République Tunisienne
Ministère de l'Enseignement Supérieur
et de la Recherche Scientifique

Université de Sfax
École Nationale d'Ingénieurs de Sfax

Ecole Doctorale
Sciences et Technologies

Thèse de *DOCTORAT*
en Ingénierie
des Systèmes Informatiques
N° d'ordre: 2017– 58

THESE

Présentée à

L'École Nationale d'Ingénieurs de Sfax

En vue de l'obtention du

DOCTORAT

Dans la discipline Ingénierie des Systèmes Informatique (ISI)

Par
Nabila MANSOURI

(Master de recherche en informatique)

APPROCHE AUTOMATIQUE À BASE DE TRAITEMENT D'IMAGES POUR L'ANALYSE COMPORTEMENTALE DE PIÉTONS ÂGÉS LORS DE LA TRAVERSÉE D'UNE RUE

Soutenu le 08 Décembre 2017, devant le jury composé de :

Mme.	**Monia TURKI** (Professeur, Université de Tunis El Manar)	Président
M.	**Atef MASMOUDI** (Maitre de conférences-HDR, Université de Sfax)	Rapporteur
M.	**François BREMOND** (Directeur de recherche INRIA, Sophia Antipolis)	Rapporteur
M.	**Frédéric LERASLE** (Professeur, Université Paul Sabatier de Toulouse)	Examinateur
Mme.	**Yousra BEN JEMAA** (Maitre de conférences-HDR, Université de Sfax)	Directeur de Thèse
M.	**Eric WATELAIN** (Maitre de conférences-HDR, Université de Valenciennes)	Directeur de Thèse

République Tunisienne
Ministère de l'Enseignement Supérieur
et de la Recherche Scientifique

Université de Sfax
École Nationale d'Ingénieurs de Sfax

Ecole Doctorale
Sciences et Technologies

Thèse de *DOCTORAT*
en Ingénierie
des Systèmes Informatiques
N° d'ordre: 2017– 58

APPROCHE AUTOMATIQUE À BASE DE TRAITEMENT D'IMAGES POUR L'ANALYSE COMPORTEMENTALE DE PIÉTONS ÂGÉS LORS DE LA TRAVERSÉE D'UNE RUE

Nabila MANSOURI

Résumé : Le trafic routier est devenu de plus en plus intense. Une telle situation avec le manque de prudence des piétons constituent deux causes majeures de l'augmentation des accidents routiers. En France, 16% des accidents de la route en 2016 impliquent au moins un piéton et chaque année, environ de 800 piétons sont tués dans un accident de la circulation. De plus, la part des plus de 65 ans dans la mortalité piétonne est en hausse de 13% entre 2014 et 2016. Ainsi, par ce projet de thèse nous proposons une approche probabiliste pour inférer le type de comportement (à risque ou sécurisé) des piétons lors de la traversé de la rue. Cette approche se compose de 2 couches principales : Une couche basse, basée sur les techniques de vision par ordinateur, pour la collecte des paramètres des piétons, du trafic et des aménagements urbains et une couche haute, basée sur le Réseau Bayésien (RB), pour l'inférence du type de comportement. Plusieurs contributions et améliorations sont proposées pour la construction d'une telle approche que ce soit au niveau de la couche basse (techniques de détection et de suivi utilisées) ou au niveau de la couche haute (gestion des incertitudes des capteurs de vision et la mise en relation des paramètres hétérogènes et variées).

Abstract: Road traffic has become more and more intense. Such as situation with the lack of attention of pedestrians are the two major causes of the increase in road accidents. In France, 16% of road accidents in 2016 involve at least one pedestrian and each year about 800 pedestrians are killed in a traffic accident. In addition, the proportion of pedestrian deaths having over 65 years old is up to 13% between 2014 and 2016. In fact, this thesis proposes a probabilistic approach to recognize pedestrians' behavior (risky or secure) when crossing the street. This approach consists of two main layers: a low layer, based on computer vision techniques, for collecting pedestrian and traffic parameters and a high layer, based on the Bayesian Network, for behavior prediction. Several contributions and improvements are proposed for the construction of such an approach, whether at the level of the low layer (detection and tracking process) or at the level of the upper layer (management of the uncertainties of the vision sensors and the correlation of heterogeneous and varied parameters).

Mots clés: Sécurité routière, ITS, Vision par ordinateur, Détection, Suivi, RB, Gestion d'incertitude.

Key-words: Road security, ITS, Computer vision, Detection, Tracking, BN, Uncertainty management.

Université de Valenciennes et du Hainaut-Cambrésis

Thèse de doctorat

Pour obtenir le grade de Docteur de l'Université de

VALENCIENNES ET DU HAINAUT-CAMBRESIS

et de

L'ECOLE NATIONALE D'INGENIEURES DE SFAX

Ecole doctorale Sciences Pour l'Ingénieur (SPI)

Discipline, spécialité: Informatique

Approche automatique à base de traitement d'images pour l'analyse comportementale de piétons âgés lors de la traversée d'une rue

Présentée par **Nabila MANSOURI**

Soutenue le 08 Décembre 2017

Equipe de recherche: Science de l'Homme et du Vivant (SHV)
Laboratoire d'Automatique, de Mécanique et d'Informatique Industrielles et Humaines (LAMIH)

JURY

Mme. Monia TURKI, Professeur, Université de Tunis El Manar (Président)
M. Atef MASMOUDI, Maitre de conférences-HDR, Université de Sfax (Rapporteur)
M. François BREMON, Directeur de recherche, Sophia Antipolis (Rapporteur)
M. Frédéric LERASLE, Professeur, Université Paul Sabatier de Toulouse (Examinateur)
Co- Directeurs de thèse
Mme. Yousra BEN JEMAA, Maitre de conférences-HDR, Université de Sfax
M. Eric WATELAIN, Maitre de conférences-HDR, Université de Valenciennes

Dédicaces

"Moi je n'arrive à transmettre que deux choses, l'énergie et le doute"

Bartabas

Le célèbre écuyer, pédagogue, metteur en scène et scénographe français, fondateur du spectacle équestre Zingaro et est depuis 2003 responsable de l'Académie du spectacle équestre de Versailles.

Je dédicace ce mémoire de thèse à toute personne qui m'a transmis ces deux facteurs promoteurs (l'énergie et le doute) pour tout travail de recherche.

Nabila MANSOURI

Remerciements

Je remercie vivement mon encadrante **Mme. Yousra BEN JEMAA**, maître de conférences-HDR à l'École Nationale d'Ingénieurs de Sfax (ENIS), pour l'excellence de son accompagnement ainsi que pour la confiance et la grande autonomie qu'elle m'a accordées tout au long de la réalisation de ce travail. Sa disponibilité illimitée, ses conseils constructifs et le suivi minutieux de toutes les particularités de mon travail méritaient d'être ardemment remerciés.

Mes remerciements vont également à **Mr. Eric WATELAIN**, maître de conférences-HDR à l'Université de Valenciennes, mon encadrant à l'université d'accueil, pour la gentillesse et la patience qu'il a manifesté à mon égard durant la réalisation de ce travail ainsi que pour tous les conseils et les documents qu'il m'a fournis. Son attitude collaborative a grandement facilité la réalisation de cette thèse, surtout pour l'acquisition de la base d'évaluation à Valenciennes et à Toulon.

J'exprime toute ma gratitude envers les membres du jury qui ont accepté de consacrer à cette thèse une partie de leurs temps extrêmement précieux.

Je commence par remercier **Mme. Monia TURKI**, Professeur de l'université de Tunis El Manar, pour avoir accepté de juger ce travail et d'en présider le jury de soutenance.

Je remercie également **Mr. Frédéric LERASLE**, Professeur à l'université Paul Sabatier de Toulouse qui m'a honoré par sa présence en qualité d'examinateur de ma thèse.

Je remercie aussi les rapporteurs de ce travail **Mr. Atef MASMOUDI**, Maitre de conférences-HDR à l'Université de Sfax et **Mr. François BREMOND** directeur de recherche à l'INRIA - Sophia Antipolis - France, pour leurs lectures attentives et leurs remarques pertinentes. Ainsi que pour l'honneur qu'ils me font en acceptant de bien

vouloir participer au jury.

Toute ma reconnaissance s'adresse aussi à l'Agence Universitaire de la Francophonie (AUF) et à Campus France pour avoir financé une grande partie de mes séjours en France suite à l'obtention des deux bourses respectivement de doctorat de l'AUF et d'excellence Eiffel.

Cette thèse de Doctorat a été réalisée dans le cadre d'une convention de cotutelle entre l'Ecole Nationale d'Ingénieurs de Sfax, Université de Sfax et l'Université de Valenciennes. Les recherches qui font l'objet de ce mémoire ont été réalisées sur deux sites : en France, au sein du Laboratoire d'Automatique, de Mécanique et d'Informatique Industrielles et Humaines (LAMIH) UMR CNRS 8201 et en Tunisie, dans le Laboratoire de Recherche en Développement et Contrôle des Applications Distribuées (ReDCAD). A cette occasion je teins à remercier tous les membres des deux laboratoires pour l'ambiance accueillante et pour tous les échanges pertinents. Mes remerciements s'adressent également à tous mes ami(e)s au sein de deux laboratoires que j'ai côtoyé quotidiennement durant ces années de thèse.

Un merci très particulier à **Mr. Cina MOTAMED**, Maitre de conférences-HDR à l'Université du Littoral Côte d'Opal pour sa collaboration précieuse et pertinente qui a fait avancé mes travaux de thèse à certains moments de blocage. Aussi que pour son chaleureux accueil lors d'un stage à l'ULCO.

Je voudrais exprimer aussi ma gratitude et ma reconnaissance envers tous les enseignants qui ont contribué à ma formation de l'école primaire jusqu'à l'université. Je remercie également tous les agents administratifs de l'Institut Supérieur d'Informatique et de Multimédia de Sfax, de l'Ecole Nationale d'Ingénieurs de Sfax et de l'Université de Valenciennes.

Je suis enfin sincèrement reconnaissante à mes fidèles et idéaux parents pour leur confiance, leur sacrifice et leur soutien inconditionnel durant tout mon cursus. Mes remerciements vont tout naturellement à mes sœurs et mes frères pour leurs encouragements et aides. Une remerciement assez particulier s'adresse à mon chér mari pour sa patience et ses sacrifices réalisés pour m'avoir me supporté pour me permettre de finaliser mes études.

Préambule

Ces derniers décades, la sécurité routière est hissée au rang de priorité dans les politiques des payés développés. L'université de Valenciennes dispose par ailleurs d'une UMR CNRS orientées vers les sciences pour l'ingénieur et spécialisée dans la sécurité pour les transports : le LAMIH (Laboratoire d'Automatique, de Mécanique et d'Informatique industrielles et Humaines). Cette dernière dispose : d'une solide expérience sur l'étude de la locomotion humaine normale et pathologique dans son équipe DEMhO, de solides expériences dans la sécurité dans les transports terrestre et l'automatique dans ses trois autres équipes ainsi que de nombreuses plateformes technologique de simulation dédiés à la sécurité. L'objectif de cette thèse s'inscrit alors, dans ce contexte de conception et d'évaluation d'un outil facilement accessible, permettant une approche écologique de l'analyse du comportement du piéton âgée lors de la traversée de rues. Cet outil s'inscrit dans la continuité de ses travaux, pour la conception d'une approche écologique, des travaux de réalité projetée à la base d'un simulateur, déjà réalisés sous la supervision de M. Eric WATELAIN, porteur de Projet SEVAP [1] au LAMIH.

Pour cela, l'approche réalisée consiste, à partir d'image vidéo issues de trois cameras du commerce, à développer une cascade de traitements informatiques automatisés pour extraire les piétons, les véhicules et les caractéristiques de l'environnement, pour en déduire le comportement globale du piéton à chaque instant, notamment en termes de prise de risques pour son intégrité. Ainsi, le développement d'un système automatique d'étude du comportement du piéton âgé in situ représente un outil non intrusif pertinent dans le cadre des travaux respectifs tant fondamentaux qu' appliqués des deux universités de tutelle.

