pour la ré-identification des personnes. Mais d'un autre côté, cette base fournie une seule image par personne, cette restriction nous a contraint à faire quelques changements dans la phase de sélection des régions saillantes de notre méthode. Un exemple illustrant quelques individus dans cette base est illustré dans la *Figure 72*.

Figure 72. Exemple illustrant les individus dans VIPeR.

3.5.2. Dans le cadre d'un réseau de caméras

Dans cette partie, nous avons évalué notre algorithme dans le contexte de ré-identification de personnes dans un réseau de caméras qui ne se chevauchent pas. Nous avons utilisé les individus détectés dans le champ de vision de la caméra 1 comme candidats, tandis que les individus détectés dans les caméras 2 et 3 sont utilisées comme individus requêtes. Ensuite, nous avons utilisé les différentes détections pour chaque individu (10 détections par individu) afin de construire l'ensemble des régions saillantes (pour ces tests chaque individu est représenté par un nombre varié de patches). Une fois l'ensemble des patches discriminants sélectionnés, la prochaine étape consiste dans l'appariement avec les individus requêtes. Nous avons alors associé l'ensemble des patches discriminants avec leurs homologues de chacun des individus requêtes en utilisant les différentes méthodes d'appariement expliquées précédemment. A la fin, l'ensemble des patches appariés est utilisé pour confirmer la ré-identification. La performance de notre système est évaluée par le calcul du métrique précision-rappel comme indiqué par les équations (29) et (30).

$$\text{rappel}_i = \frac{\text{nombre de vrais positive}}{\text{nombre de cible de l'individu i}} \tag{29}$$

$$\text{précision}_i = \frac{\text{nombre de vrais positive}}{\text{nombre de vrais positives} + \text{nombre de faux positives}} \quad (30)$$

Nous avons aussi utilisé les pourcentages de précision et de rappel qui sont calculés comme indiquer par les équations (31) et (32).

$$\text{rappel} = \frac{\sum_{i=1}^{n} \text{rappel}_i}{n} \quad (31)$$

$$\text{précision} = \frac{\sum_{i=1}^{n} \text{précision}_i}{n} \quad (32)$$

Où : $rappel_i$ est le nombre de bons appariements pour la personne i divisé par le nombre de bons appariements possible pour cette personne, et la $précision_i$ est le nombre de bons appariements pour la personne i divisé par le nombre total d'appariements retournés pour cette personne.

La première série de tests est utilisée pour évaluer la performance de notre algorithme dans le contexte de la ré-identification de personnes dans un réseau de caméras. Nous avons utilisé la base de données décrite précédemment et nous avons fait une centaine de requêtes pour les différents individus. Nous avons varié le nombre de patches pour chaque individu. Ensuite, nous avons calculé les pourcentages de vrai et faux positive pour chaque individu et nous avons représenté les matrices de confusion pour chaque procédure d'appariement. Ces résultats sont montrés dans la section 3.5.2.1.1.

La deuxième série de tests est utilisée pour évaluer notre algorithme en termes de précision-rappel comme fait dans [38]. Nous avons formulé une centaine de requêtes pour les différents individus et nous avons fait varier le nombre de patches par individu. Ensuite, nous avons utilisé les différentes méthodes d'appariement et nous avons calculé la métrique précision-rappel pour chaque cas. Les résultats de ces tests sont illustrés dans la section 3.5.2.1.1.

La troisième série de tests est utilisé pour évaluer la stabilité de notre algorithme en utilisant les Courbes Cumulatives de Correspondance (CMC). Nous avons utilisé le même protocole de test comme pour les précédentes séries de tests et nous avons tracé les

courbes cumulatives de correspondance pour les différentes méthodes d'appariement. Les résultats de ces tests sont illustrés dans la section 3.5.2.1.1.

Une fois que nous avons évalué les trois méthodes d'appariement séparément. Nous avons réalisé une autre série de tests pour comparer les performances des différentes méthodes. Nous avons utilisé le même protocole de test comme décrit précédemment et nous avons utilisé les pourcentages de précision-rappel ainsi que les courbes cumulatives de correspondance (CMC) pour comparer les différents appariements. Les résultats de ces tests sont illustrés dans la section 3.5.2.1.2.

La quatrième série de tests est utilisée pour montrer l'apport de notre procédure de sélection de patches. Pour ce faire, nous avons réalisé une série de tests en utilisant le même protocole de test décrit précédemment, et nous avons utilisé le métrique précision-rappel pour comparer la performance de notre algorithme avec deux méthodes d'association ; une association par descripteurs globaux (sans utilisation de patches) et une association par une sélection aléatoire de patches. Cette comparaison nous a permis de voir l'apport de notre méthode basée sur la sélection de patches discriminants. Les résultats de ces tests sont illustrés dans la section 3.5.2.1.3.

La dernière série de tests est utilisé pour comparer notre algorithme avec les méthodes existantes de ré-identification de personnes dans un réseau de caméras. Nous avons commencé par l'utilisation de la première base de données pour comparer notre algorithme avec celui proposé dans [104] et [6]. Nous avons utilisé le même protocole de test et nous avons utilisé la métrique précision-rappel pour la comparaison des performances. Nous avons aussi comparé notre méthode à plusieurs méthodes existantes en utilisant la base VIPeR. Dans cette série de tests, nous étions obligé de faire quelques changements dans la phase de sélection des régions saillantes vu les restrictions imposées par la base qui fournit une seule image par personne. Ces restrictions nous ont obligés à partir avec un petit handicap en passant directement à la deuxième partie de notre procédure de sélection sans pouvoir choisir les patches les plus stables. Les résultats de ces tests sont illustrés dans la section 3.5.2.1.3.

3.5.2.1. Résultats

3.5.2.1.1. Différentes méthodes d'appariement

Dans la première série de tests, nous avons évalué la performance de notre algorithme avec la représentation des matrices de confusion. Nous avons évalué les trois différents appariements avec différents nombre de patches sélectionnés. Les figures (*Figure 73*,

Figure 74 et *Figure 75*) montrent les matrices de confusion obtenues avec 10 patches pour chaque individu.

Id	P1	P2	P3	P4	P5	P6	P7	P8	P9
P1	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
P2	0,00	0,90	0,00	0,00	0,00	0,10	0,00	0,00	0,00
P3	0,00	0,00	0,90	0,00	0,00	0,00	0,10	0,00	0,00
P4	0,00	0,00	0,00	0,80	0,10	0,00	0,10	0,00	0,00
P5	0,10	0,10	0,00	0,00	0,80	0,00	0,00	0,00	0,00
P6	0,00	0,00	0,00	0,00	0,00	0,80	0,00	0,20	0,00
P7	0,00	0,00	0,00	0,00	0,10	0,00	0,90	0,00	0,00
P8	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,80	0,20
P9	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00

Figure 73. Matrice de confusion pour un appariement direct avec la distance EMD et un ensemble de 10 patches discriminants pour chaque individu.

Id	P1	P2	P3	P4	P5	P6	P7	P8	P9
P1	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
P2	0,00	0,80	0,00	0,00	0,00	0,20	0,00	0,00	0,00
P3	0,00	0,00	0,75	0,00	0,00	0,00	0,25	0,00	0,00
P4	0,00	0,00	0,00	0,80	0,10	0,00	0,10	0,00	0,00
P5	0,05	0,10	0,00	0,00	0,85	0,00	0,00	0,00	0,00
P6	0,05	0,10	0,00	0,00	0,00	0,70	0,00	0,15	0,00
P7	0,05	0,00	0,00	0,00	0,05	0,00	0,90	0,00	0,00
P8	0,00	0,00	0,00	0,00	0,00	0,05	0,00	0,75	0,20
P9	0,00	0,00	0,00	0,00	0,00	0,05	0,00	0,00	0,95

Figure 74. Matrice de confusion pour un appariement avec apprentissage d'un SVM et un ensemble de 10 patches discriminants pour chaque individu.

Id	P1	P2	P3	P4	P5	P6	P7	P8	P9
P1	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
P2	0,00	0,95	0,00	0,00	0,00	0,05	0,00	0,00	0,00
P3	0,00	0,00	0,90	0,00	0,00	0,05	0,05	0,00	0,00
P4	0,00	0,00	0,00	0,90	0,05	0,00	0,05	0,00	0,00
P5	0,10	0,05	0,00	0,00	0,85	0,00	0,00	0,00	0,00
P6	0,00	0,00	0,00	0,00	0,00	0,95	0,00	0,05	0,00
P7	0,00	0,05	0,00	0,00	0,10	0,00	0,85	0,00	0,00
P8	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,90	0,10
P9	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00

Figure 75. Matrice de confusion pour un appariement par apprentissage profond et un ensemble de 10 patches discriminants pour chaque individu.

Ces matrices de confusion montrent que notre algorithme est capable d'obtenir des taux élevés de ré-identification qui dépassent les 80% pour la majeure partie des individus. Ces taux peuvent atteindre les 100% pour certains individus. Nous avons obtenu des taux élevés pour les différentes méthodes d'appariements.

Dans la deuxième série de tests, nous avons utilisé le métrique rappel-précision pour évaluer la performance de notre algorithme. Les figures (*Figure 76*, *Figure 77* et *Figure 78*) montrent le rappel et la précision obtenue pour chaque méthode d'appariement 10 patches pour chaque individu.

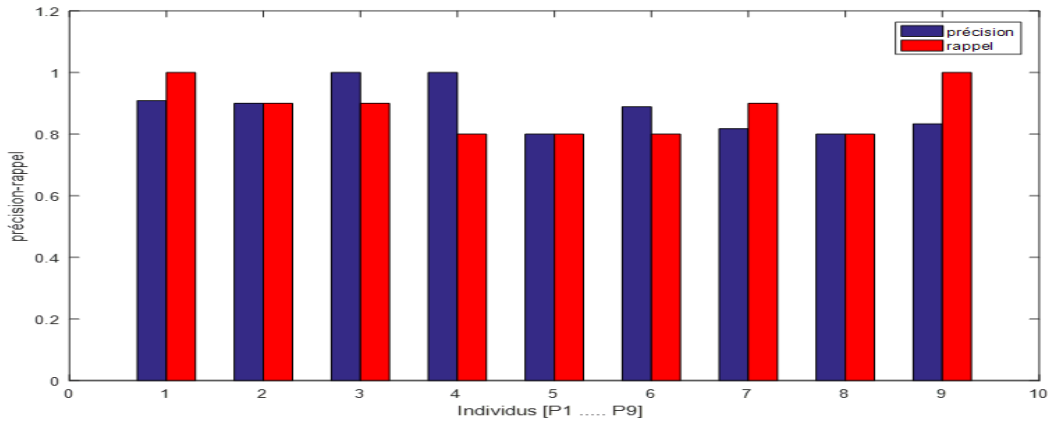


Figure 76. Rappel-précision pour un appariement direct avec la distance EMD et un ensemble de 10 patches discriminants pour chaque individu.

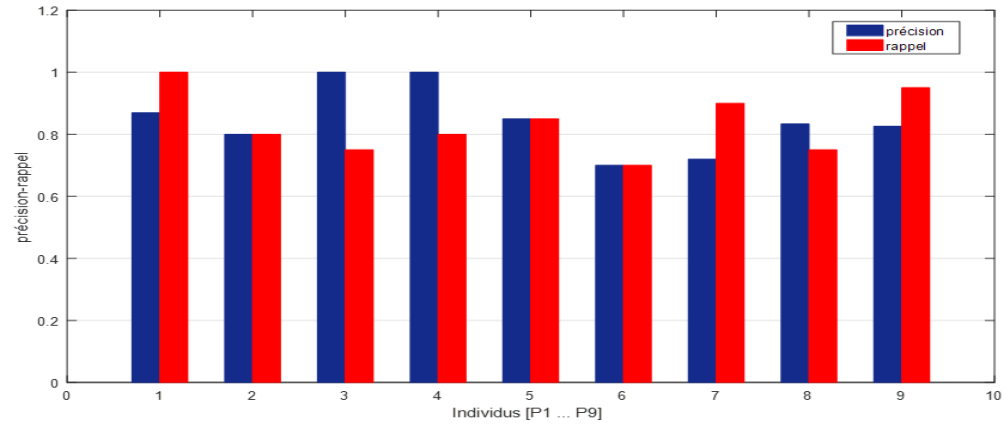


Figure 77. Rappel-précision pour un appariement avec apprentissage d'un SVM et un ensemble de 10 patches discriminants pour chaque individu.

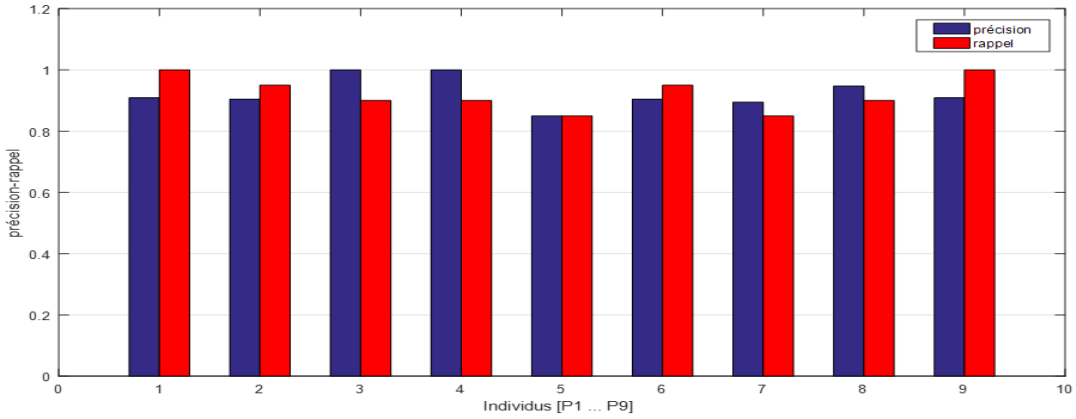


Figure 78. Rappel-précision pour un appariement par apprentissage profond et un ensemble de 10 patches discriminants pour chaque individu.

Les diagrammes de rappel-précision montrent que notre algorithme est capable d'obtenir un taux élevé de rappel tous en gardant une bonne précision. Les taux de rappel variés entre 80% et 100% tandis que la précision garde des valeurs variant entre 70% et 100% pour les différents individus. Ces résultats sont obtenus avec les différentes méthodes d'appariements.

La troisième série de tests est utilisé pour évaluer la stabilité de notre algorithme. Pour ce faire, nous avons représenté les courbes CMC pour les différents appariements. Ces courbes sont montrées dans les figures (*Figure 79*, *Figure 80* et Figure **81**) pour des tests de ré-identification avec 10 patches par individu.

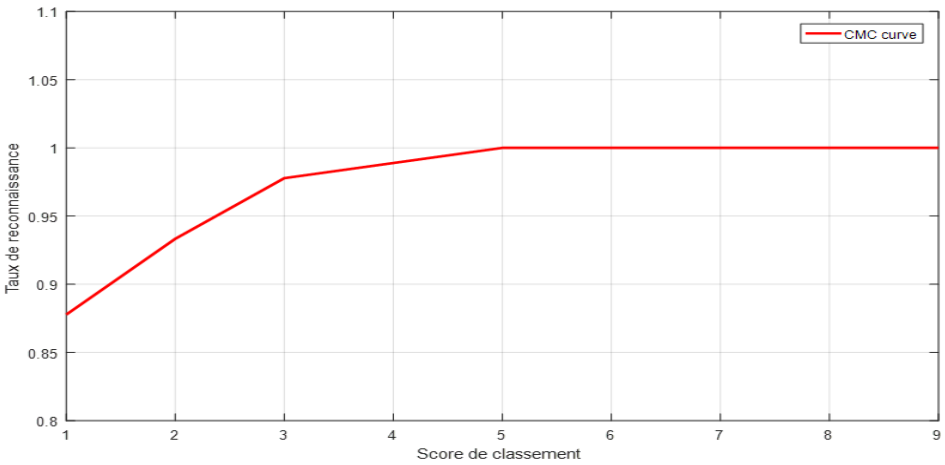


Figure 79. La courbe CMC pour un appariement direct avec la distance EMD et un ensemble de 10 patches discriminants pour chaque individu.

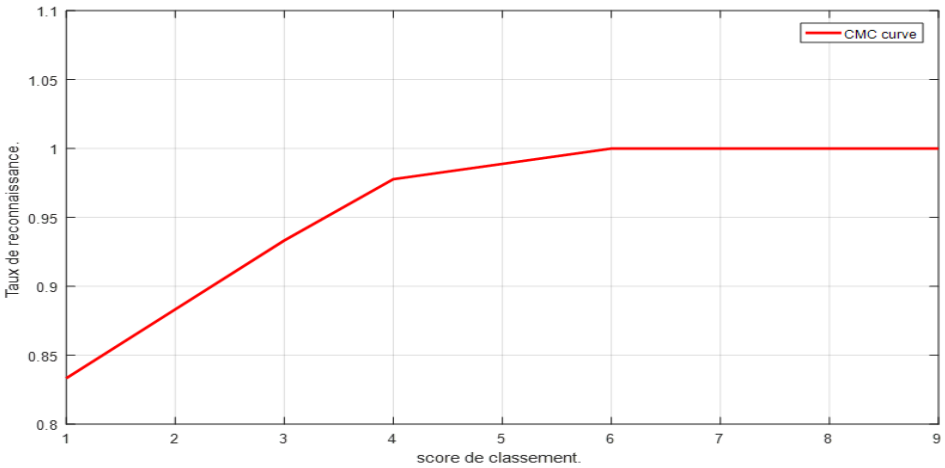


Figure 80. La courbe CMC pour un appariement avec apprentissage d'un SVM et un ensemble de 10 patches discriminants pour chaque individu.