1. Le programme PISTES : " PIétonS et comportement de TraverséE de rueS" financé par la fondation sécurité routière. Le projet PISTES est composé par 4 sous-projets à savoir :
   - SICAP "SImulation de traversée de carrefour par des Piétons",
   - AUCRE " Aménagement Urbain et Conformité à la REglementation ",
   - SEVAP " Sénior en Ville A Pied ",
   - SEPIA "SEcurité du Piéton Agé",
   - HYPCOPIA " Prise de médicaments HYPnotiques et COmportement du PIéton A risque "

Le LAMIH de l'Université de Valenciennes est porteur du projet SEVAP.

En fait, ce projet de thèse est l'objet d'une co-tutelle entre l'Ecole Nationale d'ingénieurs de Sfax (ENIS)- Université de Sfax et le LAMIH de l'Université de Valenciennes et du Hainaut-Cambrésis (UVHC). Ce projet de doctorat conduira à la délivrance par chacune des deux universités en coopération du titre de docteur qu'elle est habilitée à délivrer, à savoir :

— Pour l'Université de Sfax : "Docteur en Ingénierie des Systèmes Informatiques (ISI)"

— Pour l'Université de Valenciennes et du Hainaut-Cambrésis : "Docteur de l'Université de Valenciennes et du Hainaut-Cambrésis en Automatique, Génie informatique, Traitement du Signal".

Résumé

Le piéton fait partie des usagers de la voie publique vulnérable dans la mesure où il ne se déplace pas protégé par "l'enveloppe physique" de son moyen de transport. Il paye ainsi un lourd tribu à la morbidité et à la mortalité des usagers de la voie publique en cas de conflit avec un véhicule. Une telle situation, avec le manque croissant de prudence des piétons, constituent deux causes majeures de l'augmentation des accidents impliquant un piéton et un véhicule. Chaque année, en France, environ 800 piétons sont tués, dont 93% en milieu urbain. Parmi les piétons, deux sous populations sont plus particulièrement à risque : les enfants et les personnes âgées. Les premiers par leur insouciance et leur manque de connaissances spécifiques et les seconds par la réduction des capacités fonctionnelles, sensorielles et cognitives liées à l'âge modifiant significativement leur comportement. En effet, le vieillissement de la population, conduisant à une réduction progressive de la résistance aux chocs des organismes et affectant les comportements lors de la traversée de rue, contribuant à conduire la part des piétons âgés de 65 ans et plus dans la mortalité piétonne à passer de 26,5% en 2002 à 52% en 2015. Il relève ainsi d'un problème de santé public majeur de s'intéresser à l'origine possible de ces chiffres pour contribuer à leur baisse.

Récemment l'accidentologie a intégré les problématiques de la communauté de vision par ordinateur, compte tenu des progrès réalisés par cette dernière en termes de développement matériels et logiciels. Par ailleurs, l'implication des techniques de vision (détection, suivi, segmentation...) ainsi que de l'Intelligence Artificielle est un atout très promoteur pour l'analyse comportementale des piétons lors de la traversée d'une rue.

Nous démarquant de l'existant, nous proposons une nouvelle architecture de système d'analyse du comportement des piétons lors d'une traversée d'une rue. L'approche proposée se base sur le traitement automatique d'images et exploite les techniques et les avancées en vision par ordinateur. Le système proposé se compose de 2 couches principales interconnectées : une couche basse et une couche haute. La couche basse est centrée sur un certain nombre d'applications de techniques de vision par ordina-

teur à des fins de collecte de données du trafic, des piétons et de l'analyse automatisée de la séquence de traversée. Les capteurs vidéo utilisés sont des caméras HD grand public. Le principal avantage des techniques de vision par ordinateur est le potentiel de collecte des données microscopiques des usagers de la route et du trafic à un degré d'automatisation et avec une précision difficilement réalisable par des techniques d'imagerie classiques manuelles ou semi-automatiques et surtout de générer des indicateurs de conflit du trafic, sans l'intervention subjective de l'être humain. Pour cette couche basse nous avons proposé 3 sous systèmes de collecte de données. La couche haute se base sur un Réseau Bayésien, pour l'inférence du type de comportement du piéton, à chaque instant de sa traversée, par la corrélation des indicateurs de conflit. La collecte des données est fondée sur les capteurs de vision.

Les observations et les événements élémentaires obtenus sont de nature incertaine à cause des limitations des techniques de détection et de suivi. Pour pallier à cette limitation, nous proposons ici une nouvelle métrique de calcul d'un coefficient de dégrée de confiance associé à ces observations. Ce coefficient est calculé en fonction de la précision qualitative de la détection de l'objet cible, dit score et la consistance temporelle de cette même observation durant un petit laps de temps (1 seconde). Ensuite, ce degré de confiance est intégré dans la couche haute que ce soit lors de la phase d'apprentissage ou d'inférence à parti de la règle de 'Jeffry'.

L'outil proposé est finalement évalué sur plusieurs traversées caractéristiques et permet de quantifier de manière automatique et pour chaque instant, le niveau de prise de risque du piéton.

Table des matières

Table des figures **xiii**

Liste des tableaux **xvii**

Liste des abréviations **xix**

Introduction générale **1**

1 Revue de littérature **5**
1.1 Introduction 5
1.2 Définitions et description 6
1.2.1 Définitions 6
1.2.2 Les accidents corporels dans la littérature internationale 7
1.2.3 Chiffres en France 8
1.3 Traversée de rue et difficultés spécifiques du piéton âgé 10
1.3.1 Description d'un piéton âgé 10
1.3.2 Difficultés spécifiques du piéton âgé 11
1.4 Méthodes de collecte de données et d'étude du comportement piétonnier 13
1.4.1 Les méthodes manuelles 13
1.4.1.1 Questionnaires, entretiens et sondages 13
1.4.1.2 Observations et expérimentations in situ 14
1.4.2 Les méthodes automatiques 15
1.4.2.1 Les méthodes basées sur les capteurs actifs 15
1.4.2.2 Les simulateurs 17
1.4.2.3 Les méthodes basées sur les capteurs passifs : La vision par ordinateur 18
1.4.3 Conclusion sur les méthodes d'études du comportement piétonnier 21
1.5 Techniques de détection des conflits du trafic 22
1.5.1 Conflit du trafic 22

1.5.2 Les indicateurs de conflit de trafic et de danger lors de la traversée d'une rue . . . 22
1.5.3 Les mesures de sécurité et les stratégies d'analyse de conflits du trafic . . . 24
1.5.4 Conclusion sur les techniques d'études des conflits du trafic et prédiction du danger . . . 27
1.6 Synthèse de l'existant et identification des contributions potentielles : Objectifs et hypothèses de travail . . . 28
1.7 Conclusion . . . 30

2 Collecte automatique de données 'microscopiques' de piétons 31
2.1 Introduction . . . 31
2.2 Processus proposé : description générale . . . 32
2.3 Détection et suivi de personnes . . . 33
2.3.1 Approches existantes pour la détection de personnes . . . 34
2.3.1.1 Approches basées sur la texture et la couleur de la peau 34
2.3.1.2 Approches basées sur le mouvement . . . 34
2.3.1.3 Approches basées sur la forme . . . 35
2.3.2 Approches existantes pour le suivi de personnes . . . 36
2.3.3 Approche proposée : suivi multiple basé sur la maximisation de similarité de spatiogramme . . . 37
2.3.3.1 Détection continue : frame par frame . . . 38
2.3.3.1.1 Histogrammes de Gradient Orienté (HOG) . . . 38
2.3.3.1.2 Interpolation . . . 39
2.3.3.2 Suivi de personnes par l'association de détections avec utilisation d'une métrique de similarité de spatiogramme 41
2.3.3.2.1 Associations de détections . . . 41
2.3.3.2.2 Calcul de spatiogramme et mesure de similarité . . . 42
2.3.4 Evaluation des performances de l'approche proposée . . . 43
2.3.4.1 Présentation des bases de données . . . 43
2.3.4.2 Protocole de tests . . . 44
2.3.4.3 Résultats quantitatifs . . . 44
2.3.4.4 Résultats qualitatifs . . . 45
2.3.5 Correction des trajectoires par filtrage . . . 47
2.4 Modélisation de la zone de traversée (passage piéton) et détection de la position réelle des piétons . . . 48
2.4.1 Modèle de passage piéton . . . 49
2.4.2 Homographie . . . 50
2.5 Données collectées pour la caractérisation du comportement du piéton . 52
2.5.1 Statistiques sur la cinématique du piéton . . . 52
2.5.2 Déplacements des piétons . . . 53
2.5.3 Estimation des vitesses instantanées et moyennes des piétons . . 54
2.5.4 Estimation des phases de traversée . . . 54
2.5.5 Estimation du temps d'attente . . . 55
2.5.6 Evaluation de la précision des données calculées . . . 55

2.5.6.1 Etude de cas et base d'évaluation 55
2.5.6.2 Résultats expérimentaux 56
2.5.6.2.1 Evaluation de la précision d'estimation des phases de traversée 56
2.5.6.2.2 Évaluation de la précision de l'estimation de la vitesse moyenne 57
2.5.6.2.3 Évaluation de la précision de l'estimation du temps d'attente 60
2.5.6.3 Synthèse et étude comparative 60
2.6 Conclusion 61

3 Estimation automatique de la classe d'âge du piéton à partir de la démarche 62
3.1 Introduction 62
3.2 Processus proposé : description générale 62
3.3 Les approches existantes pour la reconnaissance de l'âge 63
3.3.1 Les approches cranio-faciales 64
3.3.2 Les approches utilisant la démarche 65
3.3.2.1 Les approches basées sur la forme de la silhouette . . . 66
3.3.2.2 Les approches basées sur le contour 69
3.4 Approche proposée pour la reconnaissance de la classe d'âge 70
3.4.1 Processus de pré-traitement 70
3.4.2 Détection des cycles de marche 71
3.4.3 Descripteur proposé : modèle de la silhouette 73
3.4.3.1 Projection transversale de la silhouette 73
3.4.3.2 Projection longitudinale de la silhouette 74
3.4.3.3 Conception du modèle de la silhouette 76
3.5 Évaluation des performances 77
3.5.1 Résultats qualitatifs : Une analyse morphologique de la démarche 77
3.5.1.1 Les vecteurs moyens de SLP et STP 77
3.5.1.2 Les variances de SLP et STP 77
3.5.1.3 Le modèle de Frieze de SLP et STP 79
3.5.2 Résultats quantitatifs 80
3.5.2.1 Base utilisée (corpus de travail) 81
3.5.2.2 Protocole de test 81
3.5.2.3 Critères de performances 82
3.5.2.4 Performances du descripteur proposé 83
3.5.2.5 Robustesse du descripteur proposé vis à vis du genre . 84
3.5.3 Étude comparative avec les approches existantes 84
3.6 Conclusion 89

4 Collecte automatique de données 'microscopiques' du trafic routier 90
4.1 Introduction 90
4.2 Processus proposé : description générale 91
4.3 Détection et suivi des véhicules 92
4.3.1 Approches existantes pour la détection et le suivi des véhicules . 92

4.3.1.1 A partir de radars et de capteurs 92
4.3.1.2 A partir des caméras 92
4.3.1.2.1 Approches basées sur le mouvement 93
4.3.1.2.2 Approches basées sur l'apparence 94
4.3.2 Approche proposée : Optimisation de l'Histogramme de Gradient Orienté 95
4.3.2.1 Les paramètres du mouvement 96
4.3.2.2 Les paramètres d'apparence 97
4.3.3 Evaluation des performances 98
4.3.3.1 Cas de détection statique 99
4.3.3.1.1 Analyse et visualisation de la représentation des caractéristiques d'une voiture par l'HOG 100
4.3.3.1.2 Performances en termes de reconnaissance des véhicules 101
4.3.3.2 Cas de détection dynamique : performances en termes d'optimisation du temps de calcul 104
4.3.3.2.1 Base dynamique utilisée 104
4.3.3.2.2 Résultats 104
4.3.3.2.3 Étude comparative 106
4.4 Modélisation de la zone de traversée (passage piéton) et détection des positions réelles des véhicules 107
4.4.1 Modèle de passage piéton 108
4.4.2 Homographie 108
4.5 Données collectées pour la caractérisation du trafic 108
4.5.1 Estimation des trajectoires et des positions des véhicules 110
4.5.2 Estimation des vitesses moyennes 110
4.5.3 Estimation des directions 110
4.5.4 Evaluation de la précision des estimations 112
4.5.4.1 Etude de cas et base d'évaluation 112
4.5.4.2 Résultats expérimentaux 112
4.5.4.2.1 Evaluation de l'estimation des trajectoires et des positions 112
4.5.4.2.2 Evaluation de l'estimation des vitesses moyennes . . . 113
4.6 Conclusion 114

5 Approche probabiliste pour l'estimation des risques pris par les piétons 116
5.1 Introduction 116
5.2 Processus proposé : description générale 117
5.3 Approches existantes d'analyse comportementale et reconnaissance de scénarios 118
5.3.1 Interprétation haut niveau des séquences d'images 118
5.3.2 Principaux travaux existants 120
5.4 Approche proposée 123
5.4.1 Architecture globale 123
5.4.2 Reconnaissance des événements et choix des variables 126
5.4.2.1 Agents pour la reconnaissance d'événement élémentaire 126

5.4.2.2 Proposition des variables observées 127
5.4.2.3 Proposition des agents à hypothèses 129
5.4.3 Estimation des paramètres du réseau bayésien 132
5.4.3.1 Apprentissage avec données complètes 132
5.4.3.2 Gestion des imperfections de la perception et apprentissages avec données incomplètes 133
5.4.3.3 Proposition d'une technique de gestion de l'incertitude et degré de confiance pour les observations des capteurs de vision 133
5.4.3.3.1 Consistance temporelle 134
5.4.3.3.2 Score de détection 134
5.4.3.3.3 Injection du degré de confiance 135
5.5 Évaluation des performances du modèle proposé pour l'estimation des risques 137
5.5.1 Protocole de test 137
5.5.2 Résultats 138
5.5.2.1 Traversée d'une rue pour les piétons jeunes 138
5.5.2.2 Traversée d'une rue pour les piétons âgés 140
5.5.2.3 Discussion 144
5.6 Conclusion 145

Conclusion et perspectives **146**

Table des figures

1.1 Répartition des décès dus aux accidents de la route par type d'usager et par zone géographique selon l'OMS [16]. 7
1.2 Mortalité rapportée à la population par million d'habitants, selon l'âge France 2015 - Europe 2014 [17]. 8
1.3 Évolution de la mortalité routière annuelle en métropole [18]. 9
1.4 Évolution de la mortalité au niveau de piétons *(Sources : Rapport ONISR Année 2002, 2009, 2011 et 2016)*. 9
1.5 Représentation du danger ressenti par les personnes de plus de 60 ans [27]. 11
1.6 Caption for LOF 12
1.7 Mortalité des séniors de 65 ans et plus [18]. 12
1.8 Variations de probabilité de traverser en fonction du TTC et de la vitesse du véhicule [46]. 17
1.9 Contribution des différents paramètres dans la prédiction de décision de traverser [46]. 17
1.10 Exemple de situation expérimentale. 18
1.11 Illustration des différents scénarios de traversée de rue [58]. 19
1.12 Présentation de quelques positions : les quatre premières sont correctes et les quatre dernières ne sont pas correctes [58]. 20
1.13 Système de protection des piétons [70]. 26

2.1 Architecture générale de l'outil automatique de collecte de données piétonnières proposé. 32
2.2 Modèles morphologiques de la personne. 35
2.3 Illustration de la discrimination entre une personne et un véhicule [94]. 36
2.4 Illustration des étapes d'extraction des Histogrammes de Gradient Orientés : cas de la détection d'une personne [89]. 39
2.5 Interpolation des détections manquantes. 40
2.6 Illustrations de détections à partir du processus HOG + interpolation. . 41
2.7 Association des détections [111]. 42

2.8 Illustration de résultats de détection de personnes : a) Base CAVIAR, b) Base CWV séquence 1 et c) Base CWV séquence 2 [111]. 46
2.9 Résultats du filtrage des trajectoires : a) Avant filtrage, et b) Après filtrage [111]. 48
2.10 Le modèle de calibration au niveau du passage piéton. 49
2.11 Points de correspondances entre le plan image et le modèle de passage piéton. 50
2.12 Exemples de transformations de trajectoires des coordonnées images en coordonnées réelles : a) Trajectoires sur plan image, b) Trajectoires projetées sur le modèle de passage piéton. 51
2.13 Trajectoires d'un piéton traversant une route, hors (à gauche) et sur (à droite) un passage piéton [13]. 53
2.14 Estimation des phases de traversée par l'analyse des positions spatiales du piéton. 56
2.15 Analyse de la variation de la vitesse instantanée pour la détection des phases de traversée de la rue. 58
2.16 Illustration des corrélations entre les vitesses moyennes estimées et les vitesses de réalités de terrain, a) Méthode M1, b) Méthode M2. 59

3.1 Processus proposé pour la reconnaissance de la classe d'âge. 63
3.2 Evolution de la morphologie du crâne chez l'être humain avec l'âge [127] : Le crâne qui se trouve au centre représente celui d'un enfant, celui qui se trouve à l'extérieur représente le crâne d'un homme mature. 64
3.3 Représentation fréquentielle de la GEI 66
3.4 Diagramme du processus de régression gaussienne [122]. 67
3.5 Diagramme du processus de pré-traitement proposé. 70
3.6 Exemple de silhouette nettoyée par l'application des opérateurs morphologiques : a) Silhouette bruitée, b) Silhouette nettoyée. 71
3.7 Diagramme du cycle de marche [152]. 71
3.8 Détection des positions des chevilles [153]. 72
3.9 Variation de la longueur du pas et détection d'un cycle. 72
3.10 Projection transversale de la silhouette. 73
3.11 Variation transversale d'une silhouette durant un cycle de marche. . . . 74
3.12 Projection longitudinale de la silhouette. 75
3.13 Variation longitudinale d'une silhouette durant un cycle de marche. . . 75
3.14 Modèle de la silhouette. 76
3.15 SLP d'un sujet jeune (bleu) et d'un âgé (rouge). 77
3.16 STP d'un sujet jeune (bleu) et d'une âgé (rouge). 78
3.17 Superposition des courbes de variances de SLP de la silhouette d'un sujet jeune (bleu) et d'un âgé (rouge). 79
3.18 Superposition des courbes de variances de STP de la silhouette d'un sujet jeune (bleu) et d'un sujet âgé (rouge). 79
3.19 Modèles de Frieze de SLP et STP des sujets jeunes et âgés. 80
3.20 Échantillon de données de la base OU-ISIR [154]. 81
3.21 La courbe ROC du descripteur proposé (SM). 83

3.22 Image moyenne d'une silhouette durant un cycle de marche : La lettre $'A'$ désigne la partie statique alors que la lettre $'B'$ représente la partie dynamique. 85
3.23 Sélection des points de contour 86
3.24 Comparison des performances du descripteur proposé (SM) par rapport à ceux existants (GEI et FED) en termes des 'rappels' et 'précisions' pour les deux classes jeune (J) et âgé (A). 87
3.25 Superposition des courbes ROC du descripteur proposé (SM) et de ceux existants (GEI et FED). 88

4.1 Présentation générale du processus proposé pour la détection des véhicules. 92
4.2 Les méthodes de détection des véhicules basées sur la vision par ordinateur [164]. 93
4.3 Illustration d'extraction des objets en mouvement : a) Image initiale, b) Image de l'avant plan initialement obtenue, c) Image de l'avant plan post-traitée, d) Résultat d'extraction d'objets en mouvement et e) Détection du véhicule. 96
4.4 Calcul des gradients horizontal et vertical d'un véhicule. 97
4.5 Échantillons de la base de données UIUC : a) Exemple positif et b) Exemple négatif. 99
4.6 Échantillons de la base de données TUD : a) Exemple positif et b) Exemple négatif. 99
4.7 Visualisation de voitures par l'HOG : a) Exemples des images aux niveaux de gris et b) Exemples des images en RVB. 100
4.8 Visualisation d'objets autres qu'une voiture par l'HOG : a) Images initiales et b) HOG. 101
4.9 Illustration de la similarité de l'HOG d'une voiture. 101
4.10 Illustration de la détection de plusieurs véhicules dans une scène statique.102
4.11 Illustration de la détection d'un seul véhicule dans une scène statique. . 103
4.12 Echantillons de la base dynamique. 104
4.13 Illustrations de la détection de véhicules : cas dynamique. a) monodétection, b) multi-détection. 105
4.14 Exemple de transformation de trajectoires des coordonnées images en coordonnées réelles pour le trafic gauche et droit : a) Trajectoires sur le plan image, b) Trajectoires projetées sur le modèle de passage piéton. . 109
4.15 Illustration de l'identification des directions : a) Exemple de directions d'un piéton et d'un véhicule, b) Exemple de direction de deux véhicules pour le trafic gauche et droit. 111
4.16 Superposition des trajectoires estimées et de celles de réalité du terrain. 113

5.1 Modèle générale de framework proposée pour la reconnaissance du comportement des piétons. 118
5.2 Processus de traitement haut niveau de l'image [195]. 119
5.3 Approche pour un système sûr proposé par [219]. 122
5.4 Structure du réseau bayésien proposée. 125
5.5 Étapes de construction d'un réseau bayésien [223]. 125

5.6 Discrétisation des observations des capteurs de vision concernant les variables du piéton. 128
5.7 Discrétisation des observations des capteurs de vision concernant la classe d'âge : a) Cas d'un piéton jeune, b) Cas de piétons âgées. 129
5.8 Discrétisation des observations des capteurs de vision concernant les variables du trafic : a) Trafic droit, b) Trafic gauche. 130
5.9 Discrétisation des agents à hypothèses en fonction des variables observées. 131
5.10 Illustration de l'architecture du modèle décisionnel proposé. 132
5.11 Estimation du comportement lors du scénario 2 de traversée d'une rue : cas d'un sujet jeune sur un passage piéton sans feu. 139
5.12 Illustration des reconnaissances des événements élémentaires et estimation du type de comportement : cas d'un sujet jeune sur un passage piétons sans feu (S2). 140
5.13 Estimation du comportement lors du scénario 3 de traversée d'une rue : cas d'un sujet jeune sur un passage piéton avec feu. 141
5.14 Estimation du comportement lors du scénario 4 de traversée d'une rue : cas d'un sujet âgé sur un passage piéton sans feu. 142
5.15 Estimation du comportement lors du scénario 5 de traversée d'une rue : cas d'un sujet âgé sur un passage piéton avec feu. 143

Liste des tableaux

1.1 Evolution de l'accidentalité routière entre 2015 et 2016 [18] 9
1.2 Comparaison entre les stratégies de protection [44]. 16
1.3 Les différents indicateurs de conflit de trafic [54] 23
1.4 Comparaison entre les méthodes de protection des piétons [13] 25

2.1 Résultats quantitatifs du processus proposé pour le suivi de personnes en termes de MODA et MOTA [111]. 45
2.2 Évaluation des performances du filtrage des trajectoires. 48
2.3 Données caractéristiques pour une traversée de passage piéton avec un feu tricolore (moyennes ± écarts types) [117]. 52
2.4 Vitesse des piétons en fonction de leur activité [108]. 52
2.5 Vitesses moyennes respectives des hommes et des femmes, population allemande [108]. 52
2.6 Résultats de validation de la vitesse moyenne en utilisant les deux méthodes proposées pour le calcul des distances. 59
2.7 Précision d'estimation du temps d'attente en fonction du R^2 et de la $RMSE$. 60
2.8 Comparaison de la précision de l'estimation de la vitesse moyenne des piétons en utilisant les techniques de vision par ordinateur (suivi de piétons et analyse de trajectoires). 61

3.1 Performances du descripteur proposé (SM). 83
3.2 Performances de descripteur proposé 'SM' vis-à-vis du genre 84
3.3 Performances du descripteur (GEI) 85
3.4 Performances du descripteur (FED). 87
3.5 Performances des différents descripteurs : étude comparative. 87