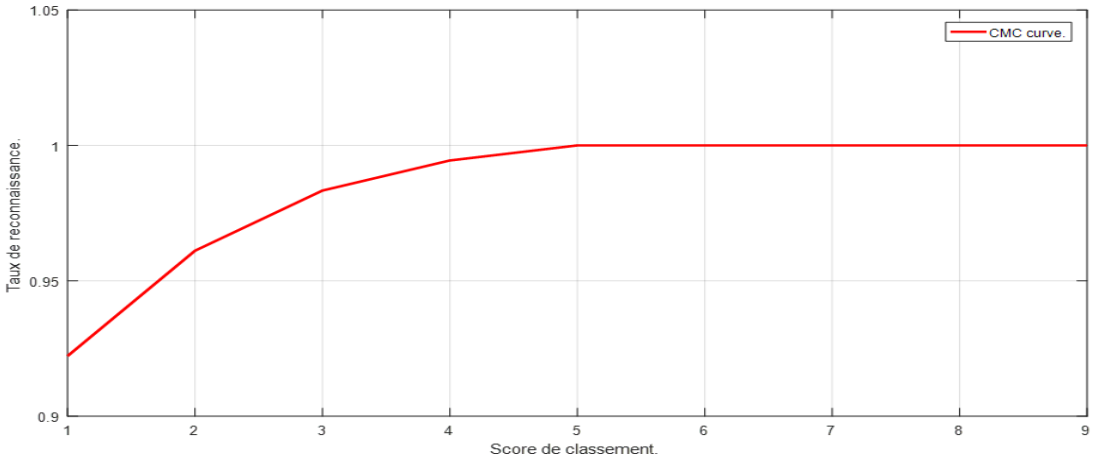


*Figure **81**. La courbe CMC pour un appariement par apprentissage profond et un ensemble de 10 patches discriminants pour chaque individu.*

Les courbes CMC obtenues convergent rapidement vers 1. Cette convergence est un critère important pour les systèmes de ré-identification car elle montre que les confusions sont apparues entre les individus qui ont des apparences approximativement similaires. Cette convergence montre également que notre algorithme est relativement stable dans les trois cas d'appariement.

3.5.2.1.2. Comparaison entre différent appariements

Dans cette section, nous avons comparé les différentes méthodes d'appariement. Pour ce faire, nous avons utilisé le Tableau 4 ainsi que les courbes CMC montrées dans la *Figure 82*.

méthode / Id	Résultats de la re-identification		
	Appariement direct	Avec SVM	Avec CNN
P1	1,00	1,00	1,00
P2	0,90	0,80	0,95
P3	0,90	0,75	0,90
P4	0,80	0,80	0,90
P5	0,80	0,85	0,85
P6	0,80	0,70	0,95
P7	0,90	0,90	0,85
P8	0,80	0,75	0,90
P9	1,00	0,95	1,00
Moyenne	0.87	0.83	0.92

*Tableau **4**. Comparaison entre différents appariements avec un ensemble de 10 patches discriminants pour chaque individu.*

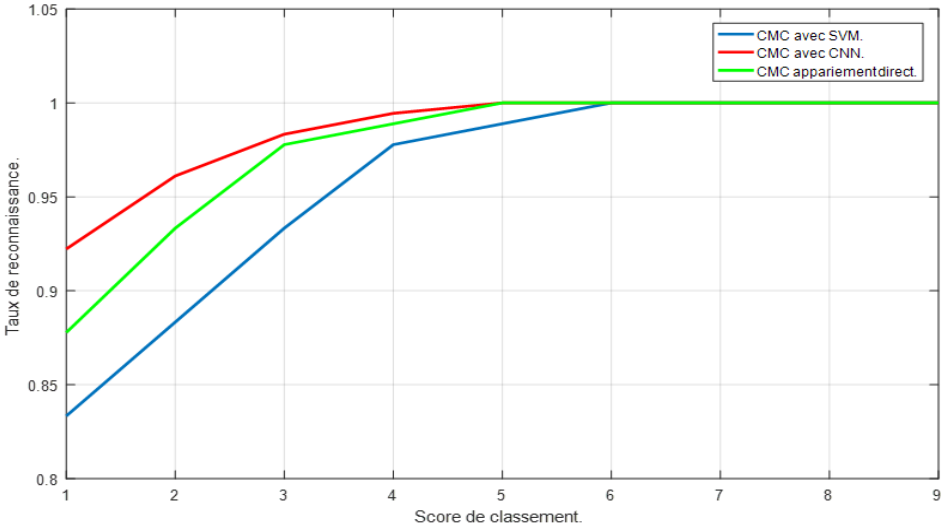


Figure 82. Comparaison entre les courbes CMC des différents appariements avec un ensemble de 10 patches discriminants pour chaque individu.

Ces résultats montrent que l'apprentissage profond donne les meilleurs taux de ré-identification dans l'ensemble, tandis que l'apprentissage d'un SVM donne les moins bons résultats. Mais globalement, tous les résultats obtenus par les différents appariements restent acceptables. En termes de stabilité les courbes CMC montrent qu'il n'y a pas une grande différence de stabilité entre les différents appariements.

3.5.2.1.3. Comparaison avec les méthodes existantes

Dans cette partie, nous allons comparer nos résultats avec les résultats obtenus par les méthodes déjà existantes de ré-identification. Pour ce faire, nous avons réalisé différentes séries de tests.

Nous avons commencé par l'étude de l'apport de notre méthode de sélection de patches dans la performance de la ré-identification. Pour ce faire, nous avons comparé les résultats obtenus par notre méthode avec une ré-identification simple par descripteur de couleur ainsi qu'une ré-identification par sélection aléatoire de patches. Les résultats de ces tests sont montrés dans le Tableau 5 et les diagrammes de rappel-précision des figures (*Figure 83* et *Figure 84*).

Id \ méthode	Résultats sans sélection de patches,		Résultats avec sélection de patches		
	Descripteur global	Patches aléatoires	Appariement direct	Avec SVM	Avec CNN
P1	0,95	0,60	1,00	1,00	1,00
P2	0,15	0,10	0,90	0,80	0,95
P3	0,85	0,65	0,90	0,75	0,90
P4	0,15	0,45	0,80	0,80	0,90
P5	0,85	0,75	0,80	0,85	0,85
P6	0,25	0,35	0,80	0,70	0,95
P7	0,45	0,25	0,90	0,90	0,85
P8	0,20	0,30	0,80	0,75	0,90
P9	0,35	0,55	1,00	0,95	1,00
Moyenne	0.47	0.44	0.87	0.83	0.92

*Tableau **5**. Apport de notre procédure de sélection de patches (résultats de ré-identification avec un ensemble de 10 patches discriminants pour chaque individu).*

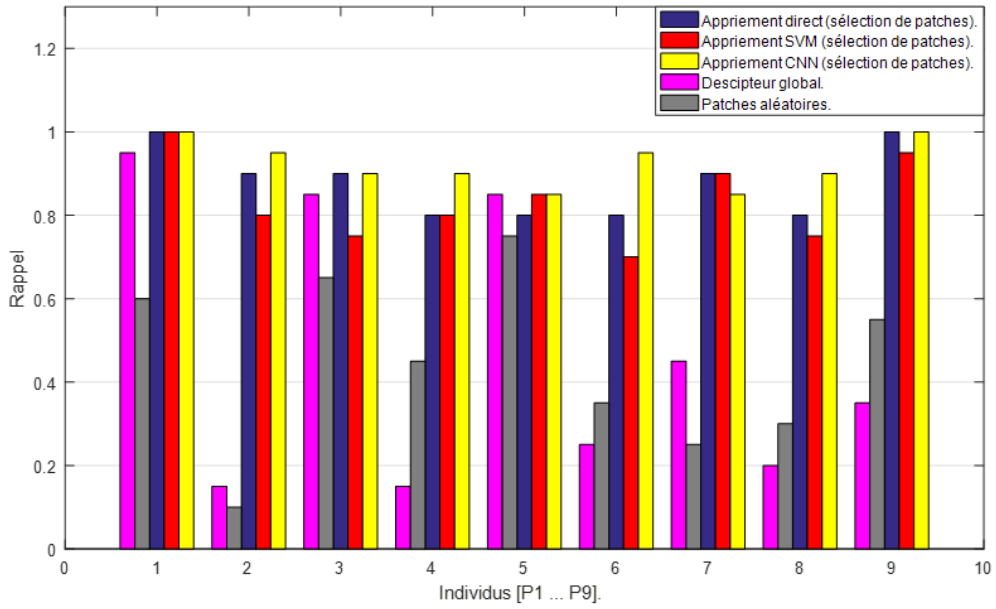


Figure 83. Apport de notre procédure de sélection de patches (rappel de la ré-identification avec un ensemble de 10 patches discriminants pour chaque individu).

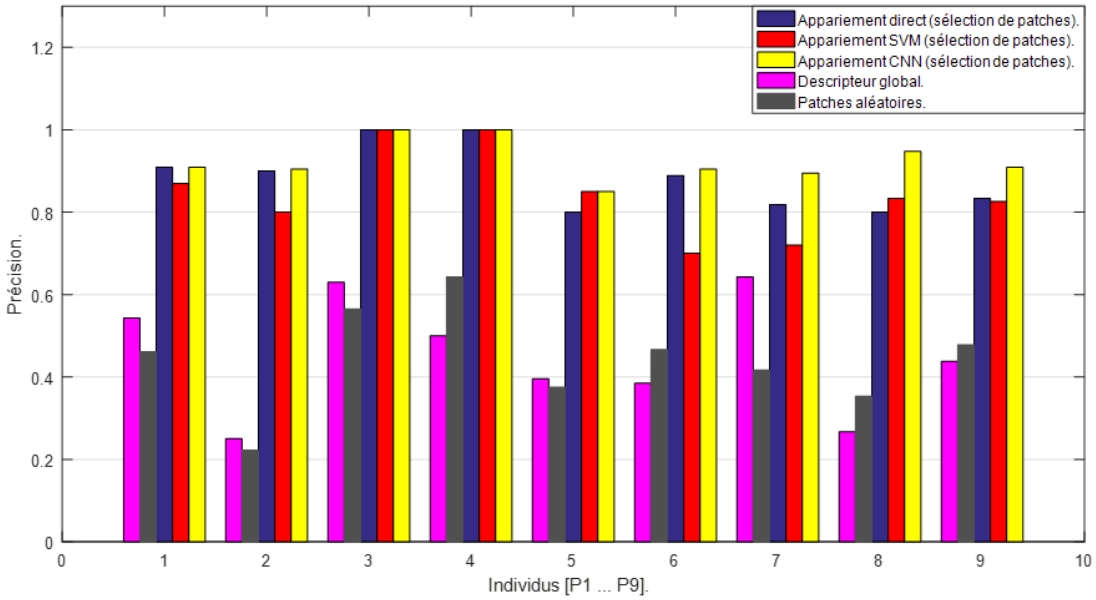


Figure 84. Apport de notre procédure de sélection de patches (précision de la ré-identification avec un ensemble de 10 patches discriminants pour chaque individu).

Ces résultats montrent une grande amélioration avec l'intégration de notre méthode de sélection de patches. Cette amélioration est nettement visible dans la ré-identification ainsi que dans le rappel et la précision. Cela confirme l'efficacité et l'apport de notre méthode de ré-identification.

Maintenant et après avoir confirmé l'apport de notre procédure de sélection de patches, nous allons comparer notre méthode avec d'autres méthodes de ré-identification proposées précédemment. Nous avons commencé par la comparaison de notre méthode avec des méthodes proposées dans le même contexte par Aziz et al [104] et [6]. Les résultats de cette comparaison sont montrés dans les diagrammes de rappel et précision des figures (*Figure 85* et *Figure 86*).

Ces résultats montrent que notre méthode donne de meilleurs résultats avec les différents appariements. Cette amélioration est plus visible dans l'utilisation d'apprentissage profond. Notre algorithme améliore les taux de rappel et de précision pour la majeure partie des individus et améliore le résultat globale de ré-identification qui sont de 78% et 80% pour Aziz et al [104] et [6] avec descripteur de couleur et points d'intérêts respectivement, tandis que notre méthode donne une moyenne de ré-identification entre 83% et 92% pour les différents appariements.

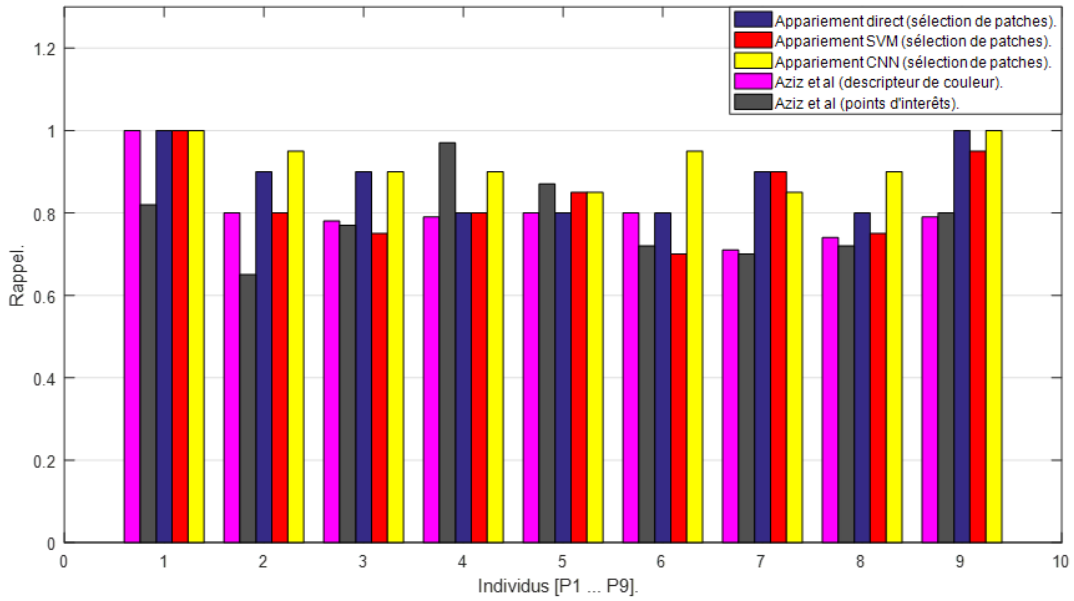


Figure 85. Comparaison avec les méthodes Aziz et al [104] et [6] (rappel de la ré-identification avec un ensemble de 10 patches discriminants pour chaque individu).

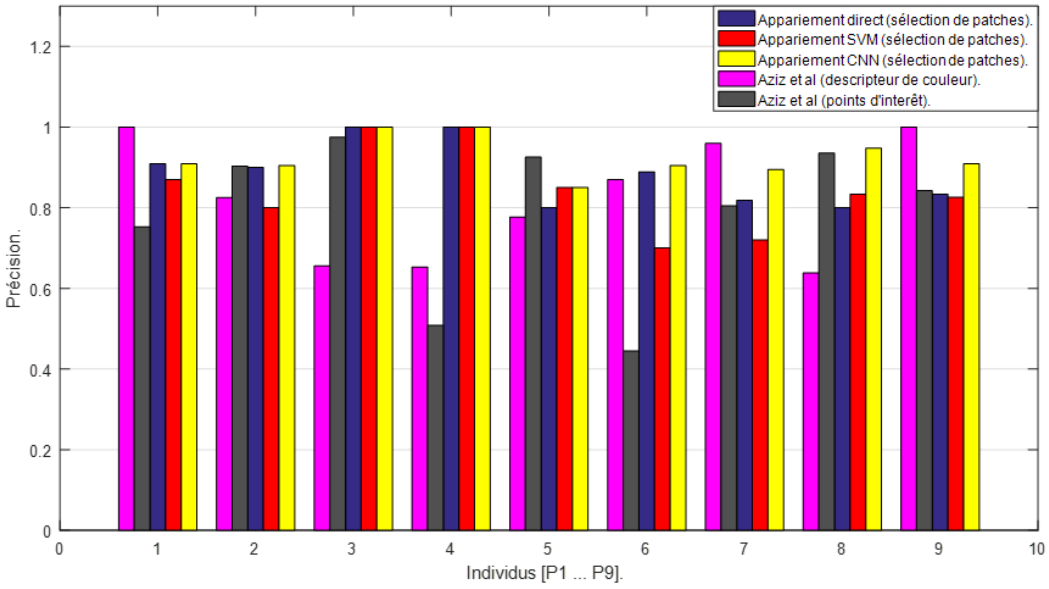


Figure 86. Comparaison avec les méthodes Aziz et al [104] et [6] (précision de la ré-identification avec un ensemble de 10 patches discriminants pour chaque individu).

La dernière série de tests que nous avons effectués dans cette partie est réalisée pour comparer notre méthode avec des méthodes existantes dans le contexte de la ré-identification des personnes sur une large échelle (ré-identification des individus parmi des centaines d'autres). Pour ce faire, nous avons utilisé la base VIPeR avec 316 individus pour l'apprentissage et 316 individus pour le test. Nous rappelons que nous avons effectué des changements dans notre méthode de sélection de patches, vu les restrictions de la base qui donne une seule image brute pour chaque individu dans chacune des caméras. Les résultats de cette comparaison sont montrés dans le Tableau 6.