4.1 Taux de reconnaissance d'un véhicule unique 102
4.2 Taux de reconnaissance multi-véhicules 102

4.3 Evaluation des performances de l'approche proposée en termes de taux et de temps de détection : cas dynamique. 106
4.4 Etude comparative des performances de notre approche en termes de taux et de temps de détection. 107
4.5 Erreur moyenne et coefficient de corrélation entre les positions estimées des véhicules et celles réelles. 112
4.6 Evaluation de l'estimation des vitesses moyennes avec la méthode M1. . 114
4.7 Evaluation de l'estimation des vitesses moyennes avec la méthode M2. . 114
4.8 Erreur moyenne pour l'estimation des vitesses des véhicules pour le trafic gauche et droit 114

5.1 Caractéristiques des séquences de test. 137
5.2 Résultats des prédictions générées par l'approche probabiliste proposée pour la séquence S2. 139
5.3 Résultats des prédictions générées par notre approche probabiliste pour la séquence S3. 141
5.4 Résultats des prédictions générées par l'approche probabiliste pour la séquence S4. 142
5.5 Résultats des prédictions générées par l'approche probabiliste pour la séquence S5. 143

Liste des abréviations

AAM	Modèle d'apparence Actif (Active Appearance Model)
ADAS	Systèmes Avancés d'Assistance au Conducteur (Advanced Driver Assistance Systems)
ASM	Modèle de Forme Active (Active Shape Model)
C	Consistance temporelle
CAVIAR	Context Aware Vision using Image-based Active Recognition
CWV	CrossWalk Valenciennes
D_c	Dégré de Confiance
FED	Feature to Exemplar Distance
FO	Nombre de sujets âgés mal classés (False Old)
FPA	Une fausse alarme positive (False Positif Alarm)
FPGA	Carte électronique programmable (Field Programmable Gate Array)
FREQ	Représentation FREQuentielle
FY	Nombre de sujets jeunes mal classés (False Young)
GEI	Image moyenne (Gait Enregie Image)
GP	Période de marche (Gait period)
GPR	Régression Gaussienne (Gaussian Process Regression)
GPU	Processeur Graphique (graphic process unit)
HMM	Modèle de Markov caché (Hidden Markov Model)
HOG	Histogram de Gradients Orientés
HSV ou TSL	Teinte, Saturation, Luminosité (Hue, Saturation, Lightness)
IA	Intelligence Artificielle
ITS	Système de Transport Intelligent

LBP	Modèle Binaire Local (Local Binary Pattern)
LPQ	Quantification locale de phase (Local Phase Quantization)
MAE	Erreur d'âge Moyenne (Mean Age Error)
MODA	Multiple Object Detection Accuracy
MOTA	Multi-Object Tracking Accuracy
MOTP	Multiple Object Tracking Precision
MV	Maximum de Vraisemblance
P-V	Piéton-Véhicule
R	Coefficient de corrélation
RB	Réseau Bayésien
RMSE	Erreur moyenne (Root Mean Squared Error)
S	Score cumulatif
SLP	Projection Longitudinale de la Silhouette
SM	Modèle de la Silhouette
STP	Projection Transversale de la silhouette
SVM	Machine à vecteur de support (Support Vector Machine)
TO	Nombre de sujets âgés correctement classés (True Old)
TPA	Véritable alarme positive (True Positif Alarm)
TR	Taux de Reconnaissance
TY	Nombre de sujets jeunes correctement classés (True Young)

Introduction générale

La marche est un mode de déplacement de base à travers le monde. En effet, nous sommes tous des piétons, on se déplace presque toujours à pied pour débuter ou terminer un trajet. Tandis que la marche est un mode de déplacement dit 'doux' [1], économique et ayant de bons effets sur la santé, elle est parmi les causes majeurs des traumatismes sur la route. Même si le déplacement à pied semble être simple et naturel pour les êtres humains ; pour traverser la rue de façon autonome et sécuritaire, les piétons doivent posséder trois aptitudes essentielles généralement acquises entre 9 et 11 ans : i) La capacité de déterminer et d'emprunter un parcours sécuritaire, ii) la faculté d'évaluer de façon réaliste la vitesse d'un véhicule et iii) les fonctions cognitives nécessaires pour juger l'intervalle de sécurité suffisant dans la circulation [2].

La réduction du nombre de victimes d'accident de la route chez les piétons est nettement moins marquée que celle observée chez les passagers des véhicules motorisés. En effet, entre 2010 et 2016, la fatalité des accidents sur les routes françaises a baissé jusqu'à -13%, mais si cela représente 37 automobilistes tués en moins soit une baisse de -17% par rapport à l'année 2010, celle des piétons est en hausse de 11%. Proportionnellement, la part des piétons tués a augmenté et ce, régulièrement, passant de 12% en 2010, à 14% en 2015 et à 16% en 2016, soit respectivement 468, 485 et 540 tués (ONISR, 2010, 2015, 2016). Les personnes âgées semblent être la catégorie la plus affectée par les accidents routiers. De plus, les personnes âgées ont tendance à marcher en conservant le regard fixé vers le sol afin d'améliorer leur équilibre, ce qui les empêchent de faire attention aux véhicules ou aux feux de circulation. Les personnes âgées de plus de 65 ans représentent 52% des piétons tués (ONIRS, 2015) alors qu'elles ne sont que 24% de la population française (INSEE, 2015). En fait, le facteur de risque de décès pour un sujet âgé de plus 65 ans est multiplié par 2 cependant celui pour les âgés de plus de 75 ans est multiplié par 5, puisque ces derniers ne représentent que 9% de la population.

La demande en matière de sécurité augmente continuellement afin de lutter contre les accidents quotidiens qui touchent les différents secteurs, en particulier la sécurité routière. Par ailleurs, les administrations ont élevé la sécurité routière au rang de priorité et plusieurs politiques de sécurité ont été mises en place. L'état encourage par

ailleurs les lois, les recherches et les études d'accidentologie par la mise en œuvre et le financement de grands projets sur la sécurité routière tel-que le projet PISTES (PIétonS et comportement de TraverséE de rueS) financé par la Fondation de Sécurité Française.

L'accidentologie est l'étude des accidents. C'est une discipline faisant appel à un spectre de compétences étendu : de l'ingénieur au médecin, en passant par le psychologue, l'expert en mécanique ou en sciences cognitives [3]. Il s'agit d'une approche pluridisplinaire pour obtenir une vision globale et exhaustive des accidents. L'évolution de l'accidentologie est retracée dans [4]. Cette discipline est très étroitement liée à la sécurité routière : son but est de fournir des données reflétant de façon objective l'incidence d'un accident et les facteurs qui le causent. Après des débuts laborieux et empiriques dans la première moitié du XXe siècle, elle est restée quelques décennies sous une forme "primitive" [5].

L'accidentologie s'est développée et passe dans sa phase "moderne" dans les années 1970. Elle commence alors à se différencier de la simple spécialité médicale qui traite les traumatismes. Un accident est alors étudié de manière plus fine et plus nuancée. Les chercheurs de disciplines et accidentologues de terrain tentent désormais de le comprendre dans sa globalité, en tenant compte de sa grande complexité [3].

En effet, les sciences cognitives et la vision par ordinateur sont deux sources prometteuses pour une telle analyse, fine et nuancée tenant compte de la complexité de l'accidentologie. La vision par ordinateur tire profit des progrès technologiques, qui ont largement contribué à améliorer les performances des microprocesseurs, surtout en matière de vitesse de traitement des données ainsi que les performances du matériel d'acquisition tel que celles des caméras à haute définition (HD).

Dans ce contexte théorique, l'objectif de ce travail de thèse est la modélisation d'un outil automatique d'analyse comportementale des piétons âgés lors de la traversée d'une rue, à base de traitement d'images. Nous nous intéressons plus particulièrement à l'interprétation du contenu des images et des séquences vidéos à des fins d'analyse des conflits du trafic et du comportement des piétons lors de traversée de la rue. Deux niveaux d'interprétation d'une image sont définis : symbolique, ou bas niveau, et sémantique, ou haut niveau. L'objectif de la recherche étant d'obtenir une interprétation de l'image au niveau sémantique, afin de communiquer des résultats fidèles aux attentes des utilisateurs humains d'un tel système fondé sur l'analyse symbolique.

Sur le plan pratique, la détermination d'un mécanisme d'analyse comportementale de piétons à base de vision par ordinateur s'affronte à plusieurs enjeux qui consistent à mettre en œuvre un descripteur indifférent aux changements qu'un environnement peut subir, dans ce cas : l'angle de prise de vue (vue frontale, latérale ...), la variation de l'éclairage etc... Ou bien les variations qui touchent les objets cibles eux-même tels que : la vitesse de déplacement, les vêtements, l'occultation totale ou partielle, ainsi que l'aspect aléatoire et adaptatif du comportement de l'être humain, etc ...

En effet, cette modélisation du problème des conflits du trafic se base alors sur le concept de l'extraction automatisée des informations pertinentes sur les objets présents dans des images et l'interprétation sémantique des séquences vidéo. Cette thématique de recherche a un intérêt grandissant surtout avec les progrès des techniques de trai-

tement bas niveau (vision par ordinateur) mais aussi de haut niveau (intelligence artificielle).

La relation cause-effet entre les observations et le comportement des piétons invite à l'utilisation d'un Réseau Bayésien (RB) pour mener ces analyses surtout que ce dernier présente la possibilité de rassembler et de fusionner des connaissances de diverses natures dans un même modèle avec une structure explicite et intuitive faisant de ce réseau un modèle adéquat. Même si le comportement du piéton est un comportement relativement aléatoire, adaptatif et évolutif au cours du temps, une décision bayésienne peut tout à fait le caractériser.

Les contributions de la thèse

La lutte contre les accidents routiers chez les personnes âgées réside dans la résolution du problème d'une analyse fiable et robuste du comportement des piétons en relation avec l'environnement, le trafic routier ainsi que les aménagements urbains en prenant en considération l'effet du vieillissement. En se basant sur ces hypothèses les contributions de cette thèse se situent à plusieurs niveaux :

— L'amélioration des techniques de vision par ordinateur pour l'étude de problème des conflits du trafic et l'automatisation de l'étude accidentologique, vue que la majorité des travaux réalisés dans ce contexte sont soit manuels, soit semi-automatiques.

— La proposition de processus automatiques pour le multi-suivi de personnes et de véhicules basés sur les Histogrammes de Gradients Orientés améliorés. Ce processus est utilisé pour la collecte des paramètres piétonniers et de trafic tels que la vitesse, la direction, le temps d'attente, les positions spatio-temporelles.... Ces paramètres seront utilisés comme indicateurs de conflit du trafic afin de caractériser par la suite le comportement des piétons et inférer le risque de la situation de traversée.

— La proposition d'un système basé sur la démarche pour la reconnaissance de l'âge des individus lors de la traversée d'une rue pour distinguer les sujets séniors des sujets plus jeunes et mettre en œuvre l'impact de l'âge sur le type de comportement.

— La dernière contribution de ce travail est au niveau du processus de reconnaissance de scénarios. En effet, l'approche développée dans le cadre de l'utilisation d'un réseau bayésien propose une adaptation efficace du coefficient de degré de confiance afin de manipuler facilement les informations incertaines des capteurs de vision voir des informations manquantes. Dans ce projet de thèse, la faisabilité de l'approche est illustrée par le problème de la reconnaissance du comportement à risque dans une situation de traversée d'une rue.

Organisation du manuscrit

Ce présent mémoire est organisé comme suit :

— le premier chapitre présente une revue de la littérature sur les études d'accidentologie et d'analyse des conflits du trafic.

— Le deuxième chapitre introduit le processus de collecte des paramètres piéton-

niers, leurs évaluation et validation à partir d'une base de données vidéo constituée dans le cadre de ce travail.

— Dans le chapitre 3 nous détaillons la phase d'estimation de la classe d'age des piétons et nous décrivons notre descripteur basé sur la démarche du piéton.

— La stratégie de collecte des paramètres du trafic ainsi que les détails de ces derniers sont présentés dans le chapitre 4. Une évaluation des performances de cette stratégie sur une base d'images de traversée d'une rue réelle est également présentée durant ce chapitre.

— Le chapitre 5 est consacré à la présentation du système complet d'analyse comportemental basé sur la corrélation des paramètres piétonniers et du trafic ainsi collectés. Ce système se base sur la mise en relation d'un traitement bas niveau (phase de collecte des données) et un traitement haut niveau (phase d'analyse sémantique et inférence du type de comportement des pitons) des séquences vidéo en exploitant une approche probabiliste basée sur un RB.