Classement	1	5	10	20	50
Sélection de patches	**25,7**	**49,60**	**73,05**	**86,43**	**96,2**
PKFM [121]	36,8	-	83,6	91,7	97,8
SWF [122]	32,97	-	75,63	86,87	96,17
QALF [123]	30,17	-	62,44	73,81	-
SalMatch [124]	30.16	-	65.54	79.15	91.49
LADF [125]	29,30	-	78,80	92,20	97,40
SDALF [39]	19,87	38,89	49,37	65,73	-
ELF [119]	12,00	31,00	41,00	58,00	-
LMNN [68]	6,23	19,65	32,63	52,25	-

*Tableau **6**. Comparaison avec les méthodes existantes de ré-identification à large échelle avec la base VIPeR sur 316 individus.*

Les résultats de ce tableau montrent que notre méthode donne des résultats très satisfaisant dans ce cadre. Notre méthode n'atteint pas les mêmes résultats que ceux obtenus par la PKFM ou la SWF qui sont nettement meilleurs. Mais, contrairement à ces méthodes qui sont faites spécialement pour ce contexte, notre méthode utilise des descripteurs simple et rapide à calculés et nos tests ont été réalisés avec un petit handicap dans la sélection des patches discriminants causé par les restrictions dans la base de données utilisée.

3.5.2.2. Discussion

Les différentes séries de tests que nous avons effectués, nous ont permis de valider plusieurs aspects de notre approche de ré-identification. Ces tests nous ont aussi permis

de vérifier la performance de notre algorithme dans deux cas différents. Le premier dans le cadre de suivi multi-caméras qui nécessite une association de personnes parmi un nombre limité d'individus et aillant plusieurs détections pour chaque individu. Le deuxième dans le cadre d'une ré-identification à large échelle qui nécessite une association entre des centaines d'individus.

Les résultats que nous avons obtenus, nous ont permis premièrement de valider notre approche. Ceci est montré par les matrices de confusion dont lesquelles nous étions capable d'avoir des taux de ré-identification supérieur à 80% pour l'ensemble des individus avec un taux moyenne qui a varié entre 83% et 92% indépendamment de la procédure d'appariement adoptée. Ce taux a même atteint 100% pour certains individus. Dans le même sens, les diagrammes de rappel-précision, nous montrent que notre algorithme est capable d'atteindre des rappels très élevés pour l'ensemble des individus tout en gardant une bonne précision.

La troisième série de tests nous a permis d'étudier la stabilité et la robustesse de notre algorithme. La stabilité et robustesse sont considérées comme deux aspects très importants, car ils permettent d'avoir la certitude de ne pas avoir de résultats trop variés d'un test à un autre. Dans notre cas, les trois approches d'appariement proposées ont montré leur stabilité. Cette stabilité est visible sur les courbes CMC qui convergent rapidement vers 1. La convergence des courbes CMC indique que les erreurs de ré-identification sont apparues entre les individus qui ont des apparences plutôt similaires.

Les résultats des tests nous ont aussi montré que l'appariement le plus performant est celui obtenu avec CNN avec une moyenne de 92% tandis que le moins performant reste le SVM avec une moyenne 83%. Cette supériorité de l'apprentissage profond est due, en partie, aux descripteurs utilisés (une simple concaténation d'histogramme pour le SVM contre un descripteur de haute dimension pour l'apprentissage profond). De la même manière, la supériorité de l'apprentissage profond par rapport à l'appariement direct peut être expliquée par la classification binaire adoptée dans le premier contre la classification multi-class adoptée dans le deuxième.

La quatrième série de tests nous a permis de montrer l'apport de notre procédure de sélection de patches dans les résultats de la ré-identification. Cet apport est nettement visible dans les résultats de la ré-identification qui se sont améliorés dans certain cas de plus de 50% par rapport à la sélection de patches aléatoires ou le cas d'un descripteur global. Cette amélioration est aussi visible dans le taux de rappel et précision. Elle confirme l'intérêt de notre procédure de sélection de régions saillantes.

Ces résultats nous ont montré également l'amélioration qu'apporte notre algorithme aux résultats de ré-identification dans le cadre de suivi multi-caméras et dans le cadre d'une ré-identification à large échelle.

3.5.3. Gestion des occultations dans le champ de vision d'une seule caméra

Dans cette partie, nous avons utilisé notre algorithme pour le traitement de l'un des problèmes les plus persistants dans le cadre de suivi mono-caméra des personnes qui est la gestion des occultations. Pour ce faire, nous avons utilisé notre algorithme pour associer les différentes détections des individus dans le champ de vision d'une seule caméra. Nous avons utilisé la même base que précédemment, mais cette fois les individus détectés dans le champ de vision de la caméra 1 sont utilisés comme candidats et requêtes. Nous avons utilisé le même protocole de test utilisé précédemment. La performance de notre système est évaluée par le calcul du métrique précision-rappel comme indiqué par les équations (29) et (30) ainsi que les pourcentages de précision et de rappel qui sont calculés comme indiqué par les équations (31) et (32).

Nous avons réalisé les mêmes séries de tests avec les mêmes protocoles utilisés pour la validation de notre approche dans le cadre d'un réseau de caméras. Les résultats des différents tests sont illustrés dans la section 3.5.3.1.

3.5.3.1. Résultats

3.5.3.1.1. Différentes méthodes d'appariement

Dans la première série de tests, nous avons évalué la performance de notre algorithme dans le cadre de la gestion des occultations dans le suivi mono-caméra. Nous avons construit les matrices de confusion pour les différentes méthodes d'appariement décrites précédemment. Les figures (*Figure 87*, *Figure 88* et *Figure 89*) montrent les matrices de confusion obtenues avec 10 patches pour chaque individu.

Id	P1	P2	P3	P4	P5	P6	P7	P8	P9
P1	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
P2	0,00	0,95	0,00	0,00	0,00	0,05	0,00	0,00	0,00
P3	0,00	0,00	0,90	0,00	0,00	0,00	0,10	0,00	0,00
P4	0,00	0,00	0,00	0,80	0,10	0,00	0,10	0,00	0,00
P5	0,05	0,10	0,00	0,00	0,85	0,00	0,00	0,00	0,00
P6	0,00	0,00	0,00	0,00	0,00	0,80	0,00	0,20	0,00
P7	0,05	0,00	0,00	0,00	0,00	0,00	0,95	0,00	0,00
P8	0,00	0,00	0,00	0,00	0,05	0,00	0,00	0,90	0,05
P9	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00

Figure 87. Matrice de confusion pour un appariement direct avec la distance EMD et un ensemble de 10 patches discriminants pour chaque individu.

Id	P1	P2	P3	P4	P5	P6	P7	P8	P9
P1	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
P2	0,00	0,75	0,00	0,00	0,05	0,20	0,00	0,00	0,00
P3	0,00	0,00	0,85	0,00	0,10	0,00	0,05	0,00	0,00
P4	0,00	0,00	0,00	0,80	0,10	0,00	0,10	0,00	0,00
P5	0,05	0,10	0,00	0,00	0,85	0,00	0,00	0,00	0,00
P6	0,00	0,00	0,05	0,00	0,00	0,80	0,00	0,15	0,00
P7	0,05	0,00	0,00	0,00	0,05	0,00	0,90	0,00	0,00
P8	0,00	0,05	0,00	0,05	0,00	0,05	0,00	0,65	0,20
P9	0,00	0,00	0,00	0,00	0,00	0,05	0,00	0,00	0,95

Figure 88. Matrice de confusion pour un appariement avec apprentissage d'un SVM et un ensemble de 10 patches discriminants pour chaque individu.

Id	P1	P2	P3	P4	P5	P6	P7	P8	P9
P1	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
P2	0,00	0,95	0,00	0,00	0,00	0,05	0,00	0,00	0,00
P3	0,00	0,00	0,90	0,00	0,00	0,05	0,05	0,00	0,00
P4	0,00	0,00	0,00	0,90	0,05	0,00	0,05	0,00	0,00
P5	0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00	0,00
P6	0,00	0,00	0,00	0,00	0,00	0,95	0,00	0,05	0,00
P7	0,05	0,05	0,00	0,00	0,10	0,00	0,80	0,00	0,00
P8	0,00	0,00	0,00	0,00	0,05	0,00	0,00	0,95	0,00
P9	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00

Figure 89. Matrice de confusion pour un appariement par apprentissage profond et un ensemble de 10 patches discriminants pour chaque individu.

Ces matrices de confusion montrent que notre algorithme est capable d'obtenir des taux très élevés de ré-identification qui dépassent les 90% pour la majeure partie des individus. Ces taux peuvent atteindre les 100% pour certains individus. Nous avons obtenu des taux élevés pour les différentes méthodes d'appariements.

3.5.3.1.2. Comparaison des différents appariements

Dans cette section, nous avons comparé les différentes méthodes d'appariement. Pour ce faire, nous avons utilisé le Tableau 7 ainsi que les diagrammes de rappel et de précision montrés dans les figures (*Figure 90* et *Figure 91*).

méthode / Id	Résultats de la re-identification		
	Appariement direct	Avec SVM	Avec CNN
P1	1,00	1,00	1,00
P2	0,95	0,75	0,95
P3	0,90	0,85	0,90
P4	0,80	0,80	0,90
P5	0,85	0,85	1,00
P6	0,80	0,80	0,95
P7	0,95	0,90	0,80
P8	0,90	0,65	0,95
P9	1,00	0,95	1,00
Moyenne	0.90	0.84	0.94

*Tableau **7**. Comparaison entre différents appariements avec un ensemble de 10 patches discriminants pour chaque individu.*

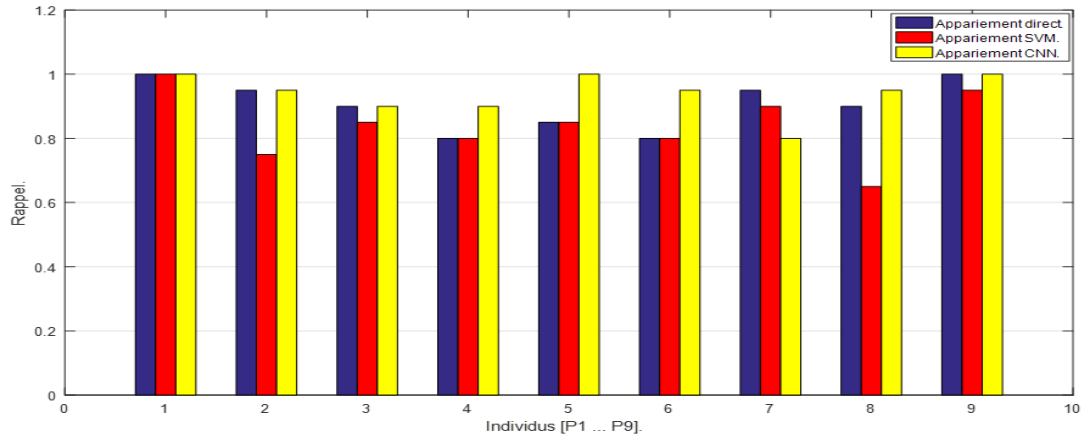


Figure 90. Comparaison entre les rappels des différents appariements avec un ensemble de 10 patches discriminants pour chaque individu.

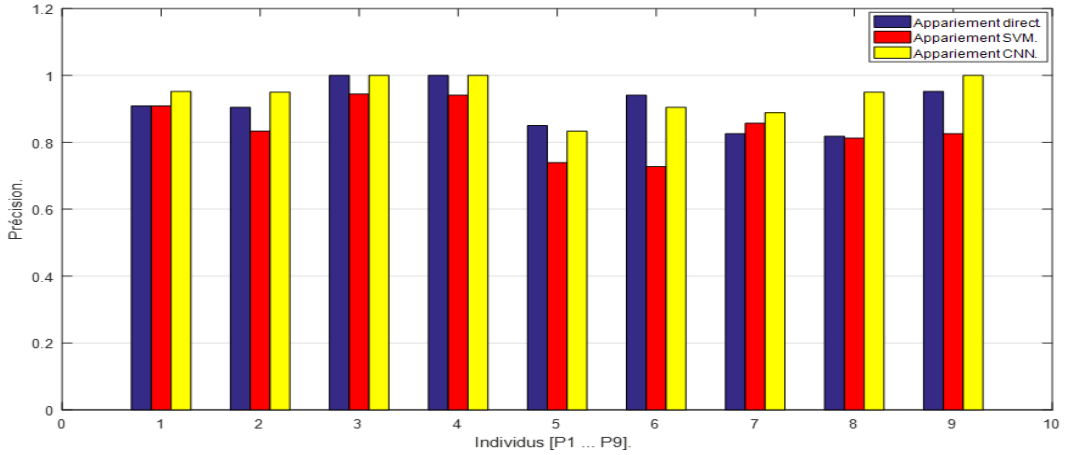


Figure 91. Comparaison entre les précisions des différents appariements avec un ensemble de 10 patches discriminants pour chaque individu.

Ces résultats montrent que l'apprentissage profond donne les meilleurs taux de ré-identification dans l'ensemble, tandis que l'apprentissage d'un SVM donne les moins bons résultats. Mais dans l'ensemble, tous les résultats obtenus par les différents appariements restent acceptables. De la même manière, les diagrammes de rappel et de précision montrent le même ordre pour les différents appariements.

3.5.3.1.3. Comparaison avec les méthodes existantes

Dans cette partie, nous allons comparer nos résultats avec les résultats obtenus par les méthodes déjà existantes de ré-identification dans le cadre de la gestion des occultations dans le suivi mono-caméra. Pour ce faire, nous allons comparer notre méthode avec d'autres méthodes de ré-identification proposées précédemment dans ce cadre(Aziz et al [104] et [6]). Les résultats de cette comparaison sont montrés dans le Tableau 8 et les diagrammes de rappel-précision des figures (*Figure 92* et *Figure 93*).

méthode / Id	Résultats de la re-identification				
	Segmentation couleur	Segmentation points d'intérêt	Appariement direct	Avec SVM	Avec CNN
P1	1,00	0,85	1,00	1,00	1,00
P2	0,87	0,92	0,95	0,75	0,95
P3	0,86	0,90	0,90	0,85	0,90
P4	0,90	0,95	0,80	0,80	0,90
P5	1,00	0,85	0,85	0,85	1,00
P6	0,81	0,82	0,80	0,80	0,95
P7	0,93	0,8	0,95	0,90	0,80
P8	0,88	0,85	0,90	0,65	0,95
P9	0,86	0,85	1,00	0,95	1,00
Moyenne	0.90	0.87	0.90	0.84	0.94

*Tableau **8**. Comparaison avec les méthodes Aziz et al [104] et [6] (résultats de la ré-identification).*

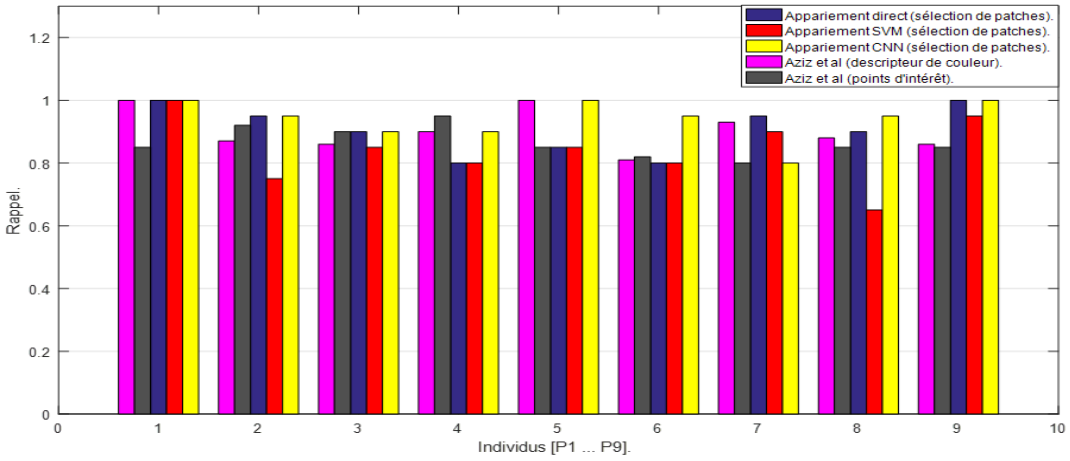


Figure 92. Comparaison avec les méthodes Aziz et al [104] et [6] (rappel de la ré-identification avec un ensemble de 10 patches discriminants pour chaque individu).

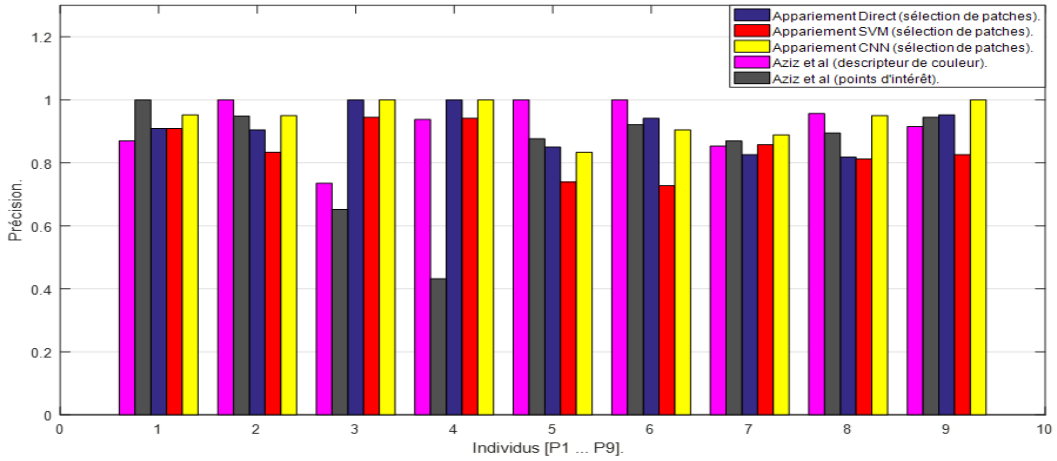


Figure 93. Comparaison avec les méthodes Aziz et al [104] et [6] (précision de la ré-identification avec un ensemble de 10 patches discriminants pour chaque individu).

Ces résultats montrent que notre méthode donne de meilleurs résultats avec l'appariement par apprentissage profond, des résultats plutôt similaires aux méthodes précédentes pour l'appariement direct, et des résultats moins bons pour l'appariement avec SVM. Notre algorithme améliore les taux de rappel et de précision pour la majeure partie des individus avec l'apprentissage profond et améliore le résultat global de ré-identification de 87% et 90% pour Aziz et al [104] et [6] avec descripteur de couleur et points d'intérêts respectivement, tandis que notre méthode atteint une moyenne de ré-identification de 94% pour l'appariement par apprentissage profond.

3.5.3.2. Discussion

Les résultats que nous avons obtenus, nous ont permis premièrement de valider notre approche dans le cadre de gestion des occultations pour le suivi mono-caméra. Ceci est montré par les matrices de confusion avec des taux de ré-identification supérieur à 90% pour la majeure partie des individus, et avec un taux moyen qui varie entre 84% et 94% indépendamment de la procédure d'appariement adoptée. Ce taux a même atteint 100% pour certains individus.

Les résultats des tests nous ont aussi montré que l'appariement le plus performant est celui obtenu par CNN avec une moyenne de 94%, tandis que le moins performant reste celui obtenu par SVM avec une moyenne 84%, celui de l'appariement direct étant de 90%. Ces tests nous ont permis aussi de vérifier que notre méthode est capable d'améliorer les résultats obtenus par l'utilisation des méthodes proposé par Aziz et al

[104] et [6] qui sont basées sur la segmentation et l'utilisation de descripteurs de couleurs et de descripteurs par points d'intérêts. Notre méthode est capable d'atteindre des performances plus élevées avec l'utilisation d'apprentissage profond et des performances plutôt similaire en utilisant un appariement direct.

3.6. Conclusion

Dans ce chapitre, nous avons proposé une méthode originale de ré-identification dans le cadre de l'association des individus suivis dans un réseau de caméras. Cette méthode est basée sur l'association des régions saillantes représentants les différents individus. Nous avons proposé un algorithme en deux étapes, la première utilisée pour la sélection des régions saillantes, et la seconde utilisée pour l'association de ces régions et la confirmation de la ré-identification.

La première phase consiste dans la sélection des régions saillantes caractéristiques des différents individus. Pour réaliser cette tâche, nous avons proposé un nouveau framework qui est utilisé pour la représentation des individus en se basant sur leur description spatiale ainsi que leur description d'apparence. Dans ce framework, chaque individu est représenté par un ensemble de détections, chacune de ces détections est représentée par un ensemble de patches, et chacun de ces patches est représenté par sa description d'apparence et ces informations spatiales. Ensuite, ce framework est utilisé pour la sélection des régions saillantes (patches discriminants) caractérisant chaque individu. Ces régions saillantes seront utilisées dans la deuxième phase pour l'association de ces individus aux nouvelles détections.

La deuxième phase de notre algorithme est la phase d'appariement qui permet d'associer les régions saillantes sélectionnées précédemment et confirmer la ré-identification. Pour la réalisation de cette phase, nous avons proposé et implémenté trois méthodes, la première basée sur un appariement direct, la seconde sur un appariement par SVM, et une troisième sur un appariement par apprentissage profond.

Nous avons effectué plusieurs tests pour valider notre algorithme dans le cadre de l'association des individus pour un système de suivi multi-caméras, et dans celui de la gestion des occultations dans un système de suivi mono-caméra. Dans les deux cas, notre algorithme de sélection de régions saillantes donne des résultats très satisfaisants.

Chapitre 4. Modélisation sémantique des trajectoires

4.1. Introduction

L'étude de comportement est l'un des problèmes majeurs de plusieurs recherches. L'un des principaux objectifs de ces études est la compréhension de comportement de clientèle dans un espace de vente pour l'amélioration de service et l'optimisation des ventes. En général, l'étude de comportement dans ce contexte passe par l'exploitation des données obtenues du système de vidéosurveillance. L'exploitation de la faramineuse quantité de données générées par les différents systèmes de surveillance nécessite le déploiement d'efforts considérables et l'implication de plusieurs opérateurs humains. Pour cela, plusieurs recherches ont été effectuées, durant la dernière décennie, pour l'automatisation de ces études. L'un des problèmes majeurs étudiés dans ce contexte est le suivi de personnes dans un réseau de caméras.

Le suivi de personnes dans un réseau de caméras est défini comme la capacité de reconstruire les trajectoires des différents individus observés dans l'ensemble des caméras couvrant l'espace surveillé. Une fois le suivi effectué, le système de suivi génère des trajectoires qui sont l'ensemble des positions occupées par un individu donné durant son passage dans l'espace surveillé. Ces positions, représentées par des données spatiales et temporelles, sont difficilement interprétable dans le cadre de la compréhension des comportements. Pour cela, une nouvelle problématique d'interprétation de ces trajectoires s'impose. Cette problématique consiste dans l'ajout de connaissance à ces trajectoires brutes pour faciliter leurs interprétations par les opérateurs humains. Une

façon naturelle pour ajouter ces connaissances consiste dans l'élaboration d'une ontologie pour la construction des trajectoires sémantiques.

Dans ce chapitre, nous proposons le quatrième et dernier module de notre système de suivi multi-caméras de personnes. Précédemment, nous avons proposé un système de suivi multi-caméras permettant de reconstruire les trajectoires des personnes suivies durant leurs passages dans l'espace surveillé. Une fois ces trajectoires obtenues, la prochaine étape consiste dans l'ajout de connaissance à ces trajectoires pour qu'elles deviennent interprétables. Pour ce faire, nous avons proposé un module pour la construction des trajectoires sémantiques. Ce module est basé sur la construction de trois ontologies : une ontologie de modélisation de l'environnement pour modéliser les parties statiques de l'espace surveillé et une ontologie de modélisation des trajectoires qui est utilisée dans la construction des trajectoires sémantiques, ainsi qu'une ontologie d'observation dans l'environnement qui permet de relier les trajectoires à l'environnement.

4.2. Trajectoires sémantiques

Dans ce chapitre, notre objectif est l'ajout de connaissances aux trajectoires afin de faciliter leurs interprétations. Une trajectoire de base est définie comme une succession de positions spatio-temporelles. Cette succession de positions représente les différents emplacements occupés par l'individu suivi durant son passage dans le champ de vision de la caméra.

Pour construire une trajectoire sémantique à partir de ces trajectoires de base, la tâche principale consiste dans l'ajout d'une dimension sémantique permettant de connecter cette trajectoire à un réseau de connaissances plus large. Pour ce faire, la succession de positions spatio-temporelles est converti à une succession d'épisodes regroupant celles qui sont sémantiquement liées (par exemple les arrêts et les mouvements). Chaque épisode est disjoint l'un de l'autre et deux épisodes du même type ne peuvent pas se succéder. Ce concept de base, segmentant les trajectoires en des épisodes, est alors enrichi avec des informations et annotations sémantiques sur l'environnement. L'ensemble des éléments utilisés dans l'annotation sémantique dépond de l'application. Une illustration de ce concept est montrée dans la Figure **94**.

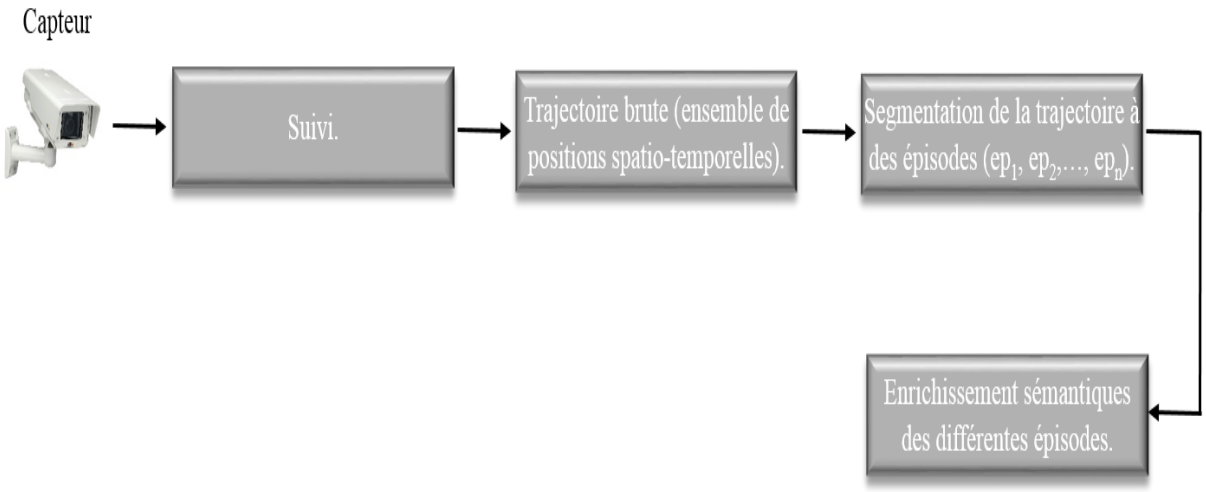


Figure 94. Illustration de la construction d'une trajectoire sémantique.

4.3. État de l'art

Dans les chapitres précédents, nous avons présenté les différents modules composant notre système de suivi multi-caméras de personnes. Ces différents modules nous permettent de reconstruire les trajectoires des individus suivis durant leurs passages dans l'espace surveillé. Les trajectoires que nous récupérons sont représentées par une succession de positions spatio-temporelles qui sont difficilement interprétable pour l'étude de comportement. Pour cela, nous avons créé un autre module responsable de l'ajout d'une dimension sémantique à ces trajectoires pour pouvoir les connecter à des bases de connaissances plus larges. Pour ce faire, nous avons construit deux ontologies : une ontologie pour les trajectoires sémantiques et ontologie pour l'environnement pour pouvoir connecter les trajectoires à l'espace surveillé.

La plupart des travaux réalisés dans le cadre de l'étude des trajectoires sémantiques [126], [127], [128] s'inspirent du travail proposé par Spaccapietra et al [129]. Dans leurs travaux, ils ont donné différentes définitions du concept de trajectoire. La première définition qu'ils ont donnée prend en compte l'ensemble des positions occupées par un objet mobile durant un intervalle de temps pour accomplir un but. Donc, la trajectoire est considérée comme le parcours de l'objet mobile qui est pertinent pour l'accomplissement du but. Dans ce cas, la segmentation de la trajectoire est faite en fonction du but à atteindre. Dans les mêmes travaux, ils proposent une autre définition pour les trajectoires, cette fois-ci ils considèrent la trajectoire comme une succession d'épisodes. Pour ce faire, ils ont défini deux concepts permettant la structuration des trajectoires : Stop et Move. Donc, la trajectoire est conceptualisée par une succession

d'épisodes d'arrêts significatifs et de mouvements ; et chaque épisode (Stop ou Move) est disjoint l'un de l'autre et deux épisodes du même type ne peuvent pas se succéder. Cette nouvelle définition permet de segmenter les trajectoires à des épisodes d'arrêts significatifs auxquelles une attention particulière est donnée et des épisodes de mouvements. Les épisodes d'arrêts sont considérés comme motivés par l'accomplissement du but mentionné précédemment et les épisodes de mouvements sont considérés comme des étapes intermédiaires. Cette conceptualisation de trajectoire est utilisée par Santipantakis et al [127] dans le cadre de la conceptualisation des couloirs maritimes, elle est illustrée par la *Figure 95*.

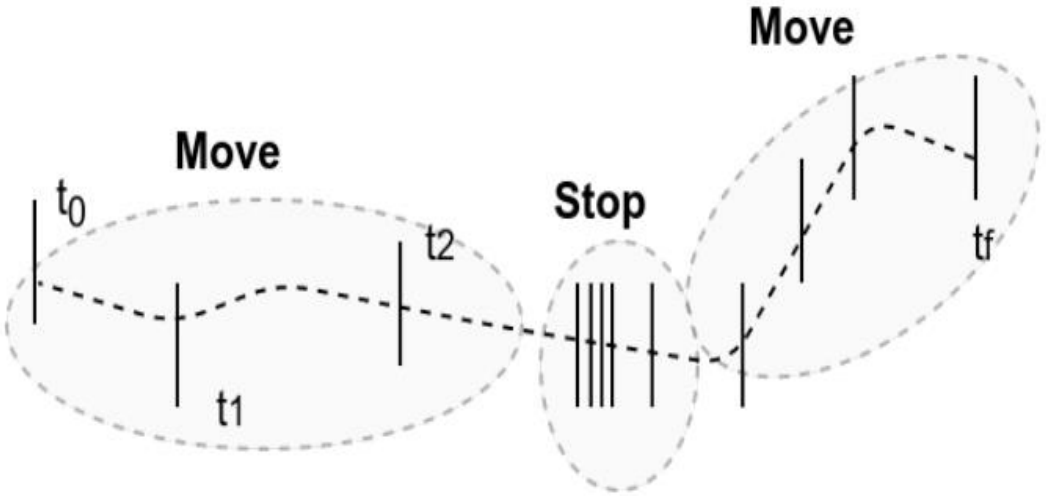


Figure 95. conceptualisation des trajectoires sémantiques (Spaccapietra et al [129]).

Cette conceptualisation de trajectoires en une succession d'épisodes peut être pertinente pour l'étude des trajectoires dans un espace vaste avec peu de points d'intérêt comme les couloirs maritimes dans [127], mais elle est beaucoup moins pertinente pour l'étude de petits espaces avec plusieurs points d'intérêt comme les surfaces de ventes. De ce fait et pour enrichir ces trajectoires, un ajout d'informations cinétique sur le mouvement est proposé par Yan et al [130] afin de décrire les différents épisodes. Ils ont ajouté les concepts : rapide et lent pour décrire le mouvement dans les différents épisodes. Dans le même contexte, Ying et al [131] ont proposé d'utiliser la forme des trajectoires sémantiques afin de comparer les trajectoires. Ils ont estimé que la distance géométrique n'est pas la meilleure façon pour calculer la similarité entre deux trajectoires.

Cette limitation dans le concept proposé précédemment, pour la conceptualisation des trajectoires, a donnée naissance à la trajectoire sémantique qui consiste dans l'ajout

d'informations et l'annotation des épisodes. Dans ce cadre, Spaccapietra et al [129] ont proposé d'annoter les entités B.E.S (Begin, End, Stop) avec des informations liées à la zone traversée et aux activités effectuées. Cet enrichissement sémantique est spécifique au problème étudié. De la même manière, Yan et al [128] ont ajouté une dimension sémantique à leurs trajectoires avec une annotation à l'aide d'un ensemble de lieux sémantique. Ils ont défini un ensemble d'éléments qui représentent les objets géographiques pertinents. Cet ensemble d'éléments est réparti en trois sous-ensembles en fonction de type de leurs extensions spatiales (points, ligne ou région). Cette étude est réalisée dans le cadre d'une application pour les déplacements extérieurs capturés par les traces GPS. Le concept de trajectoire sémantique proposé ici est une extension du concept de trajectoire où chaque trace spatio-temporelle est étendue par l'ajout d'un ensemble d'annotations contenant des références à des lieux sémantiques avec lesquels l'objet a eu une interaction. Une illustration de cette ontologie est montrée dans la *Figure 96*.

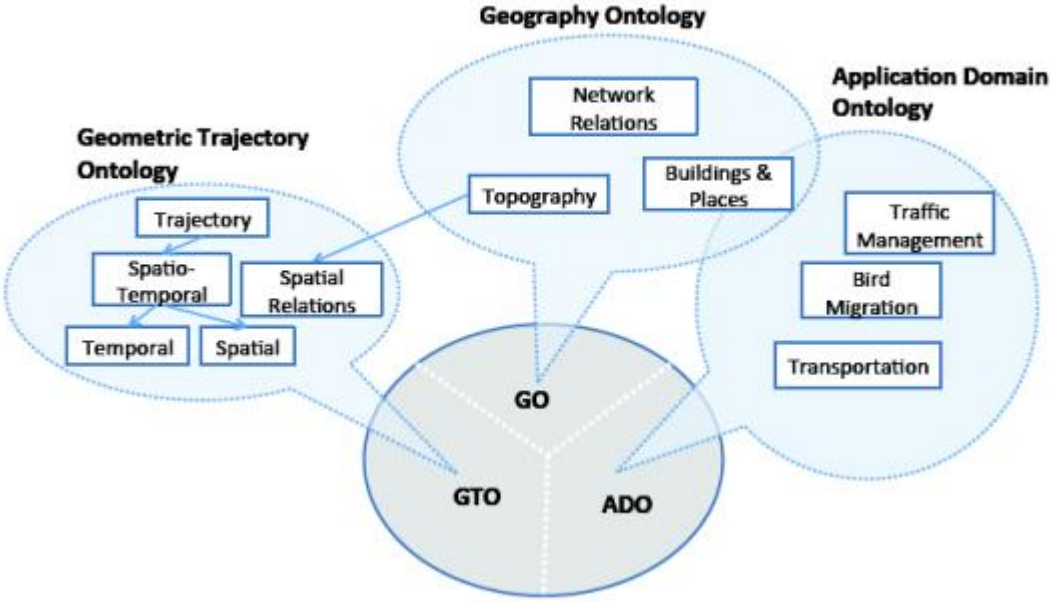


Figure 96. Enrichissement sémantique des trajectoires sémantiques (Yan et al [128]).