Chapitre

1 Revue de littérature

1.1 Introduction

Plus d'un cinquième des personnes tuées sur les routes chaque année dans le monde ne sont ni des automobilistes, ni des motocyclistes, ni même des cyclistes, mais des piétons. Pourtant, les accidents et les décès de piétons sont souvent évitables moyennant des interventions qui ont fait la preuve de leur efficacité. Cependant, dans beaucoup d'endroits, la sécurité des piétons ne fait pas l'objet de toute l'attention qu'elle mérite [6].

Pour intervenir efficacement afin de protéger et de promouvoir la sécurité des piétons, il faut bien comprendre la nature des facteurs de risque d'accident et les potentielles stratégies de mise en œuvre d'un système de sécurité routière. Différents projets technologiques et scientifiques (Systèmes de Transport Intelligents (ITS), Systèmes Avancés d'Assistance au Conducteur (ADAS), réalité projetée, des analyses statistiques simples, apprentissage,...) tentent de contribuer à ce problème.

Ce chapitre à pour objectif de donner une idée générale sur les accidents impliquant des piétons et les facteurs de risque au niveau international et en particulier en France (cas d'étude dans cette thèse). Nous mettrons l'accent également sur les approches et les méthodes de collecte ainsi que d'analyse du comportement du piétons lors de traversée d'une rue. Les indicateurs de conflit entre un piéton et des véhicules ainsi que les stratégies de protection (prédiction, détection) des usagers de la route sont également présentées. Ce chapitre se clôturera par une synthèse de l'existant et une présentation des hypothèse et des objectifs généraux de cette thèse.

1.2 Définitions et description

1.2.1 Définitions

- Le piéton

Un piéton est une personne ayant comme mode de déplacement le plus souvent la marche. Il s'agit d'un mode de transport dit "doux" car il est d'un faible coût économique et ne génère pas d'émissions polluantes [1]. La marche est une mobilité durable ou dite aussi "écomobilité" [7] car elle ne consomme pas d'énergie fossile mais se base sur la force du corps.

La marche à pieds est vivement encouragée et réputée durant les années 2000. Ainsi, représentant 22% des déplacements, elle se révèle davantage pratiquée dans les grandes agglomérations en 2008 que 14 ans auparavant [8]. En fait, le déplacement à pied est d'abord "un phénomène urbain" : le cas de la ville de Nice, par exemple, montre que les déplacements à pied sont légèrement plus nombreux que ceux effectués en voiture (49,3% et 47,7% respectivement), et les personnes âgés de plus de 75 ans représentent 66,3% des piétons [9].

- Accident

Lors de l'execution de ses taches de la vie quotidienne la personne est exposée à nombreux risques d'accidents tels que la chute, les brûlures ou encore un accident de la route. Ces artéfacts peuvent être tous décrits par le terme générique "accident" [10]. Il est important alors de définir et différencier les terminologies suivantes :

- Accident de la circulation

Un accident de la circulation est : "une collision ou un incident impliquant au moins un véhicule routier en mouvement, sur une voie de circulation publique ou sur une voie de circulation privée à laquelle le public a le droit d'accéder, et qui fait au moins un blessé ou un mort." Cette définition concerne les collisions entre véhicules routiers, entre véhicules routiers et piétons, entre véhicules routiers et animaux ou obstacles fixes, ainsi que les accidents n'impliquant qu'un seul véhicule [11].

- Accident véhicule-piéton

Un piéton accidenté est "toute personne, autre qu'un conducteur impliquée dans un accident corporel [12]." Un conducteur accidenté est "toute personne impliquée dans un accident corporel et qui conduisait un véhicule routier au moment de l'accident."

- Gravité d'un accident

Quatre ordres de gravité sont utilisés mondialement pour décrire les conséquences d'un accident [12] :

— Accident matériel : tout accident n'entraînant que des dommages matériels [13].
— Accident corporel : tout accident dans lequel au moins une personne est tuée ou blessée [13].
— Accident mortel : tout accident corporel entraînant la mort d'une personne. Un accident est dit mortel si la victime décède sur le coup ou des suites de l'accident, dans les 30 jours [12].

— Accident non mortel : tout accident corporel autre qu'un accident mortel. Il y a alors au moins une personne blessée. Deux cas se présentent [13] :

1. Personne gravement blessée : hospitalisation supérieure à 24 heures,
2. Personne légèrement blessée : personne blessée, mais non gravement blessée.

1.2.2 Les accidents corporels dans la littérature internationale

En se basant sur les statistiques d'accidentologie actuelles, à l'horizon 2030 les accidents de la route passeront au cinquième rang des causes de décès si aucune mesure n'est prise très rapidement. En fait, les traumatismes dus aux accidents de la circulation représentent la huitième cause de décès dans le monde et la première cause de décès chez les jeunes âgés de 15 à 29 ans [14].

Environ 2 millions d'accidents et 6 700 décès surviennent aux intersections chaque année aux États-Unis, ce qui représente 26% de toutes les collisions. En Inde, les piétons représente 62% de la mortalité par les accidents routiers [15]. Les traumatismes de la route impliquent 39% des piétons dans l'Union Européenne [16, 17].

Généralement, "les usagers de la route vulnérables" qui sont les piétons, les cyclistes, les motocyclistes représentent la moitié des décès dus aux accidents de la circulation enregistrés dans le monde. Les parts de ces usagers dans la mortalité routière sont respectivement 22 % pour les piétons, 5 % des cyclistes et 23 % des motocyclistes [16].

Le risque de mourir d'un accident de la route est le plus élevé dans le continent africain et le plus faible dans le continent européenne (Figure 1.1).

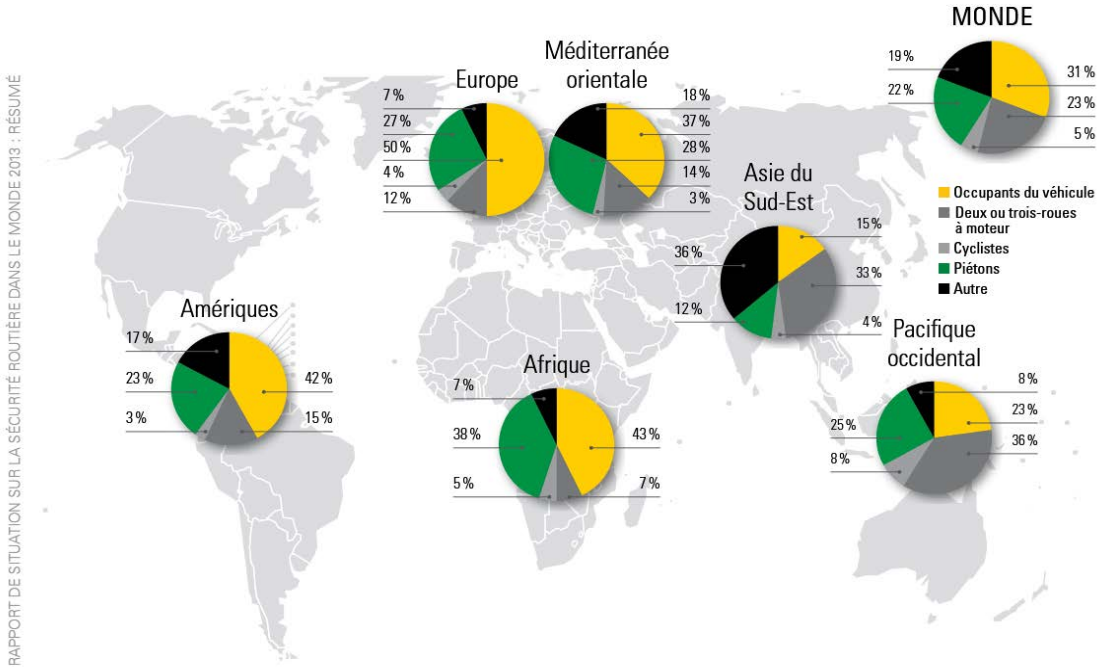


FIGURE 1.1 – Répartition des décès dus aux accidents de la route par type d'usager et par zone géographique selon l'OMS [16].

Néanmoins, on constate d'importantes disparités entre les pays d'une même région, les plus marquantes étant relevées en Europe. En effet, la mortalité rapportée pour les piétons en France est très proche des valeurs moyennes pour l'Europe, voir même la dépasse pour certains catégories d'âges (Figure 1.2) [16]. De ce fait, mettre l'accent sur les chiffres de mortalités en France, se révèle une tache très importante.

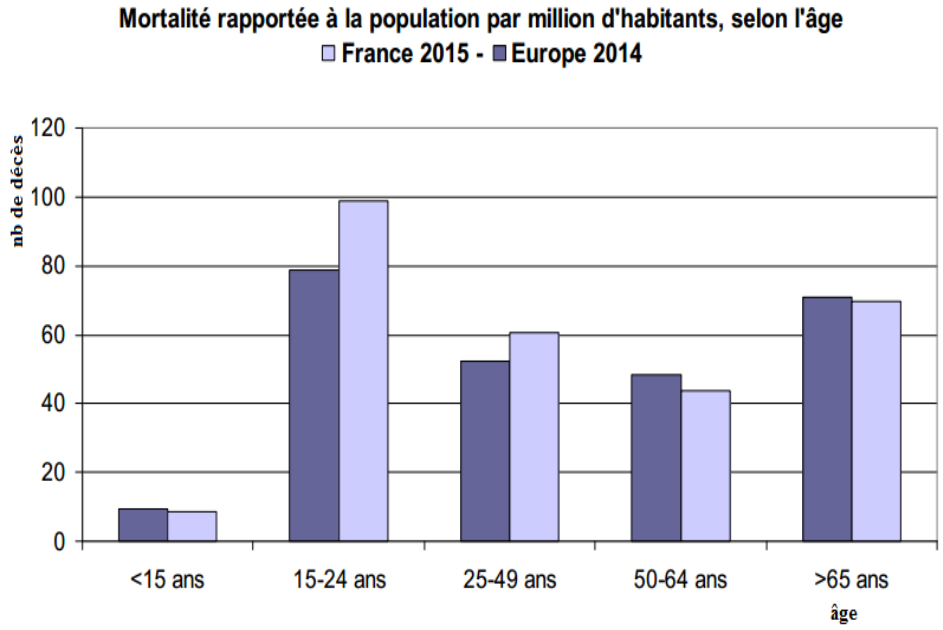


FIGURE 1.2 – Mortalité rapportée à la population par million d'habitants, selon l'âge France 2015 - Europe 2014 [17].

1.2.3 Chiffres en France

Les données sur les accidents de la circulation présentent de grandes similarités entre les pays à niveau de développement comparables. Le cas de la France sera pris comme exemple principal pour illustrer l'état des connaissances sur les accidents véhicule-piéton.

En se référant au bilan des accidents routiers de l'année 2016 (Figure 1.3), délivré par l'Observatoire National Interministériel de Sécurité Routière (ONISR) on constate que la mortalité routière annuelle en métropole ne suit pas l'objectif de diminution des fatalités des accidents fixé pour les années 2010/2020. Par ailleurs, une augmentation est marquée à partir de l'année 2013 [18].

Le tableau 1.1 met l'accent sur des statistiques très récentes et porte sur l' évolution de l'accidentalité routière entre 2015 et 2016. On note, notamment une augmentation de 1,1% des accidents corporels et que la mortalité routière est en augmentation de 0,2% entre 2015 et 2016, soit 8 décès de plus.