4.4. Description de la méthodologie proposée

L'objectif de ce module est la conceptualisation des trajectoires obtenues précédemment pour faciliter leur interprétation. Les modules que nous avons présentés dans les chapitres précédents permettent d'obtenir les trajectoires des personnes suivies durant leurs passages dans la scène surveillée. Ces trajectoires sont représentées par un ensemble de détections ou de localisations de ces personnes. Ces localisations spatio-

temporelles peuvent être représentées par les coordonnées (x, y, z, t). Ces trajectoires sont difficilement interprétables pour l'analyse du comportement. Pour cela, l'objectif de ce module est l'ajout de connaissances sémantique à ces trajectoires.

Pour construire les trajectoires sémantiques à partir des trajectoires que nous avions, nous nous sommes inspirés du travail réalisé par Vandecasteele [132]. Dans ses travaux, l'auteur a développé un système basé sur le développement d'une ontologie qui permet l'automatisation des raisonnements spatiaux à partir des connaissances d'experts dans la sécurité maritime. Il a créé un système pour associé des alertes en fonction des positions des navires ; ce qui lui a permis de créer des profils de comportement et identifier des scénarios en se basant sur ces positions. Notre travail rentre dans le même cadre ; mais dans le cadre de la création de profils de comportement pour des clients dans surface de vente. Pour la conceptualisation de nos trajectoires, nous avons choisi de ne prendre en compte que les interactions entre les individus suivis et les objets statiques dans la surface de ventes et d'ignorer les interactions entre les individus. C'est un choix qui peut être discuté car la dimension sociale n'est pas sans importance dans la perception de l'atmosphère d'un lieu de vente et donc dans l'analyse du comportement des personnes s'y trouvant [133].

En s'inspirant des travaux cités précédemment, nous avons choisi de construire des trajectoires sémantiques à partir de nos trajectoires avec une problématique sur plusieurs phases : l'extraction des informations sur les individus suivis, la modélisation de l'environnement (dans notre cas une surface de vente) et l'établissement des relations entre les individus et les éléments de l'environnement. Cette problématique illustrée dans la *Figure 97*.

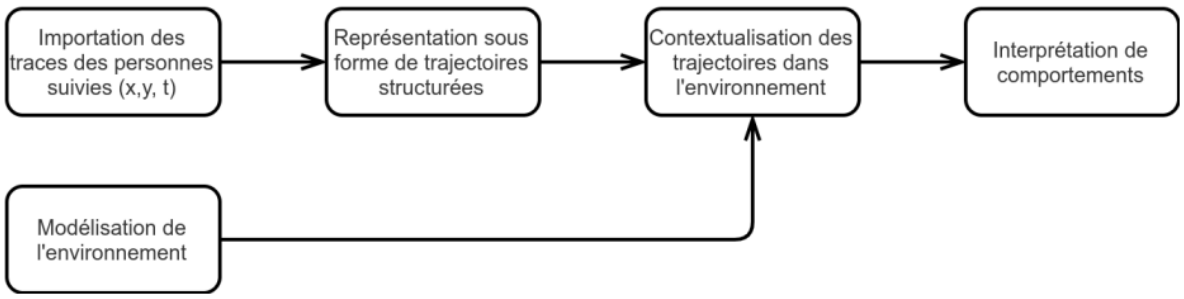


Figure 97. Problématique de modélisation ontologique des trajectoires de clients dans une surface de ventes.

Donc, notre problématique de base qui consiste dans la transformation des trajectoires spatio-temporelles à des trajectoires sémantiques plus facile à interpréter peut être imbriquée en deux sous-problématiques : la représentation par des connaissances de l'évolution des positions des individus suivis et la mise en relation de cette évolution avec l'environnement par la modélisation de l'environnement. Pour résoudre ces sous-problématiques, nous avons proposé une ontologie composée de plusieurs modules qui permet la séparation des connaissances modélisées en plusieurs sous-ontologies. La structure générale de l'ontologie proposée est montrée par la *Figure 98*.

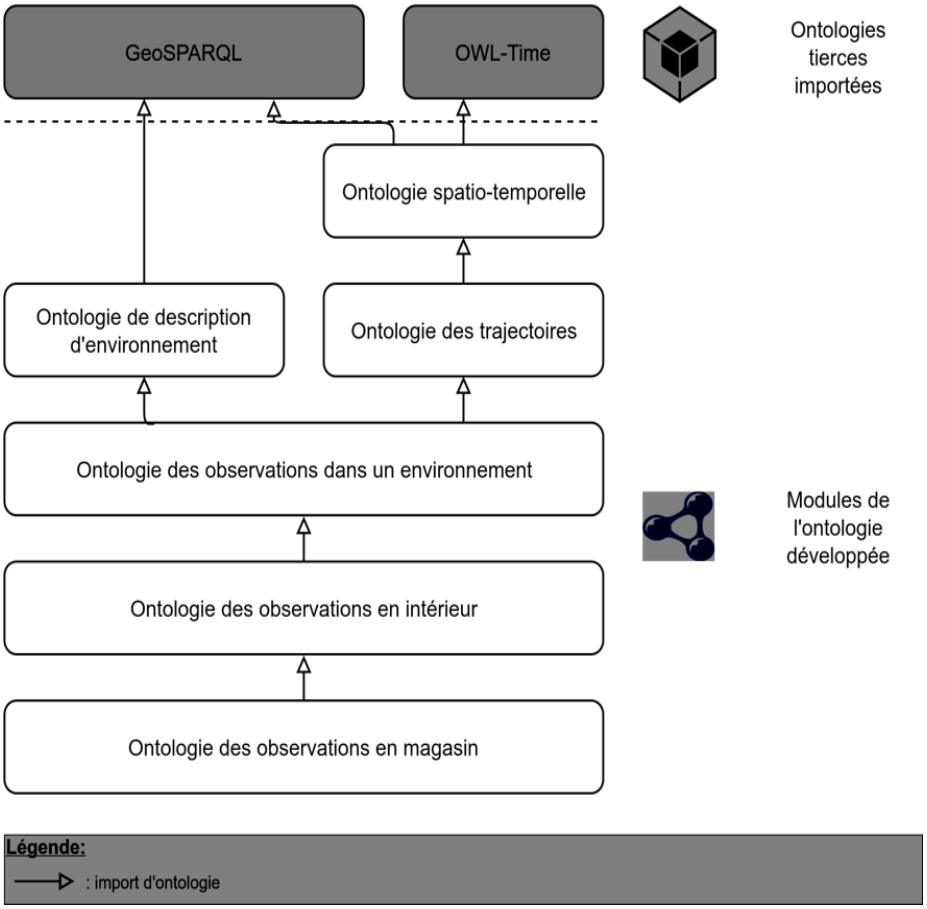


Figure **98. Problématique de modélisation ontologique des trajectoires de clients dans une surface de ventes.**

Dans cette structure, nous trouvons en haut les ontologies importées et utilisées dans notre ontologie qui sont la GeoSPARQL pour géométrie et OWL-Time pour le temps ; ensuite viennent les ontologies utilisées pour la modélisation des trajectoires et de l'environnement. Ces ontologies contiennent les connaissances de base qui permet de modélisées l'environnement et les observations. A la fin, nous trouvons les ontologies de haut niveau qui permettent de représenter les observations dans l'environnement.

4.4.1. Modélisation de l'environnement

Comme nous l'avant mentionner précédemment, une partie importante de notre travail consiste dans la modélisation de l'environnement. Dans notre étude, nous considérant que l'environnement est une surface de vente. Pour cela, nous avons commencé par la création d'une base de connaissances contenant les éléments statiques dans notre environnement. Cette base de connaissances contient :

- Les instances des concepts représentants l'environnement de type surface de vente : bâtiments, étages, pièces, caméras, rayon, gondole, …

Dans ce travail, nous avons considéré l'environnement comme un espace peuplé par des entités statiques possédant une extension spatiale et une composante sémantique (un exemple d'un environnement à modélisé est montré par la *Figure 99*.

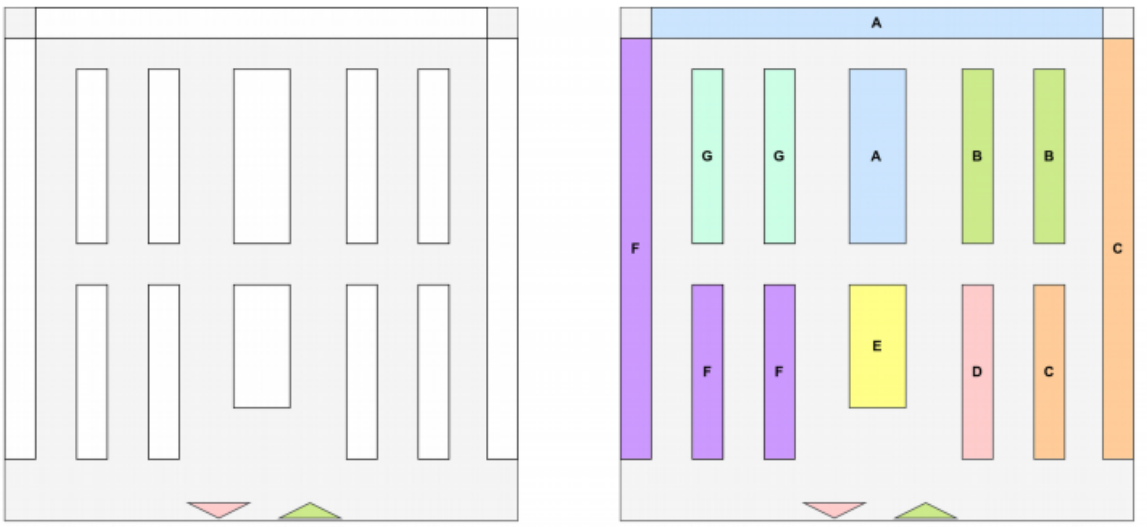


Figure **99. Problématique de modélisation ontologique des trajectoires de clients dans une surface de ventes.**

Les extensions spatiales des éléments de l'environnement sont obtenues par la spécialisation du concept ***Feature*** de ***GeoSPARQL***. En partant du concept ***Feature*** de ***GeoSPARQL***, nous avons créé le concept ***EnvironmentObject*** qui représente tous les éléments statiques de l'environnement. Ce concept représentant les éléments de l'environnement peut être spécialisé à son tour à des objets physiques et des objets non-physiques comme dans l'ontologie fondamentale DOLCE proposée par Yang et al [134] avec les concepts de ***PhysicalObject*** et ***NonPhysicalObject*** (lui-même spécialisé par ***MentalObject*** et ***SocialObject***). Donc, nous avons créé deux nouveaux concepts pour décrire les entités de l'environnement qui sont : ***EnvironmentMentalObject*** et

EnvironmentPhysicalObject ; ensuite les éléments spécifiques de l'environnement qui ont une existence matérielle sont décrits par des concepts qui sont subsumés par le concept ***EnvironmentPhysicalObject***, tandis que les concepts décrivant les éléments qui n'ont pas d'existence matérielle sont subsumés par le concept ***EnvironmentMentalObject***. Ces entités sont néanmoins significatives pour la compréhension de l'environnement. Par exemple, savoir qu'une zone d'un supermarché est un rayon particulier permet de mieux comprendre les trajectoires qui y passent, même si ce que nous considérons comme le rayon n'est pas une entité physique délimitée. Ensuite, nous avons créé deux autres concepts : le concept ***Space*** et le concept ***Passage***. Le premier représentant tous les espaces délimités dans l'environnement et le deuxième représente les entités permettant le passage d'un espace délimité à un autre. Le concept ***Passage*** est subdivisé en deux concepts ***Access*** pour les passages réels et ***Frontier*** pour les passages virtuels entre deux espaces. En même temps, le concept ***Space*** est subdivisé en trois concepts ***OutdoorSpace*** pour les espaces extérieures et ***IndoorSpace*** pour les espaces intérieurs, ainsi qu'un concept de ***ClearArea*** représentant les espaces vides qui ne contiennent pas d'obstacles pour le mouvement. Cette notion peut être utilisée dans la validation des trajectoires. À la fin, le concept ***ClearArea*** est à son tour subdivisé en deux concepts ***OutdoorClearArea*** pour les espaces extérieurs vides et ***IndoorClearArea*** pour les espaces intérieurs vides. En même temps, le concept ***Access*** est subdivisé en ***UnidirectionalAccess*** pour les accès unilatéral et ***BidirectionalAccess*** pour les accès bilatéraux. Cette ontologie de description de l'environnement est illustrée par la *Figure 100*.

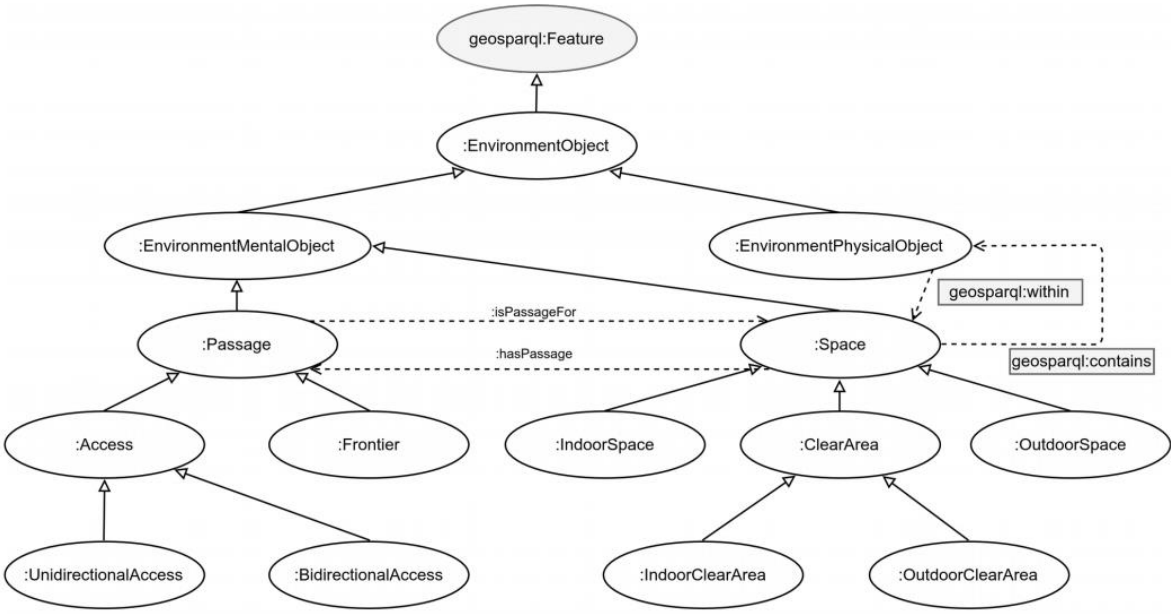


Figure **100. Ontologie de description de l'environnement.**

4.4.2. Modélisation des trajectoires

Le deuxième module que nous avons construit dans notre ontologie est destiné à l'enrichissement sémantique des trajectoires. Dans ce cadre, nous nous sommes inspirés du modèle que nous avons présenté dans l'état de l'art.

Pour cela nous avons commencé par la construction d'une ontologie spatio-temporelle. Cette ontologie est faite pour représenter les entités avec des extensions spatiales et des extensions temporelles. Pour ce faire, nous avons créé un concept ***SpatioTemporalEntity*** qui spécialise les concepts de ***TemporalEntity*** d'OWL-Time et ***Feature*** de GeoSPARQL. Donc, une instance de ce concept va avoir une dimension temporelle et une dimension spatiale. Les entités temporelles que nous avons utilisées dans la construction de notre ontologie ont pour géométrie : le ***point*** pour les positions des objets dans l'espace et la ***polyligne*** pour les trajectoires. ***SpatioTemporalEntity*** est donc subdivisé en les concepts *TimePoint* et *TimeLine*. Ces derniers subsument ***InstantPoint*** et ***IntervalPoint*** ainsi que ***InstantLine*** et ***IntervalLine***. Les concepts spatio-temporels que nous avons proposés respectent l'axiome contraignant suivant :

$$TimePoint \subseteq SpatioTemporalEntity \wedge \forall hasGeometry.Point$$

$$TimeLine \subseteq SpatioTemporalEntity \wedge \forall hasGeometry.Polyline$$

L'illustration de cette ontologie spatio-temporelle est montrée dans la *Figure 101*.

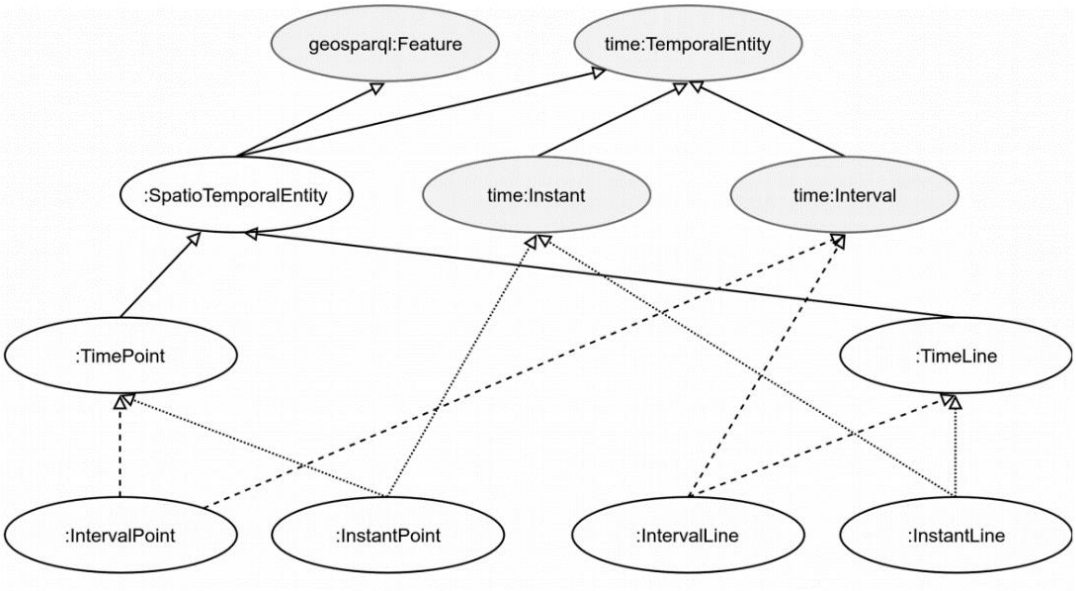


Figure **101. Ontologie spatio-temporelle proposée.**

L'ontologie spatio-temporelle que nous avons proposée précédemment est ensuite utilisée par l'ontologie des trajectoires afin de représenter les aspects spatio-temporels des différents concepts. Pour cette partie, nous nous sommes inspiré des travaux qui se fondent sur le modèle de trajectoires sémantiques de Spaccapietra et al [129]. Pour cela, nous avons créé une ontologie qui permet de représenter les trajectoires sur deux niveaux : le niveau des traces brutes et le niveau sémantique.