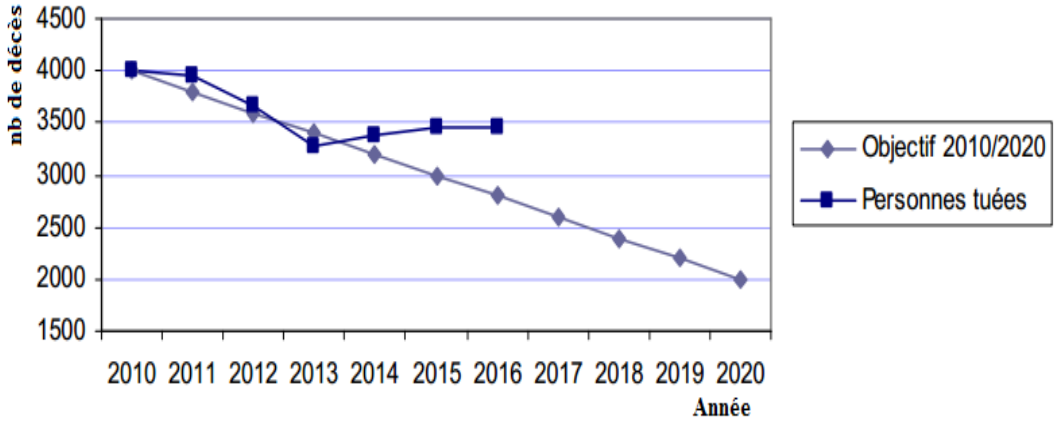


FIGURE 1.3 – Évolution de la mortalité routière annuelle en métropole [18].

TABLE 1.1 – Evolution de l'accidentalité routière entre 2015 et 2016 [18]

Bilan de l'année 2016	Accidents corporels	Tués à 30 jours	Blessés	Blessés hospitalisés
Année 2016 estimations	57 251	3 469	72 199	27 214
Année 2015	56 603	3 461	70 802	26 595
Différence 2016/2015	648	8	1 397	619
Evolution 2016/2015	**1,1%**	**0,2%**	**2,0%**	**2,3%**

D'après les statistiques de l'ONISR, durant plus de 10 ans en France, on note une augmentation de 5% des piétons tués sur les routes (Figure 1.4). Il est à noter aussi que 16% des accidents de la route en 2016 impliquent au moins un piéton. Ce pourcentage est en augmentation de 3% par rapport à l'année 2011.

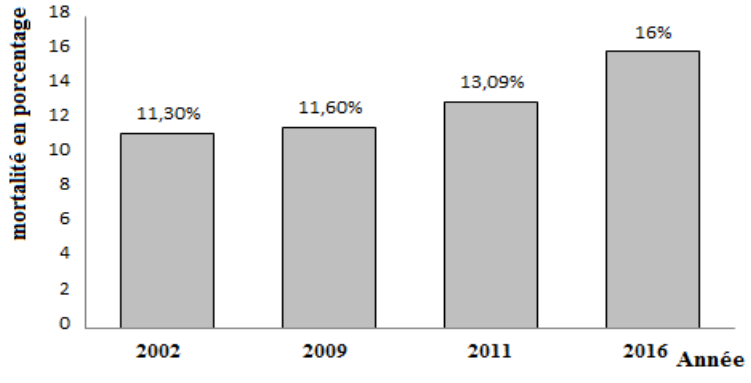


FIGURE 1.4 – Évolution de la mortalité au niveau de piétons *(Sources : Rapport ONISR Année 2002, 2009, 2011 et 2016)*.

Les bilans des accidents routiers à l'échelle mondiale ainsi qu'en France montre une hausse continue des fatalités des accidents routiers sur les usagers de la route en générale et plus spécifiquement pour les piétons. La cause de cette hausse n'est pas encore bien définie. Cependant, l'idée d'un décalage entre le comportement des piétons et l'adaptation de l'environnement et/ou du comportement des automobilistes pourrait expliquer ce phénomène. L'augmentation importante du nombre de véhicules à moteur et leur fréquence d'utilisation dans le monde – mais aussi une prise en compte insuffisante des besoins des piétons dans la conception des voies de circulation et l'aménagement du territoire conduisent a ce que les piétons soient de plus en plus exposés à un risque d'accident [19]. La vulnérabilité des piétons est encore plus grande lorsque la législation routière est insuffisamment appliquée [20].

Différentes hypothèses sont donc possibles, néanmoins il est important de garder en mémoire que le vieillissement de la population, impliquant une augmentation de la part des piétons âgés, pourrait également être une explication.

Dans ce qui suit, nous mettrons en évidence les difficultés spécifiques d'une personne âgés et leurs impacts sur son comportement lors de la traversée d'une rue.

1.3 Traversée de rue et difficultés spécifiques du piéton âgé

1.3.1 Description d'un piéton âgé

Les personnes avançant en âge sont souvent décrites par une terminologie très variée.

Le seuil à partir duquel une personne est dite âgée est variable. Selon le dictionnaire de Pierre Richelet, les "vieux" sont les personnes âgées de plus de 40 ans [21]. Cependant, dans les données de l'Institut National de la Statistique et des Études Économiques (INSEE) ils sont la catégorie des " plus de 60 ans " [22]. Alors que l'Organisation Mondiale de la Santé, considère que les personnes âgées sont ceux ayant "plus de 65 ans". D'un point de vue socio-économique, le repère est l'âge de la retraite.

Dans cette étude, en se référant au bilan des accidents routiers en général et en France en particulier, nous avons choisi de nous intéresser aux seniors de plus de 60 ans. En effet, à cet âge commence à apparaître plus fréquemment un état de fragilité, avec des conséquences en termes d'augmentation des difficultés à se déplacer, à se tenir en équilibre naturellement et des capacités à surveiller l'état de la rue ainsi qu'à estimer les dangers.

"En 2050, la France métropolitaine comptera entre 61 et 79 millions d'habitants selon les différentes hypothèses de fécondité, de mortalité et de migrations. Près d'un habitant sur trois aura plus de 60 ans, contre un sur cinq en 2005." On prévoit un doublement du nombre de personnes âgées de plus de 85 ans sur la période 2005-2020. Même si l'espérance de vie en bonne santé connaît une croissance plus rapide que celle de l'espérance de vie, le nombre de personnes dépendantes est en constante augmentation.

Ce rapide état des lieux montre clairement que le vieillissement de la population sera un enjeu majeur pour le futur de la société française tout comme d'autres pays

industrialisés.

Les personnes âgées, qu'on définit dans cette étude, comme ayant 60 ans et plus, représentent 18,8 % de la population française en 2015, celles âgées de 75 ans et plus ne représentent que 9,3% de cette population [22]. Ces derniers représentent pourtant 52% de la mortalité piétonne en 2015 [17].

1.3.2 Difficultés spécifiques du piéton âgé

Il est vraisemblable que les seniors soient de plus en plus mobiles. Ainsi, l'enjeu que représentent les seniors en termes de sécurité routière, déjà important aujourd'hui, devrait croître dans les années à venir. Néanmoins, la durée moyenne des déplacements quotidiens diminue avec l'âge : elle passe de 57 minutes en moyenne pour les 65-74 ans à 28 minutes pour les plus de 80 ans, et le territoire de vie se rétrécit selon l'Enquête Nationale de Transport et de Déplacement de 2008 [23]. Le renoncement de se déplacer et de quitter chez eux, par les personnes âgées est dû à plusieurs facteurs tels que des facteurs environnementaux, des limitations fonctionnelles ou psychologiques, mais elle est accentuée par les sentiments de peur de chuter, d'être accidentées, d'être bousculées, de l'imprévu. En fait, l'avancée en âge provoque une baisse des capacités motrices [24, 25, 26] et cognitives induisant une augmentation des risques de chute chez les seniors. La potentiel du danger augmente considérablement chez les piétons seniors surtout en fonction du type de trafic routier (Figure 1.5). Les gros véhicules présentent une source de danger majeur pour les sujets seniors. De même, le trafic dans les deux sens semble être une menace pour les sujets âgés avec un pourcentage de 97,1%.

	Nb	% obs.
gros véhicules (précisez : bus, camion,etc.)	35	100,0%
voitures mal garées	34	97,1%
voitures venant des deux sens	34	97,1%
deux-roues	33	94,3%
voitures venant en sens unique	31	88,6%
autre(précisez)	6	17,1%
Total	35	

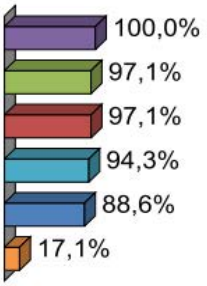


FIGURE 1.5 – Représentation du danger ressenti par les personnes de plus de 60 ans [27].

Le risque de décéder suite à un accident double tous les 10 ans après 60 ans [28]. En effet, pour un même accident, les conséquences s'aggravent avec l'âge, ce qui en fait un enjeu pour la qualité de vie et l'autonomie des seniors. Ces affirmations sont schématisées par la figure 1.6 qui représente le nombre de piétons tués et blessés hospitalisés (BH) pour un million de personnes par classe d'âge en 2014.

On constate sur la figure 1.6 que les chiffres de mortalités des piétons concernent davantage les tranches d'âge élevées.

2. source : Fichiers BAAC 2014, ONISR

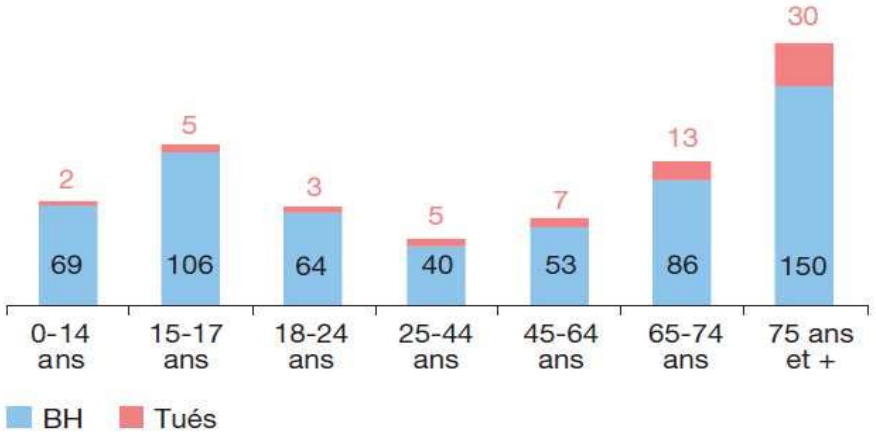


FIGURE 1.6 – Nombre de piétons tués et blessés hospitalisés (BH) pour un million de personnes par classe d'âge, 2014. [2]

Les piétons âgés de 65 et plus représentent la catégorie d'âge la plus touchée par les accidents routiers et ces chiffres sont en augmentation contenue (Figure 1.7).

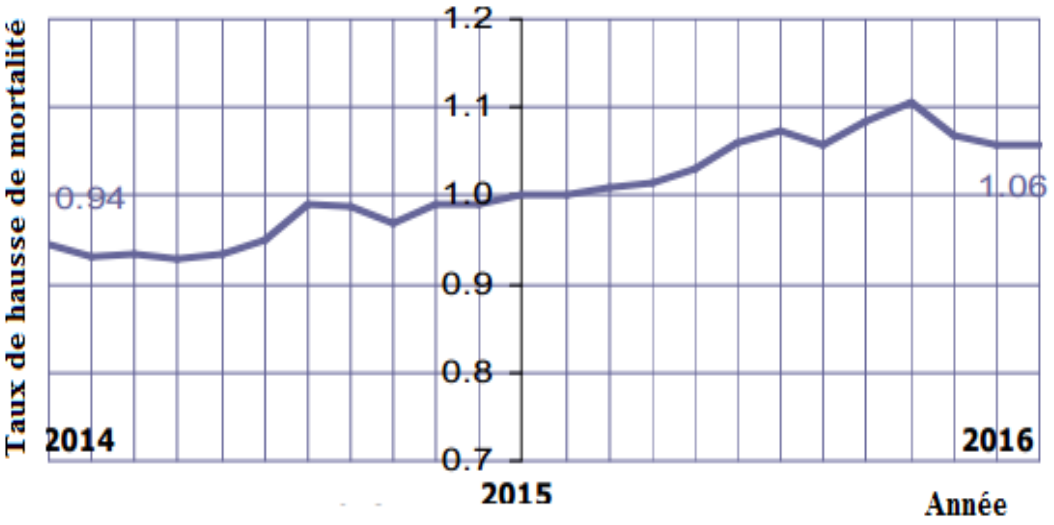


FIGURE 1.7 – Mortalité des séniors de 65 ans et plus [18].

La capacité de traverser la rue de manière efficiente devient plus difficile avec l'avancée en âge et d'autant plus lorsque la complexité de la tâche augmente [29] que se soit par les croisements, les carrefours et par le type (urbaine, non urbaine) de la zone de traversée, l'intensité du trafic et/ou l'action fait et pour le conducteur. En fait, les piétons âgés sont sur-représentés dans les accidents de croisement à mi-section, en particulier sur de routes larges à voies multiples, en ville animée avec un trafic bidirectionnel. L'embarquement ou la sortie des transports publics ainsi que les véhicules effectuant une manœuvre de recul peuvent aussi causer des accidents mortels pour ce type de piétons [29, 30, 31, 32]. Les piétons les plus âgés auraient plus de difficulté à

faire face à la circulation rapide et dense, aux rues larges, aux feux pour piétons ou à la lecture des signaux routiers [29, 33].

La vulnérabilité physique du piéton s'accroit avec l'avancée en âge et les personnes âgés représentent la catégorie la plus touchée par les accidents routiers. De ce fait, leurs vie est de plus en plus mise en danger et la sécurité des seniors est engagée sur l'espace public.

Prendre des mesures de sécurité routière profitables à tous est vitales pour les sujets plus âgées est constitue alors un enjeu de santé public majeur. Un projet pionnier dans ce domaine est le programme PISTE" PIétonS et comportement de TraverséE de rues" proposé et supporté par la Fondation de Sécurité Routière française [27]. Dans cette même optique nous nous intéresserons alors dans ce projet de thèse à l'analyse comportementale des piétons âgées lors de la traversée d'une rue en relation avec le trafic routier existant afin de prendre les décisions et les précautions nécessaires à la sécurité des cette catégorie de piétons.

1.4 Méthodes de collecte de données et d'étude du comportement piétonnier

L'analyse du comportement des piétons lors de la traversée d'une rue est une thématique de recherche en plein essor. Cette analyse se base essentiellement sur un ensemble de données et de paramètres à des degrés de détails différents. En effet, plusieurs types de travaux ont fait l'objet d'une collecte de données microscopiques et/ou macroscopiques pour les piétons.

Une variété d'études ont abordé ce sujet en s'appuyant sur : i) des analyses statistiques effectuées sur des données collectées manuellement, à l'aide de questionnaires, entretiens et sondages ou à partir d'observations directes et d'expérimentations in situ, et ii) des méthodes automatiques se basant sur des expérimentations en simulateurs ou des mesures à partir de capteurs de vision.

1.4.1 Les méthodes manuelles

1.4.1.1 Questionnaires, entretiens et sondages

Le principe de base de ces approches consiste à proposer une série de questions écrites ou orales aux participants. De nombreuses études utilisent, partiellement ou totalement, cette méthode d'analyse comportementale, qu'elles fassent ou non référence à la réglementation. Ces études visent le plus souvent à comprendre les raisons ayant conduit à la décision de traverser, que ce soit dans un contexte sécurisé ou avec une certaine prise de risque.

La "Théorie du Comportement Planifié" (TCP) [34] a été utilisée avec succès dans le passé pour tenir compte des intentions des piétons lors de traversée d'une rue dans des situations risquées. Cette théorie constitue une base pour quelques études postérieures [35, 36]. L'étude [36] étend les investigations en examinant l'importance relative des composantes du modèle pour prédire l'intention de traverser pour quatre groupes

d'âges différents, hommes, femmes, conducteurs et non-conducteurs. La conception repose sur une sélection de situations, chacune suivie d'une série de questions demandant aux participants de s'imaginer dans une situation de traversée d'une rue (par exemple, [35]). Une série de mesures générales et des questions sur les variables démographiques ont également été incluses, mais seuls l'âge, le sexe et le statut du conducteur étaient jugés pertinents pour ces analyses.

L'approche réalisée dans [37] étudie le comportement des piétons par le suivi de son basculement de tête pour détecter son niveau de vigilance lors de la traversée et l'impact de la peur de chuter et de l'âge. Les informations sont collectées à partir d'une vidéo capturée dans des conditions très strictes (plein soleil, absence de pluie). Ainsi, l'effet des informations sur l'âge, le "taux" de peur de chuter, le nombre de fois où le piéton traverse cette route et sa situation matérielle sont étudiés. A l'issu de ce travail, les auteurs présentent quelques interprétation à savoir, les piétons âgés sont plus touchés par les accidents, les femmes ont une vitesse plus lente ce qui diminue leur marge de sécurité et la signalisation routière facilite l'action de traversée. Ces conclusions restent cependant insuffisantes dans la mesure où elles se basent sur des informations subjectives (questionnaires) et ne couvrent pas la totalité de la situation.

Les questionnaires, les entretiens et les sondages offrent l'avantage de réaliser des statistiques à partir d'échantillons importants intégrant une multitude de scénarios de traversée d'une rue. Ce type d'étude contribue plus particulièrement à comprendre le comportement des piétons à partir de la psychologie et de la sociologie, ayant pour objectif la maitrise de prise de décision des piétons et leur perception de risque. Cependant, ces méthodes rencontrent plusieurs limites, dont les biais de réponses qui dépendent de plusieurs facteurs tels que : les normes sociales et démographiques. En outre, ces méthodes possèdent comme limitation générale d'être le plus souvent réalisées en dehors d'un contexte réel de traversée de rue et de ne pas pouvoir rendre compte de l'action immédiate et aléatoire du piéton vis-à-vis d'un danger concret.

1.4.1.2 Observations et expérimentations in situ

Les observations et les expérimentations in situ ont comme objectifs d'analyser le comportement des piétons lors de traversée d'une rue à base de suivi et d'annotation de leurs actions dans une situation de traversée réelle. L'analyse comportementale est effectuée en fonction de variables fixées soit à priori (cas des expérimentations in situ) soit à posteriori (cas des observations).

Les observations (ou "videotape") consistent à collecter une base de vidéo de traversée qui enregistre le comportement de piéton vis-à-vis d'un paramètre bien déterminé, généralement un aménagement urbain. En effet, l'infrastructure de la route et la section de traversée constitue souvent le paramètre le plus étudié. C'est par exemple le cas de l'approche présentée dans [38] par laquelle les auteurs établissent une classification des intersections en 6 catégories d'aménagements et constatent que les classes d'intersections présentant le plus d'aménagements atteignent les moyennes de piétons blessés les plus élevées dans la période 1999-2008.

En outre, [39] exploitent les enregistrements vidéo tirés d'observations d'une zone

de traversée dans le centre-ville de Hong Kong pour étudier le passage piétons comme étant un aménagement urbain cible. Une autre étude effectuée à New Delhi consiste à filmer les piétons au niveau des 3 intersections de rue équipées avec des feux. Cette dernière à pour objet de modéliser l'effet des feux de circulation sur le comportement des piétons ainsi que le temps d'attente avant la traversée. Ces paramètres sont quantifiées en fonction du gap sécuritaire pour chaque individu. L'étude a montré une grande hétérogénéité entre les gaps acceptables pour les différents participants qui dépend principalement des paramètres démographiques (âge et genre) et des aménagements urbains disponibles.

Le "videotape" ou l'annotation visuelle de vidéo est aussi utilisé par [40] pour examiner les interactions entre les véhicules tournants et les piétons traversant les intersections. L'objectif est de développer un modèle d'interaction "véhicule-piéton" et classer les intersections en fonction du risque qu'elles imposent aux mouvements des piétons en raison des véhicules tournants pendant le feu vert. Le comportement des piétons et des véhicules à l'intersection a été aussi l'objet d'une étude élaborée dans [41]. Une caméra est installée au niveau de l'intersection. Depuis son installation, la caméra a été témoin de quinze accidents piétons-véhicules. Les séquences vidéo ont été observées et annotées pour quantifier les trajectoires des piétons et des véhicules. Des mesures détaillées de l'intersection ont été prises pour l'analyse des accidents.

Bien que, les expérimentations in situ permet des avancées dans ce domaine. Elles se basent sur des paramètres et scénarios fixés à priori, en plus les participants sont conscients qu'ils seront filmés voir même entrainés sur les scénarios des expérimentations.

Un autre mode d'observation des piétons consiste à réaliser des mesures en conditions réelles. La trajectoire empruntée, la vitesse, l'accélération ou une analyse détaillée de la démarche ou de la foulée, ont fait l'objet de plusieurs travaux. Ces études sont effectuées « en direct », ou en temps différé sur un enregistrement vidéo des scènes [41, 42, 43, 42]. Dans ce dernier cas, les informations sont plutôt de nature quantitative.

On constate une utilisation exhaustive de la vidéo comme support pour l'analyse du comportement du piéton lors de traversée d'une rue. Néanmoins cette utilisation se limite le plus souvent à une simple observation et annotation manuelle sans l'intégration d'aucune technique de vision par ordinateur pour la segmentation, la détection, le suivi des objets cible et/ou la collecte de données à analyser. Ces éléments constituent les limitations de ces méthodes qui restent laborieuses en termes de temps de calcul et faibles en termes de précision.

1.4.2 Les méthodes automatiques

1.4.2.1 Les méthodes basées sur les capteurs actifs

Des études récentes ont utilisé les nouvelles technologies pour détecter et suivre les piétons afin d'analyser leurs comportements lorsqu'ils traversent la rue. A titre d'exemple, on peut citer l'approche élaborée dans [44] qui présente un système pour détecter les piétons dans la zone la plus dangereuse. Cette zone est définie comme la région située derrière un obstacle fixe (généralement une voiture garée). En effet,

une étape de détection et de classification des obstacles est effectuée en interprétant les impulsions lasers fournies par un scanner laser fixé à l'intérieur de la voiture. Un obstacle fixe est représenté par une poly-ligne chevauchée, ayant une taille réduite et la distance entre le centroide de poly-ligne courant et ceux des anciens est inférieure à un seuil fixe. Après fixation de la zone de recherche, les auteurs utilisent les ondelettes de Haar avec l'algorithme d'Adaboost pour la détection et la classification des piétons présents dans cette zone. Ils s'intéressent aux piétons partiellement occultés par les voitures garées au bord de la route. Si la direction du piéton suit celle de la frontière de la rue vers son centre, une alerte est envoyée vers le conducteur indiquant la possibilité d'un accident.

Les performances de cette approche sont illustrées dans le tableau 1.2

TABLE 1.2 – Comparaison entre les stratégies de protection [44].

Test	Détection correcte	Faux rejet	Fausse acceptation
A	24	11	1
B	5	0	0
C	8	0	0

La reconnaissance des piétons est effectuée sur une zone limitée ce qui augmente sa performance temporelle et réduit les fausses alertes.

Une autre étude [45] utilise également des systèmes GPS par satellite pour assurer 2 fonctions principales. La première fonction suit et affiche la variation de la vitesse du véhicule. Le système avertit le conducteur lorsque la limite de vitesse a été atteinte ou est dépassée d'une certaine valeur. Alors que le rôle de la deuxième fonction est l'affichage des panneaux routiers dans le véhicule en fonction d'un système lié à un satellite qui envoie ces signaux sur le tableau de bord ou sur le pare-brise. L' affichage est transparent pour ne pas perturber la vision du conducteur [45]).

Ces approches, basées sur un scanner laser ou un système satellitaire, reflètent l'importance d'utilisation des nouvelles technologies pour assurer la sécurité des usagers de la route. Cependant, leur application dans le monde réel reste limitée et coûteuse.

Un pré-requis pour la réussite du développement d'un système de sécurité routière est la détection et le suivi fiable des piétons d'une part et la détection des signes de communication entre le piéton et le conducteur d'autre part. Dans ce sens le travail [46] cherche à répondre à deux questions préliminaires : Comment la distance entre le piéton et le véhicule et le temps de collision (Time To Collision : TTC) qui dépend de la vitesse des voitures influencent le comportement du piéton ? Quelles sont les paramètres que le conducteur utilise pour prédire l'intention du piéton ? En effet, pour déterminer les deux variables TTC et la distance, une étude expérimentale sur une population de 30 personnes a été réalisée. Ces paramètres sont déterminés à partir d'un capteur radar. A l'issue de cette expérimentation les relations de probabilité de traverser, TTC et vitesse de véhicule sont dégagées (Figure 1.8).

La plupart des piétons ont entre 3 et 7 secondes comme TTC pour traverser devant une voiture. La probabilité de traverser pour un $TTC = 6\ s$ est de 70% pour la vitesse de 35 $km.h^{-1}$ contre 90% pour 45 $km.h^{-1}$ et 60 $km.h^{-1}$. Les auteurs concluent alors que plus la vitesse est élevée plus la probabilité de traverser devant la voiture est élevée.