Tout d'abord, pour la représentation des trajectoires par des traces brutes, nous avons créé un concept de ***RawTrajectory*** qui est subsumé par ***IntervalLine*** pour modéliser les trajectoires. Une instance de ce concept est une entité qui a une extension géométrique correspondant à une ligne qui reliée toutes les positions parcourues par l'individu durant sa présence dans l'espace surveillé et une extension temporelle correspondant à l'intervalle allant de l'instant de la première observation de l'individu jusqu'à l'instant de sa sortie de l'espace surveillé ; ensuite, chaque observation d'une position de l'objet est représentée par une instance de ***ObjectPositionTrace*** et chaque trace spatio-temporelle peut être enrichie avec des données grâce à l'association d'une instance du concept ***ObjectPositionTrace*** à une instance du concept ***TraceData*** avec la propriété d'objet ***hasExtraData***. L'ontologie du domaine qui intègre notre ontologie peut alors définir ses propres concepts spécifiant *TraceData* en fonction de ses capacités d'observation. Chaque trace est reliée à la trace qui la suit et à celle qui la précède avec les propriétés ***hasNextTrace*** et ***hasPreviousTrace*** et chaque instance de ***RawTrajectory*** est liée aux traces qui la composent avec la propriété ***hasTrace*** (spécifiée par ***hasFirstTrace*** et ***hasLastTrace*** qui doivent avoir chacune un objet pour chaque instance de *RawTrajectory*). L'illustration de cette modélisation au niveau des traces brutes des trajectoires est montrée dans la *Figure 102*.

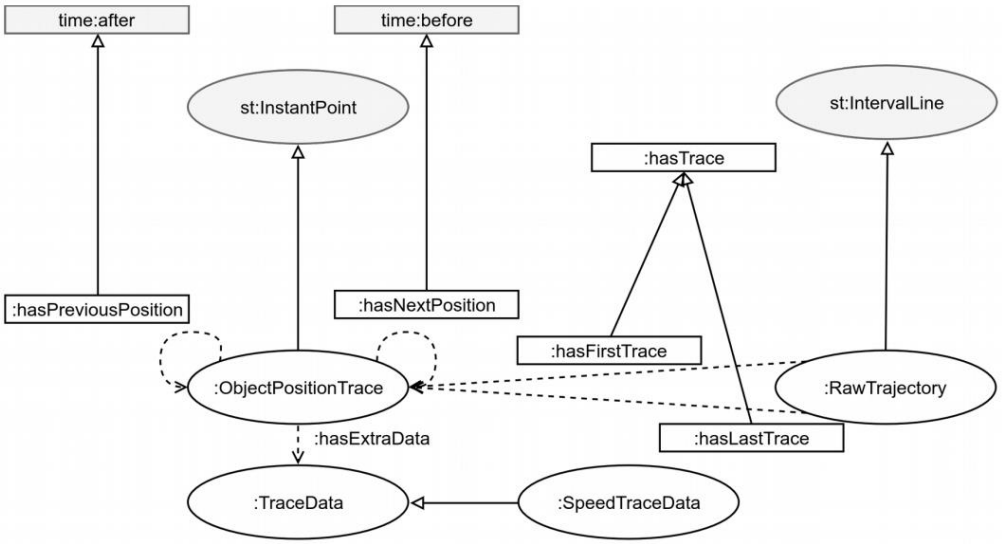


Figure 102. **Modélisation au niveau des traces brutes des trajectoires.**

Après avoir représenté les trajectoires avec un ensemble de traces brutes, la phase suivante consiste dans l'enrichissement de ces traces par une modélisation sémantique pour construire les trajectoires sémantiques. Pour ce faire, nous avons introduit le concept de ***SemanticTrajectory***. Le modèle de trajectoires sémantiques que nous proposons s'inspire du modèle que nous avons présenté dans l'état de l'art. Donc, une instance du concept ***SemanticTrajectory*** est composée d'épisodes ***TrajectoryEpisode***. Chaque épisode représente une séquence de positions qui a un sens (arrêt ou mouvement). Cette segmentation de trajectoires permet de structurer les traces de l'individu suivi (***MobileObject***) en épisodes. Nous n'avons pas repris les mêmes contraints sur les épisodes classiques de ***Stop*** et ***Move*** ; vu que nous considérons les épisodes de ***Move*** étant représentant une séquence de mouvement significative pouvant être suivie par un autre épisode de ***Move***. Donc, il est possible de segmenter les épisodes de mouvement en des phases qui ont sens ; Cette modélisation de la structure d'une trajectoire est motivée par la conviction que le déplacement d'un individu dans une zone de l'environnement peut être considéré comme marque d'intérêt dans certaines conditions (si sa vitesse est faible par exemple).

Cette ontologie ne représente que les connaissances relatives à la structure d'une trajectoire sémantique et à ses extensions spatiales et temporelles. Elle est prévue pour être intégrée dans une ontologie permettant de modéliser les observations annotées sémantiquement grâce à l'environnement. Pour ce faire, un point d'extension est prévu avec le concept de ***TrajectoryEvent***. Ce concept peut être associé à un épisode avec la propriété ***hasEvent*** et peut être spécialisé par une ontologie du domaine où les

trajectoires sont observées pour représenter des événements significatifs dans l'interprétation du comportement manifesté lors du déplacement de l'individu. L'illustration de cette modélisation au niveau sémantique des trajectoires est montrée dans la *Figure 103*.

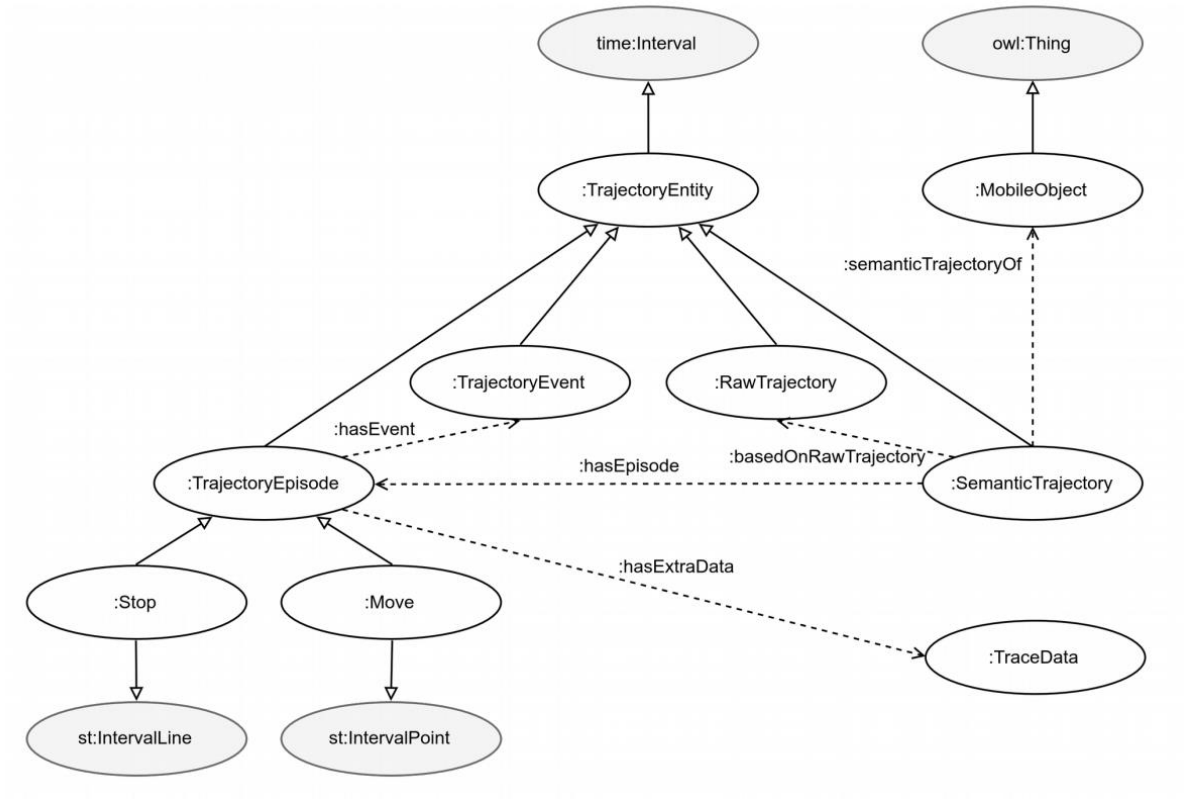


Figure 103. Modélisation au niveau sémantique des trajectoires.

4.4.3. Modélisation des observations dans l'environnement

Dans les deux sous-sections précédentes, nous avons créé deux ontologies une pour la modélisation de l'environnement et l'autre pour la modélisation des trajectoires. Dans cette partie, nous allons créer une ontologie qui permettra de faire le lien entre l'environnement et les trajectoires pour pouvoir faire quelques interprétations de bas-niveau des trajectoires. Pour ce faire, nous avons défini le concept ***EnvironmentRelationEvent*** qui spécialise ***TrajectoryEvent***. Ce concept permet l'enrichissement sémantique des trajectoires par leur association à des éléments de l'environnement. Cette relation est établie par la propriété d'objet ***relatesTo***.

Dans cet exemple d'ontologie, nous proposons trois concepts d'évènements : ***SpaceInclusionEvent***, ***ProximityInteractionEvent*** et ***PassageCrossingEvent*** (chacun défini par des restrictions sur l'objet de ***relatesTo*** devant être, respectivement,

de type ***Space***, ***EnvironmentPhysicalObject*** ou ***Passage***). L'illustration de cette modélisation des observations dans l'environnement est montrée dans la *Figure 104*.

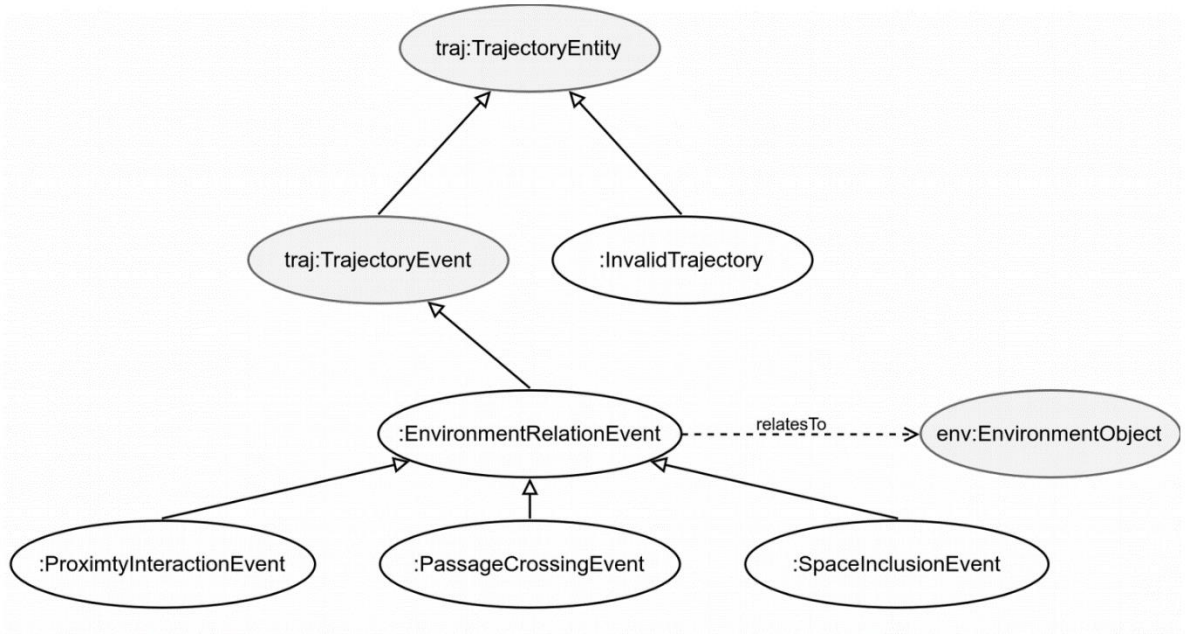


Figure 104. Modélisation des observations dans l'environnement.

4.5. Conclusion

Dans ce chapitre, nous avons proposé le quatrième et dernier module de notre système de suivi multi-caméras de personnes. Ce module est responsable de l'ajout de connaissances aux trajectoires brutes obtenues précédemment pour faciliter leur interprétation. Pour cela, nous avons proposé un module basé sur la construction de trajectoires sémantiques à l'aide d'ontologies.

Ce module de construction des trajectoires sémantiques est basé sur la création de trois ontologies : une ontologie de modélisation de l'environnement pour modéliser les parties statiques de l'espace surveillé et une ontologie de modélisation des trajectoires qui est utilisée dans la construction des trajectoires sémantiques, ainsi qu'une ontologie d'observation dans l'environnement qui permet de relier les trajectoires à l'environnement.

La première ontologie qui est utilisée pour la modélisation de l'environnement est basée sur l'introduction de concepts permettant la description des objets statiques dans une scène. Elle est utilisée dans la modélisation des différents objets statiques qui se trouvent dans l'environnement surveillé. La deuxième ontologie est responsable de la transformation des trajectoires brutes en des trajectoires sémantiques par la segmentation

de ces trajectoires en épisodes qui ont un sens. La dernière ontologie est responsable de l'association des épisodes de la trajectoire structurée avec des éléments de l'environnement modélisé précédemment pour donner un sens aux épisodes de la trajectoire.

Ce module que nous avons proposé est basé sur le formalisme des ontologies pour associer des connaissances aux trajectoires. Il couvre à la fois la représentation d'un environnement et l'observation des événements de bas-niveaux issues d'un système de suivi multi-caméras. L'originalité de ce travail peut s'observer sur plusieurs points. On peut noter la généralité de notre ontologie pour l'interprétation des observations. En effet, notre modèle est conçu pour intégrer des mécanismes généraux concernant les connaissances modélisées et il laisse aux ontologies spécialisées la possibilité de définir les concepts spécifiques liés aux éléments observés et/ou aux entités de l'environnement. La structure de notre ontologie, basée sur les épisodes significatifs, permet l'annotation sémantique des épisodes d'arrêts et de mouvements et laisse l'interprétation des comportements de haut-niveau aux applications sémantiques basées sur notre ontologie. Pour les futurs travaux, nous envisageons de faire aboutir cette démarche en développant un mécanisme d'inférence de comportements de haut-niveau à partir des événements liés aux observations de bas-niveau. Dans un esprit de modularité et d'extensibilité, ceci se fera sans doute avec une ontologie des comportements proposant des concepts généraux à étendre par des ontologies de domaines et des règles SWRL permettant d'associer des comportements à des trajectoires.

Conclusion générale et perspectives

Dans le cadre de cette thèse, nous avons exploré les problématiques liées à l'étude du comportement en vue de caractériser les profils de clients dans une surface de ventes. D'une part, La mise en disposition d'études statistiques pour caractériser les différents profils de clients ne permettront pas seulement l'augmentation des ventes. Cependant, mais encore l'amélioration du service proposé ; mais d'autre part, la récupération de données suffisantes pour réaliser ces études avec une observation classique nécessite l'engagement de grands efforts financières. Cette causalité a conduit vers la recherche de nouvelles méthodes de collecte de données plus fiables et moins couteuses. L'approche la plus pertinente pour y répondre est l'exploitation des réseaux de caméras de surveillance déjà présents sur place. Par conséquent, la vision par ordinateur constitue une solution efficace pour développer des outils d'analyse automatiques.

En effet, les comportements les plus basiques caractérisant le profil d'un client sont directement liés à ces déplacements. Ces derniers peuvent constituer une base pour la création de profils reliant les types d'achat à l'environnement, le genre, ou encore la disposition des produits. Dans ce travail, nous avons alors développé une plate-forme d'interprétation des vidéos de surveillance pour l'analyse de mouvements. Cette dernière nous a permis à la fois de compter et de suivre les individus évoluant dans la scène.

Cette problématique d'étude des déplacements d'individus est communément connue comme le suivi de personnes. Elle peut intervenir dans le cadre de suivi mono-caméra ou encore le suivi multi-caméras. Pour cette dernière, plusieurs travaux de recherches, qui reposent sur des approches centralisées ou distribuées, ont été effectués.

Les travaux présentés dans cette thèse constitués un système distribué de suivi, et s'articule auteur de quatre axes principaux :

- Le comptage des personnes traversant une ligne d'intérêt.
- Le suivi mono-caméra de personnes.
- La ré-identification des personnes.
- La création des trajectoires sémantiques.

Comptage et suivi mono-caméra de personnes

Cette première partie est axée sur la conception d'un module de comptage et d'un module de suivi mono-caméra.

Dans la littérature, deux problématiques de comptage sont considérées, les comptages ROI et les comptages LOI. Nous avons adopté cette dernière dans le cadre de nos travaux par la conception d'une méthode originale basée sur l'exploitation des cartes de profondeur d'une kinect, et l'utilisation d'une grille afin de mesurer les entrées. Nous avons effectué plusieurs séries de tests afin de valider notre approche. En effet, les résultats obtenus ont montrés que notre algorithme permet la bonne détection de la majeure partie des individus avec un taux de détection dépassant les 95%. Une fois les individus détectés, et dans la perspective de déterminer la direction de leurs mouvement, nous avons associé des trackeurs aux détections. Pour cela, nous avons conçu une méthode reposant sur l'utilisation d'un modèle cinétique associé à un MeanShift. Là aussi, nous avons effectué plusieurs tests de validation, et les résultats obtenus ont montrés la capacité de notre algorithme à obtenir des taux très satisfaisants de suivi dépassant les 90%. La combinaison des deux phases décrites ci-dessus nous ont permis d'incrémenter les compteurs d'entrées et sorties afin de valider notre stratégie de comptage.

Dans la partie de suivi mono-caméra, nous avons développé un module responsable de la récupération des parcours d'individus pour chacune des caméras. Nous avons proposé une méthode de trois phases, une phase de détection, une phase de suivi, et une phase d'association. Nous avons commencé par l'implémentation et la validation de la phase de détection basée sur la détection et la segmentation de mouvement. Les résultats obtenus sont très satisfaisants dépassant les 90% de bonnes détections. Ensuite, Ces détections sont utilisées pour l'initialisation et la mise à jour de la liste des personnes suivies, où chaque individu est associé à un trackeur par filtre particulaire. La validation de cette phase, sur la base PETS 2009, a démontré la grande efficacité de ces trackeurs pour les suivis à courte durée. Cependant ces mêmes tests ont montré les limites de cette approche pour les suivis sur des longues durées. Ces limites nous ont poussés à implémenter un post-traitement responsable de la fusion des segments de trajectoires. Pour cela, nous avons utilisé une méthode de ré-identification que nous avons validée dans le cadre de la gestion des occultations et qui a montré une grande efficacité en la matière.

Ré-identification de personnes

Dans cette partie, nous avons proposé une méthode originale exploitant les différences locales pour la ré-identification. Elle sert à associer les individus suivis d'individus à travers les caméras du réseau.

Pour cela, nous avons proposé une méthode composée d'une étape d'apprentissage et d'une étape d'appariement. La première a pour objectif la sélection des patches saillant pour les individus à l'aide d'un nouveau framework de représentation d'individus et de sélection de régions saillantes. Dans ce dernier, chaque individu est représenté par un ensemble de détections, chaque détection est représentée par un ensemble de patches, et chaque patch est représenté par sa description spatiale et sa description d'apparence. Une fois les individus représentés, une procédure de sélection de patches saillants est appliquée. Les patches saillants retenus sont alors utilisés dans l'appariement.

Nous avons effectué plusieurs séries de tests de validation ; tout d'abord, pour la validation de notre méthode dans le cadre de la ré-identification multi-caméras. Nous avons validé cette approche par la construction des matrices de confusion qui ont montrées des taux de ré-identification très satisfaisants dépassant les 90% pour l'appariement par apprentissage profond. Nous avons par la suite tracé des courbes CMC montrant la stabilité et la robustesse de notre méthode. Nous avons comparé nos résultats avec d'autres méthodes existantes dans la littérature et nous avons constaté les améliorations apportées par notre approche dans le cadre du suivi. Nous avons alors effectué des tests pour valider notre approche dans le cadre d'une application de gestion d'occultations pour le suivi mono-caméra ; là encore, notre méthode a montrée des résultats très satisfaisants qui dépassent les 90% de taux de ré-identification.

Représentation sémantique des trajectoires

Dans cette partie, nous avons proposé un module dans le but est l'ajout de connaissances aux trajectoires spatio-temporelles brutes pour la construction de trajectoires sémantiques plus compréhensibles pour l'utilisateur.

Nous avons développé une plate-forme de construction de trajectoires sémantiques en se basant sur trois ontologies : une ontologie d'environnement, une ontologie de trajectoire, et une ontologie qui relié les trajectoires à l'environnement. Nous avons commencé par la construction d'une ontologie de description de l'environnement, cette ontologie qui hérite des concepts des ontologies géométrique est utilisé pour l'ajout des

connaissances sur les parties statiques de la scène. Ensuite vient la deuxième ontologie qui est celle des trajectoires, elle permet la conversion de la succession de positions spatio-temporelles à une succession d'épisodes qui représentent des intervalles de trajectoires sémantiquement reliés et auxquelles un sens peut être donné. Une fois les trajectoires segmentées à des épisodes et l'environnement décrit, nous avons construit une troisième ontologie qui relié les épisodes de trajectoires aux objets statique de l'environnement pour donner un sens à ces épisodes.

L'état de l'art, que nous avons fait, nous a montré que cette approche d'ajout de connaissances au suivi n'a jamais était étudié jusqu'à présent dans le cadre de l'analyse de comportement. L'utilisation des ontologies permettra la connexion des études géométriques de trajectoires aux connaissances d'experts dans le domaine du comportement.

Afin de poursuivre ce travail et comme perspectives, nous allons tout d'abord commencer par la considération des limites montrées par notre méthodes de ré-identification dans le cadre de la ré-identification à grande échèle, dont laquelle nous avons eu des résultats très satisfaisant mais qui sont moins importants que certaines méthodes existantes. Pour cela, nous comptons intégrer dans notre framework de sélection de patches discriminants une procédure de pondération qui permettra l'utilisation de plusieurs descripteurs dans la sélection des patches discriminants et par conséquent l'exploitation de cette pondération dans l'appariement des patches. Cette variation de descripteurs permettra la différenciation des patches qui apparaissaient jusqu'à présent comme des patches similaires.

Nous comptons aussi développer le rôle du module de gestion qui est basé sur l'utilisation des ontologies. Pour cela, nous comptons intégrer les ontologies développées précédemment dans le contrôle et l'orientation de suivi par la définition de concepts permettant, en premier temps, la validation des positions retourné par le modèle de suivi avec l'utilisations des concepts de zones vides dans notre algorithme de suivi. Ensuite, l'orientation de suivis par la définition préalable des zones probables de la prochaine détection d'un individu suivi par rapport à sa position actuel et l'historique des suivis.

Articles

Revues internationales

- **IGUERNAISSI, Rabah**, MERAD, Djamal, AZIZ, Kheireddine, *et al.* People tracking in multi-camera systems: a review. *Multimedia Tools and Applications*, 2018, p. 1-21.

- MERAD Djamal, DRAP Pierre, LUFIMPU-LUVIYA Yannick, **IGUERNAISSI, Rabah**, *et al.* Purchase behavior analysis through gaze and gesture observation. *Pattern Recognition Letters*, 2016, vol. 81, p. 21-29.

- MERAD, Djamal, AZIZ, Kheir-Eddine, **IGUERNAISSI, Rabah**, *et al.* Tracking multiple persons under partial and global occlusions: Application to customers' behavior analysis. *Pattern Recognition Letters*, 2016, vol. 81, p. 11-20.

- AZIZ, Kheireddine, MERAD, Djamal, **IGUERNAISSI, Rabah**, *et al.* Head detection based on skeleton graph method for counting people in crowded environments. *Journal of Electronic Imaging*, 2016, vol. 25, no 1, p. 013012.

Conférences internationales

- **IGUERNAISSI, Rabah**, MERAD, Djamal, et DRAP, Pierre. People Counting based on Kinect Depth Data. In : *ICPRAM*. 2018. p. 364-370.

- **IGUERNAISSI, Rabah**, MERAD, Djamal, et DRAP, Pierre. People's Re-identification Across Multiple Non-overlapping Cameras by Local Discriminative Patch Matching. In: *International Conference Image Analysis and Recognition*. Springer, Cham, 2017. p. 190-197.

- MAHIDDINE, Amine, **IGUERNAISSI, Rabah**, MERAD, Djamal, *et al.* 3D Registration of Multi-modal Data Using Surface Fitting. In : *ICPRAM (2)*. 2015. p. 71-78.

Références

[1] I. Collin-Lachaud and R. Vanheems, "Naviguer entre espaces virtuel et réel pour faire ses achats: exploration de l'expérience de shopping hybride," *Recherche et Applications en Marketing (French Edition),* vol. 31, pp. 43-61, 2016.

[2] M. L. Meuter, M. J. Bitner, A. L. Ostrom, and S. W. Brown, "Choosing among alternative service delivery modes: An investigation of customer trial of self-service technologies," *Journal of marketing,* vol. 69, pp. 61-83, 2005.

[3] B. Larivière, D. Bowen, T. W. Andreassen, W. Kunz, N. J. Sirianni, C. Voss, *et al.*, "“Service Encounter 2.0”: An investigation into the roles of technology, employees and customers," *Journal of Business Research,* vol. 79, pp. 238-246, 2017.

[4] I. Barth and B. Anteblian, "Comprendre les courses ordinaires: enjeux et implications pour les grandes surfaces alimentaires," *Décisions Marketing,* pp. 49-61, 2010.

[5] D. Merad, P. Drap, Y. Lufimpu-Luviya, R. Iguernaissi, and B. Fertil, "Purchase behavior analysis through gaze and gesture observation," *Pattern Recognition Letters,* vol. 81, pp. 21-29, 2016.

[6] D. Merad, K.-E. Aziz, R. Iguernaissi, B. Fertil, and P. Drap, "Tracking multiple persons under partial and global occlusions: Application to customers' behavior analysis," *Pattern Recognition Letters,* vol. 81, pp. 11-20, 2016.

[7] L.-L. Yannick, P. Sebastien, M. Djamel, and F. Bernard, "Combining regression and classification methods for age band estimation from human faces," in *Image and Signal Processing and Analysis (ISPA), 2013 8th International Symposium on*, 2013, pp. 136-141.

[8] M. Evans, C. J. Osborne, and J. Ferryman, "Multicamera object detection and tracking with object size estimation," in *Advanced Video and Signal Based Surveillance (AVSS), 2013 10th IEEE International Conference on*, 2013, pp. 177-182.

[9] Y. Wang, "Distributed multi-object tracking with multi-camera systems composed of overlapping and non-overlapping cameras," The University of Nebraska-Lincoln, 2013.

[10] L. Patino and J. Ferryman, "Multicamera trajectory analysis for semantic behaviour characterisation," in *Advanced Video and Signal Based Surveillance (AVSS), 2014 11th IEEE International Conference on*, 2014, pp. 369-374.

[11] M. C. Liem and D. M. Gavrila, "Joint multi-person detection and tracking from overlapping cameras," *Computer Vision and Image Understanding,* vol. 128, pp. 36-50, 2014.

[12] K. Kim and L. S. Davis, "Multi-camera tracking and segmentation of occluded people on ground plane using search-guided particle filtering," in *European Conference on Computer Vision*, 2006, pp. 98-109.

[13] F. Fleuret, J. Berclaz, R. Lengagne, and P. Fua, "Multicamera people tracking with a probabilistic occupancy map," *IEEE transactions on pattern analysis and machine intelligence*, vol. 30, pp. 267-282, 2008.

[14] T. T. Santos and C. H. Morimoto, "Multiple camera people detection and tracking using support integration," *Pattern Recognition Letters*, vol. 32, pp. 47-55, 2011.

[15] D. Baltieri, R. Vezzani, R. Cucchiara, A. Utasi, C. Benedek, and T. Szirányi, "Multi-view people surveillance using 3D information," in *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on*, 2011, pp. 1817-1824.

[16] W. Chen, L. Cao, X. Chen, and K. Huang, "An equalized global graph model-based approach for multi-camera object tracking," *IEEE Transactions on Circuits and Systems for Video Technology*, 2016.

[17] M. Hofmann, D. Wolf, and G. Rigoll, "Hypergraphs for joint multi-view reconstruction and multi-object tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3650-3657.

[18] C. R. del-Blanco, R. Mohedano, N. Garcia, L. Salgado, and F. Jaureguizar, "Color-based 3D particle filtering for robust tracking in heterogeneous environments," in *Distributed Smart Cameras, 2008. ICDSC 2008. Second ACM/IEEE International Conference on*, 2008, pp. 1-10.

[19] M. Bredereck, X. Jiang, M. Körner, and J. Denzler, "Data association for multi-object tracking-by-detection in multi-camera networks," in *Distributed Smart Cameras (ICDSC), 2012 Sixth International Conference on*, 2012, pp. 1-6.

[20] L. Wen, Z. Lei, M.-C. Chang, H. Qi, and S. Lyu, "Multi-Camera Multi-Target Tracking with Space-Time-View Hyper-graph," *International Journal of Computer Vision*, vol. 122, pp. 313-333, 2017.

[21] Z. Zhang and J. Cheng, "Multi-Camera Tracking Helmet System," *Journal of Image and Graphics*, vol. 1, 2013.

[22] W. Du and J. Piater, "Multi-camera people tracking by collaborative particle filters and principal axis-based integration," *Computer Vision–ACCV 2007*, pp. 365-374, 2007.

[23] C.-C. Huang and S.-J. Wang, "A Bayesian hierarchical framework for multitarget labeling and correspondence with ghost suppression over multicamera surveillance system," *IEEE Transactions on Automation Science and Engineering*, vol. 9, pp. 16-30, 2012.

[24] S. M. Khan and M. Shah, "A multiview approach to tracking people in crowded scenes using a planar homography constraint," in *European Conference on Computer Vision*, 2006, pp. 133-146.

[25] R. Eshel and Y. Moses, "Tracking in a dense crowd using multiple cameras," *International journal of computer vision,* vol. 88, pp. 129-143, 2010.

[26] L. Guan, J.-S. Franco, and M. Pollefeys, "Multi-view occlusion reasoning for probabilistic silhouette-based dynamic scene reconstruction," *International journal of computer vision,* vol. 90, pp. 283-303, 2010.

[27] D. Arsic, E. Hristov, N. Lehment, B. Hornler, B. Schuller, and G. Rigoll, "Applying multi layer homography for multi camera person tracking," in *Distributed Smart Cameras, 2008. ICDSC 2008. Second ACM/IEEE International Conference on*, 2008, pp. 1-9.

[28] T. Kroeger, R. Dragon, and L. Van Gool, "Multi-view tracking of multiple targets with dynamic cameras," in *German Conference on Pattern Recognition*, 2014, pp. 653-665.

[29] J. Yao and J.-M. Odobez, "Multi-camera multi-person 3D space tracking with MCMC in surveillance scenarios," in *European Conference on Computer Vision, Workshop on Multi Camera and Multi-modal Sensor Fusion Algorithms and Applications (ECCV-M2SFA2)*, 2008.

[30] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE transactions on pattern analysis and machine intelligence,* vol. 32, pp. 1627-1645, 2010.

[31] L. Leal-Taixe, G. Pons-Moll, and B. Rosenhahn, "Branch-and-price global optimization for multi-view multi-target tracking," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012, pp. 1987-1994.

[32] X. Jiang, M. Körner, D. Haase, and J. Denzler, "A graph-based MAP solution for multi-person tracking using multi-camera systems," in *Computer Vision Theory and Applications (VISAPP), 2014 International Conference on*, 2014, pp. 343-350.

[33] D. Baltieri, R. Vezzani, and R. Cucchiara, "3D Body Model Construction and Matching for Real Time People Re-Identification," in *Eurographics Italian Chapter Conference*, 2010, pp. 65-71.

[34] W. Du and J. Piater, "Data fusion by belief propagation for multi-camera tracking," in *Information Fusion, 2006 9th International Conference on*, 2006, pp. 1-8.

[35] M. Taj and A. Cavallaro, "Distributed and decentralized multicamera tracking," *IEEE Signal Processing Magazine,* vol. 28, pp. 46-58, 2011.

[36] S. Bąk, G. Charpiat, E. Corvee, F. Bremond, and M. Thonnat, "Learning to match appearances by correlations in a covariance metric space," in *European Conference on Computer Vision*, 2012, pp. 806-820.

[37] X. Wang, G. Doretto, T. Sebastian, J. Rittscher, and P. Tu, "Shape and appearance context modeling," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, 2007, pp. 1-8.

[38] O. Hamdoun, F. Moutarde, B. Stanciulescu, and B. Steux, "Person re-identification in multi-camera system by signature based on interest point

descriptors collected on short video sequences," in *Distributed Smart Cameras, 2008. ICDSC 2008. Second ACM/IEEE International Conference on*, 2008, pp. 1-6.

[39] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person re-identification by symmetry-driven accumulation of local features," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, 2010, pp. 2360-2367.

[40] S. Bak, E. Corvee, F. Brémond, and M. Thonnat, "Person re-identification using spatial covariance regions of human body parts," in *Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on*, 2010, pp. 435-440.

[41] W. Chen, L. Cao, X. Chen, and K. Huang, "A novel solution for multi-camera object tracking," in *Image Processing (ICIP), 2014 IEEE International Conference on*, 2014, pp. 2329-2333.

[42] K.-E. Aziz, D. Merad, and B. Fertil, "People re-identification across multiple non-overlapping cameras system by appearance classification and silhouette part segmentation," in *Advanced Video and Signal-Based Surveillance (AVSS), 2011 8th IEEE International Conference on*, 2011, pp. 303-308.

[43] D. S. Cheng, M. Cristani, M. Stoppa, L. Bazzani, and V. Murino, "Custom Pictorial Structures for Re-identification," in *BMVC*, 2011, p. 6.

[44] A. Alahi, P. Vandergheynst, M. Bierlaire, and M. Kunt, "Cascade of descriptors to detect and track objects across any network of cameras," *Computer Vision and Image Understanding,* vol. 114, pp. 624-640, 2010.

[45] D.-N. T. Cong, L. Khoudour, C. Achard, C. Meurie, and O. Lezoray, "People re-identification by spectral classification of silhouettes," *Signal Processing,* vol. 90, pp. 2362-2374, 2010.

[46] H. Bouma, S. Borsboom, R. Hollander, and M. Worring, "Re-identification of persons in multicamera surveillance under varying viewpoints and illumination," in *Proceedings of SPIE, the International Society for Optical Engineering*, 2012, pp. 8359-26.

[47] W. Nie, A. Liu, Y. Su, H. Luan, Z. Yang, L. Cao, *et al.*, "Single/cross-camera multiple-person tracking by graph matching," *Neurocomputing,* vol. 139, pp. 220-232, 2014.

[48] K. Jungling and M. Arens, "View-invariant person re-identification with an implicit shape model," in *Advanced Video and Signal-Based Surveillance (AVSS), 2011 8th IEEE International Conference on*, 2011, pp. 197-202.

[49] B. Leibe, A. Leonardis, and B. Schiele, "Robust object detection with interleaved categorization and segmentation," *International journal of computer vision,* vol. 77, pp. 259-289, 2008.

[50] R. Mazzon, S. F. Tahir, and A. Cavallaro, "Person re-identification in crowd," *Pattern Recognition Letters,* vol. 33, pp. 1828-1837, 2012.

[51] W. Brendel, M. Amer, and S. Todorovic, "Multiobject tracking as maximum weight independent set," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, 2011, pp. 1273-1280.

[52] W. Li, Y. Wu, M. Mukunoki, and M. Minoh, "Common-near-neighbor analysis for person re-identification," in *Image Processing (ICIP), 2012 19th IEEE International Conference on*, 2012, pp. 1621-1624.

[53] A. Colombo, J. Orwell, and S. Velastin, "Colour constancy techniques for re-recognition of pedestrians from multiple surveillance cameras," in *Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications-M2SFA2 2008*, 2008.

[54] R. Mazzon and A. Cavallaro, "Multi-camera tracking using a multi-goal social force model," *Neurocomputing,* vol. 100, pp. 41-50, 2013.

[55] S. Bak, E. Corvee, F. Brémond, and M. Thonnat, "Person re-identification using haar-based and dcd-based signature," in *Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on*, 2010, pp. 1-8.

[56] B. J. Prosser, W.-S. Zheng, S. Gong, T. Xiang, and Q. Mary, "Person re-identification by support vector ranking," in *BMVC*, 2010, p. 6.

[57] D. Figueira, L. Bazzani, H. Q. Minh, M. Cristani, A. Bernardino, and V. Murino, "Semi-supervised multi-feature learning for person re-identification," in *Advanced Video and Signal Based Surveillance (AVSS), 2013 10th IEEE International Conference on*, 2013, pp. 111-116.

[58] M. Koestinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012, pp. 2288-2295.

[59] Z. Lin and L. S. Davis, "Learning pairwise dissimilarity profiles for appearance recognition in visual surveillance," in *International symposium on visual computing*, 2008, pp. 23-34.

[60] W. R. Schwartz and L. S. Davis, "Learning discriminative appearance-based models using partial least squares," in *Computer Graphics and Image Processing (SIBGRAPI), 2009 XXII Brazilian Symposium on*, 2009, pp. 322-329.

[61] T. Avraham, I. Gurvich, M. Lindenbaum, and S. Markovitch, "Learning implicit transfer for person re-identification," in *Computer Vision–ECCV 2012. Workshops and Demonstrations*, 2012, pp. 381-390.

[62] C. Nakajima, M. Pontil, B. Heisele, and T. Poggio, "Full-body person recognition system," *Pattern recognition,* vol. 36, pp. 1997-2006, 2003.

[63] V. S. Kenk, S. Kovačič, M. Kristan, M. Hajdinjak, and J. Perš, "Visual re-identification across large, distributed camera networks," *Image and vision computing,* vol. 34, pp. 11-26, 2015.

[64] C. C. Loy, C. Liu, and S. Gong, "Person re-identification by manifold ranking," in *Image Processing (ICIP), 2013 20th IEEE International Conference on*, 2013, pp. 3567-3571.

[65] H. Q. Minh, L. Bazzani, and V. Murino, "A unifying framework for vector-valued manifold regularization and multi-view learning," in *ICML (2)*, 2013, pp. 100-108.

[66] C.-H. Kuo, C. Huang, and R. Nevatia, "Multi-target tracking by on-line learned discriminative appearance models," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, 2010, pp. 685-692.

[67] L. Bazzani, M. Cristani, A. Perina, and V. Murino, "Multiple-shot person re-identification by chromatic and epitomic analyses," *Pattern Recognition Letters,* vol. 33, pp. 898-903, 2012.

[68] W.-S. Zheng, S. Gong, and T. Xiang, "Reidentification by relative distance comparison," *IEEE transactions on pattern analysis and machine intelligence,* vol. 35, pp. 653-668, 2013.

[69] O. Tuzel, F. Porikli, and P. Meer, "Region covariance: A fast descriptor for detection and classification," *Computer Vision–ECCV 2006,* pp. 589-600, 2006.

[70] O. Javed, K. Shafique, Z. Rasheed, and M. Shah, "Modeling inter-camera space–time and appearance relationships for tracking across non-overlapping views," *Computer Vision and Image Understanding,* vol. 109, pp. 146-162, 2008.

[71] B. Antić, D. Letić, D. Ćulibrk, and V. Crnojević, "K-means based segmentation for real-time zenithal people counting," in *Image Processing (ICIP), 2009 16th IEEE International Conference on*, 2009, pp. 2565-2568.

[72] E. Bondi, L. Seidenari, A. D. Bagdanov, and A. Del Bimbo, "Real-time people counting from depth imagery of crowded environments," in *Advanced Video and Signal Based Surveillance (AVSS), 2014 11th IEEE International Conference on*, 2014, pp. 337-342.

[73] C.-H. Chen, T.-Y. Chen, D.-J. Wang, and T.-J. Chen, "A cost-effective people-counter for a crowd of moving people based on two-stage segmentation," *Journal of Information Hiding and Multimedia Signal Processing,* vol. 3, pp. 12-25, 2012.

[74] H. Fu, H. Ma, and H. Xiao, "Scene-adaptive accurate and fast vertical crowd counting via joint using depth and color information," *Multimedia tools and applications,* vol. 73, pp. 273-289, 2014.

[75] T. Van Oosterhout, S. Bakkes, and B. J. Kröse, "Head Detection in Stereo Data for People Counting and Segmentation," in *VISAPP*, 2011, pp. 620-625.

[76] X. Zhang, J. Yan, S. Feng, Z. Lei, D. Yi, and S. Z. Li, "Water filling: Unsupervised people counting via vertical kinect sensor," in *Advanced Video and Signal-Based Surveillance (AVSS), 2012 IEEE Ninth International Conference on*, 2012, pp. 215-220.

[77] L. Zhang and L. van der Maaten, "Structure preserving object tracking," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, 2013, pp. 1838-1845.

[78] L. Del Pizzo, P. Foggia, A. Greco, G. Percannella, and M. Vento, "Counting people by RGB or depth overhead cameras," *Pattern Recognition Letters,* vol. 81, pp. 41-50, 2016.

[79] Y. Cong, H. Gong, S.-C. Zhu, and Y. Tang, "Flow mosaicking: Real-time pedestrian counting without scene-specific learning," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 1093-1100.

[80] A. Ellin and G. Dolsak, "The design and application of rotary encoders," *Sensor Review,* vol. 28, pp. 150-158, 2008.

[81] J. Barandiaran, B. Murguia, and F. Boto, "Real-time people counting using multiple lines," in *Image Analysis for Multimedia Interactive Services, 2008. WIAMIS'08. Ninth International Workshop on*, 2008, pp. 159-162.

[82] W. Luo, X. Zhao, and T.-K. Kim, "Multiple object tracking: A review," *arXiv preprint arXiv:1409.7618,* vol. 1, 2014.

[83] A. Andriyenko and K. Schindler, "Multi-target tracking by continuous energy minimization," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, 2011, pp. 1265-1272.

[84] A. Milan, K. Schindler, and S. Roth, "Detection-and trajectory-level exclusion in multiple object tracking," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 3682-3689.

[85] J. Berclaz, F. Fleuret, E. Turetken, and P. Fua, "Multiple object tracking using k-shortest paths optimization," *IEEE transactions on pattern analysis and machine intelligence,* vol. 33, pp. 1806-1819, 2011.

[86] L. Leal-Taixé, M. Fenzi, A. Kuznetsova, B. Rosenhahn, and S. Savarese, "Learning an image-based motion context for multiple people tracking," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3542-3549.

[87] W. Hu, X. Li, W. Luo, X. Zhang, S. Maybank, and Z. Zhang, "Single and multiple object tracking using log-Euclidean Riemannian subspace and block-division appearance model," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 34, pp. 2420-2440, 2012.

[88] L. Zhang and L. van der Maaten, "Structure preserving object tracking," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 1838-1845.

[89] L. Zhang and L. Van Der Maaten, "Preserving structure in model-free tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 36, pp. 756-769, 2014.

[90] M. Yang, T. Yu, and Y. Wu, "Game-theoretic multiple target tracking," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, 2007, pp. 1-8.

[91] A. Andriyenko, K. Schindler, and S. Roth, "Discrete-continuous optimization for multi-target tracking," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012, pp. 1926-1933.

[92] B. Yang and R. Nevatia, "An online learned crf model for multi-target tracking," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012, pp. 2034-2041.

[93] Y. Shi and W. C. Karl, "Real-time tracking using level sets," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, 2005, pp. 34-41.

[94] S.-H. Bae and K.-J. Yoon, "Robust online multi-object tracking based on tracklet confidence and online discriminative appearance learning," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 1218-1225.

[95] S. Chen, A. Fern, and S. Todorovic, "Multi-object tracking via constrained sequential labeling," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1130-1137.

[96] S. Tang, M. Andriluka, and B. Schiele, "Detection and tracking of occluded people," *International Journal of Computer Vision*, vol. 110, pp. 58-69, 2014.

[97] H. Izadinia, I. Saleemi, W. Li, and M. Shah, "2T: Multiple people multiple parts tracker," in *European Conference on Computer Vision*, 2012, pp. 100-114.

[98] S. Tang, M. Andriluka, A. Milan, K. Schindler, S. Roth, and B. Schiele, "Learning people detectors for tracking in crowded scenes," in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 1049-1056.

[99] H. Pirsiavash, D. Ramanan, and C. C. Fowlkes, "Globally-optimal greedy algorithms for tracking a variable number of objects," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, 2011, pp. 1201-1208.

[100] K. Nummiaro, E. Koller-Meier, and L. Van Gool, "Object tracking with an adaptive color-based particle filter," *Pattern Recognition*, pp. 353-360, 2002.

[101] A. Shimada, D. Arita, and R.-i. Taniguchi, "Dynamic control of adaptive mixture-of-Gaussians background model," in *Video and Signal Based Surveillance, 2006. AVSS'06. IEEE International Conference on*, 2006, pp. 5-5.

[102] S. Bak, E. Corvee, F. Bremond, and M. Thonnat, "Person Re-identification Using Haar-based and DCD-based Signature," in *Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on*, 2010, pp. 1-8.

[103] R. Zhao, W. Ouyang, and X. Wang, "Learning mid-level filters for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 144-151.

[104] K. E. Aziz, D. Merad, and B. Fertil, "People re-identification across multiple non-overlapping cameras system by appearance classification and silhouette part segmentation," in *Advanced Video and Signal-Based Surveillance (AVSS), 2011 8th IEEE International Conference on*, 2011, pp. 303-308.

[105] N. Martinel and G. L. Foresti, "Multi-signature based person re-identification," *Electronics Letters,* vol. 48, pp. 765-767, 2012.

[106] L. Bazzani, M. Cristani, A. Perina, M. Farenzena, and V. Murino, "Multiple-Shot Person Re-identification by HPE Signature," in *Proceedings of the 2010 20th International Conference on Pattern Recognition*, ed. Washington, DC, USA: IEEE Computer Society, 2010, pp. 1413-1416.

[107] N. Martinel, C. Micheloni, and G. L. Foresti, "Kernelized saliency-based person re-identification through multiple metric learning," *IEEE Transactions on Image Processing*, vol. 24, pp. 5645-5658, 2015.

[108] R. Zhao, W. Ouyang, and X. Wang, "Person re-identification by salience matching," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 2528-2535.

[109] A. Datta, L. M. Brown, R. Feris, and S. Pankanti, "Appearance modeling for person re-identification using weighted brightness transfer functions," in *Pattern Recognition (ICPR), 2012 21st International Conference on*, 2012, pp. 2367-2370.

[110] A. Colombo, J. Orwell, and S. Velastin, "Colour constancy techniques for re-recognition of pedestrians from multiple surveillance cameras " *in Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications-M2SFA2 2008,* 2008.

[111] O. Javed, K. Shafique, Z. Rasheed, and M. Shah, "Modeling inter-camera space-time and appearance relationships for tracking across non-overlapping views," *Comput. Vis. Image Underst.,* vol. 109, pp. 146-162, February 2008.

[112] N. Martinel, C. Micheloni, and G. L. Foresti, "A pool of multiple person re-identification experts," *Pattern Recognition Letters,* vol. 71, pp. 23-30, 2016.

[113] T. Avraham and M. Lindenbaum, "Learning appearance transfer for person re-identification," in *Person Re-Identification*, ed: Springer, 2014, pp. 231-246.

[114] W. Li, Y. Wu, M. Mukunoki, and M. Minoh, "Common-near-neighbor analysis for person re-identification," in *2012 19th IEEE International Conference on Image Processing*, 2012, pp. 1621-1624.

[115] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re-identification by local maximal occurrence representation and metric learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2197-2206.

[116] S. Pedagadi, J. Orwell, S. Velastin, and B. Boghossian, "Local fisher discriminant analysis for pedestrian re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3318-3325.

[117] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097-1105.

[118] A. Schumann and E. Monari, "A soft-biometrics dataset for person tracking and re-identification," in *Advanced Video and Signal Based Surveillance (AVSS), 2014 11th IEEE International Conference on*, 2014, pp. 193-198.

[119] D. Gray, S. Brennan, and H. Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," in *Proc. IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS)*, 2007, pp. 1-7.

[120] D. Baltieri, R. Vezzani, and R. Cucchiara, "3dpes: 3d people dataset for surveillance and forensics," in *Proceedings of the 2011 joint ACM workshop on Human gesture and behavior understanding*, 2011, pp. 59-64.

[121] D. Chen, Z. Yuan, G. Hua, N. Zheng, and J. Wang, "Similarity learning on an explicit polynomial kernel feature map for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1565-1573.

[122] N. Martinel, C. Micheloni, and G. L. Foresti, "Saliency Weighted Features for Person Re-identification," in *ECCV Workshops (3)*, 2014, pp. 191-208.

[123] L. Zheng, S. Wang, L. Tian, F. He, Z. Liu, and Q. Tian, "Query-adaptive late fusion for image search and person re-identification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1741-1750.

[124] R. Zhao, W. Ouyang, and X. Wang, "Unsupervised salience learning for person re-identification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3586-3593.

[125] Z. Li, S. Chang, F. Liang, T. S. Huang, L. Cao, and J. R. Smith, "Learning locally-adaptive decision functions for person verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3610-3617.

[126] R. Wannous, J. Malki, A. Bouju, and C. Vincent, "Trajectory ontology inference considering domain and temporal dimensions—Application to marine mammals," *Future Generation Computer Systems,* vol. 68, pp. 491-499, 2017.

[127] G. M. Santipantakis, G. A. Vouros, A. Glenis, C. Doulkeridis, and A. Vlachou, "The datAcron ontology for semantic trajectories," in *European Semantic Web Conference*, 2017, pp. 26-30.

[128] Z. Yan, D. Chakraborty, C. Parent, S. Spaccapietra, and K. Aberer, "SeMiTri: a framework for semantic annotation of heterogeneous trajectories," in *Proceedings of the 14th international conference on extending database technology*, 2011, pp. 259-270.

[129] S. Spaccapietra, C. Parent, M. L. Damiani, J. A. de Macedo, F. Porto, and C. Vangenot, "A conceptual view on trajectories," *Data & knowledge engineering,* vol. 65, pp. 126-146, 2008.

[130] Z. Yan, C. Parent, S. Spaccapietra, and D. Chakraborty, "A hybrid model and computing platform for spatio-semantic trajectories," in *Extended Semantic Web Conference*, 2010, pp. 60-75.

[131] J. J.-C. Ying, E. H.-C. Lu, W.-C. Lee, T.-C. Weng, and V. S. Tseng, "Mining user similarity from semantic trajectories," in *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks*, 2010, pp. 19-26.

[132] A. Vandecasteele, "Modélisation ontologique des connaissances expertes pour l'analyse de comportements à risque: application à la surveillance maritime," Paris, ENMP, 2012.

[133] P. Kotler, "Atmospherics as a marketing tool," *Journal of retailing*, vol. 49, pp. 48-64, 1973.

[134] L. Yang and M. Worboys, "A navigation ontology for outdoor-indoor space:(work-in-progress)," in *Proceedings of the 3rd ACM SIGSPATIAL international workshop on indoor spatial awareness*, 2011, pp. 31-34.