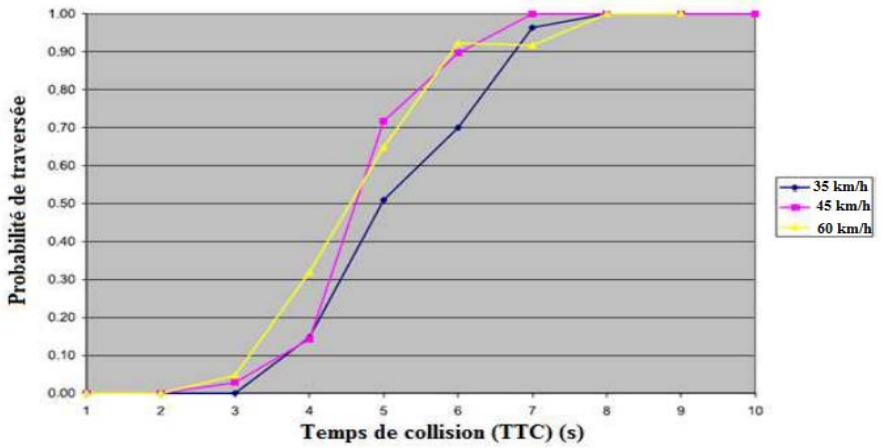


FIGURE 1.8 – Variations de probabilité de traverser en fonction du TTC et de la vitesse du véhicule [46].

Afin de connaitre les paramètres que le conducteur utilise pour prédire l'intention du piéton (traverse ou attend), un ensemble d'expériences est réalisé [46]. Il s'agit à chaque fois de cacher un paramètre des candidats (tête, jambes, trafic, la totalité de la personne) et de demander au conducteur d'essayer de prédire le comportement du piéton. L'importance de ces paramètres dans la prise de décision est illustrée par la figure 1.9.

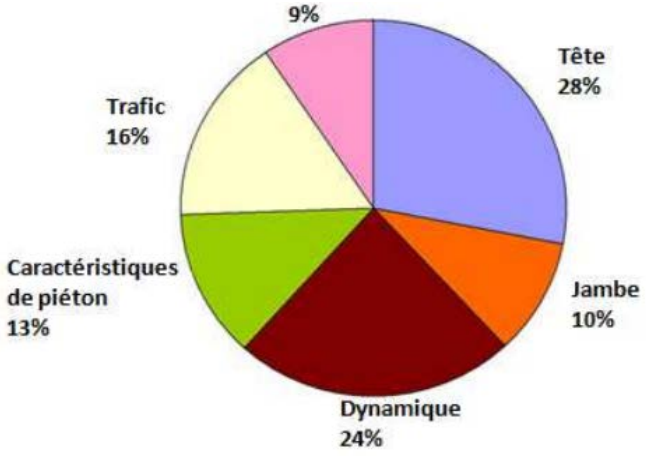


FIGURE 1.9 – Contribution des différents paramètres dans la prédiction de décision de traverser [46].

1.4.2.2 Les simulateurs

La conception de simulateurs se base sur la réalité projetée ou simulée au sein du laboratoire. La réalité simulée par ordinateur vise à rendre le milieu urbain reproductible en laboratoire. Ainsi, cette méthode offre l'avantage de pouvoir moduler facilement le trafic dans un environnement défini et de rendre possible l'analyse du comportement du piéton face à des situations à risque.

Plusieurs travaux dans la littératures ont tiré profit des progrès dans la réalisation d'environnements artificiels afin de concevoir des simulateurs de traversée d'une rue et analyser le comportement du piéton. Comme instance, on cite le travail réalisé par [27] qui estime le temps de traversée nécessaire pour qu'un piéton âgé effectue une traversée sur la base d'une étude en laboratoire.

Le simulateur français de l'IFSTTAR (ex. [47, 48, 49, 50, 51, 52, 53]) intègre 3 écrans et utilise un simulateur composé de 4 PC pour gérer et contrôler la projection des scénarios de test. Ces études emploient des scènes fixes pour représenter une rue à sens unique de $4,20\ m$ avec un trafic composé par une moto et deux véhicules. L'objet principal de ces études est de vérifier si le piéton effectue une traversée sécuritaire en fonction des gaps de sécurité et des marges entre les véhicules. Elles cherchent, en effet, à analyser la capacité du piéton à manipuler les gaps sécuritaires (Figure 1.10) en fonction des vitesses des véhicules.

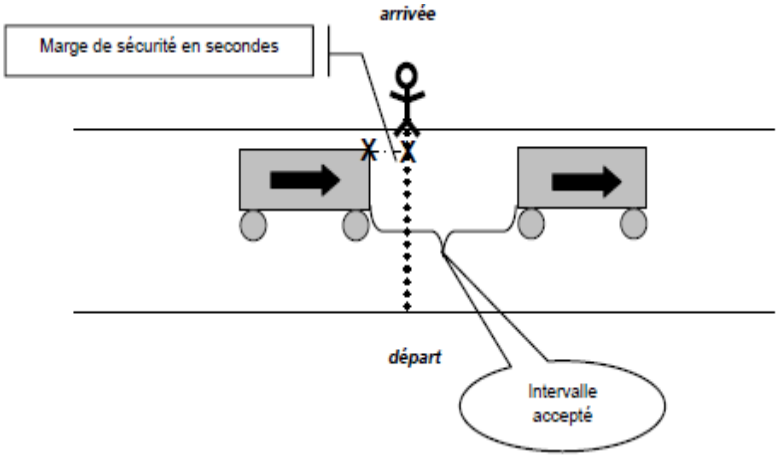


FIGURE 1.10 – Exemple de situation expérimentale.

Malgré les analyses importantes et les contributions fournies par ces approches, ces simulations restent loin du contexte réel, notamment avec l'entrainement des piétons avant de tester et d'utiliser des scénarios prédéfinis.

1.4.2.3 Les méthodes basées sur les capteurs passifs : La vision par ordinateur

Il y a eu un développement accéléré des techniques de vision par ordinateur ces derniers années. Il devient de plus en plus accessible de détecter, de suivre et de classer automatiquement les usagers de la route par l'analyse automatisée des données vidéo. Des nombreuses applications éprouvées et produits commerciaux ont émergé en réponse aux besoins croissants en termes de sécurité des habitats et de vidéo-surveillance. Les techniques de vision par ordinateur dans les applications de la sécurité routière sont particulièrement attrayante [54] et quelques travaux et interventions côté piétons et/ou conducteur de la part de quelques chercheurs et/ou entreprise de production de véhicules telles que Renault [55] et Honda [56] commencent à être disponibles.

La vision par ordinateur appliquée à la sécurité routière consiste à extraire des données sur les piétons à partir de traitement automatique d'images et de vidéo. Les étapes principales de ce processus sont l'extraction des caractéristiques, la détection et le suivi des objets cibles (piétons et véhicules) et la collecte et l'analyse de données [54].

Dans ce contexte, un travail initial a été introduit dans [57] pour déterminer si le piéton est dans une situation de danger critique et envoyer un message d'avertissement au conducteur. Le piéton est détecté et suivi comme la cible unique en mouvement. Ensuite, sa vitesse est calculée en soustrayant les positions de chaque centre de gravité entre les images sélectionnées. Enfin, la direction de déplacement est obtenue en analysant le vecteur de mouvement du piéton à l'aide du code de chaîne de 'Freeman'.

La principale limitation de ce travail est qu'il repose sur un tracker classique (soustraction de l'arrière-plan) qui présente de nombreux défauts et traite le comportement des piétons indépendamment de l'environnement, en particulier la présence de véhicules.

Dans ce même contexte, un autre travail est présenté dans [58] pour une prédiction a priori de l'intention de traverser la rue chez le piéton. Les auteurs utilisent alors, les histogrammes de HOG pour déterminer si le piéton est dans une position de traversée légale ou non. L'étude expérimentale s'est basée sur 5 scénarios (respectivement Figure 1.11 .a), 1.11 .b), 1.11 .c), 1.11 .d) et 1.11 .e)) dans lesquels le déplacement du piéton est caractérisé par une direction prédéfinie.

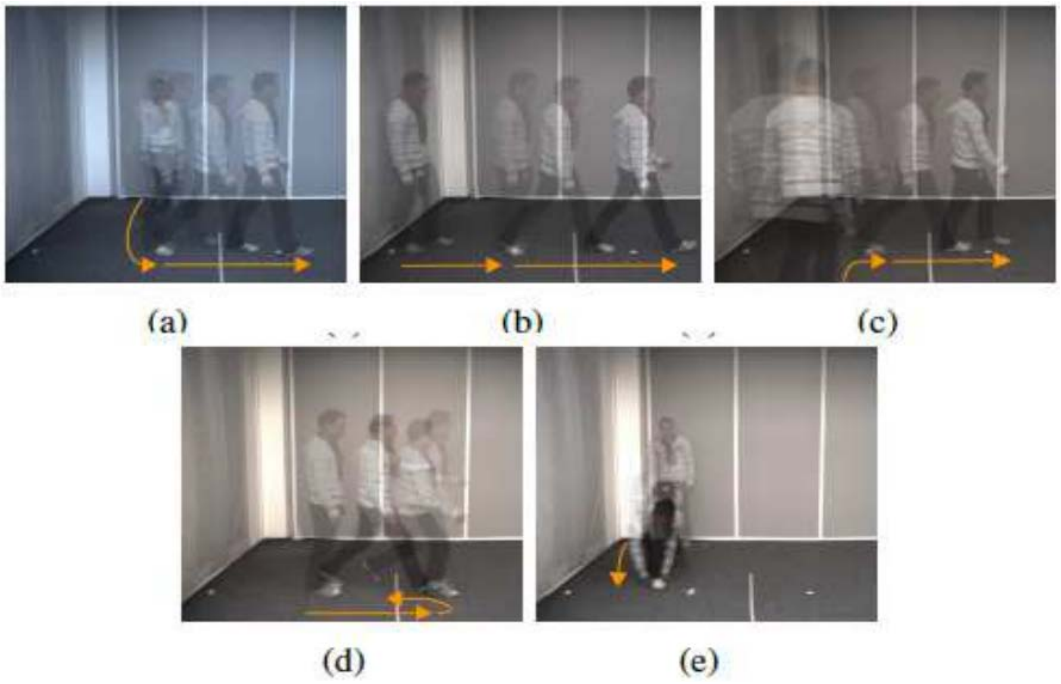


FIGURE 1.11 – Illustration des différents scénarios de traversée de rue [58].

En se basant sur ces scénarios, la silhouette prend des formes précises. Un mouvement correcte doit donc correspondre à une forme précise de la silhouette tel qu' illustré

par la figure 1.12.

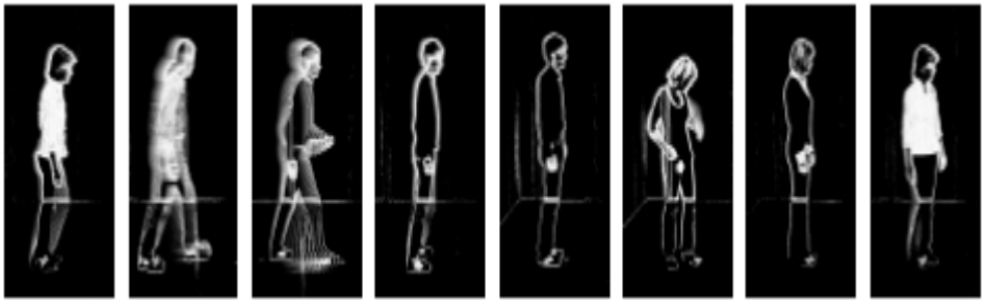


FIGURE 1.12 – Présentation de quelques positions : les quatre premières sont correctes et les quatre dernières ne sont pas correctes [58].

Cette étude s'intéresse uniquement aux comportements des piétons indépendamment du trafic routier. Cette lacune ne permet pas d'avoir une vue d'ensemble de la situation. En fait, ce travail est considéré comme une étape préliminaire dans l'étude de la traversée d'une rue.

La détection de piétons à distance et l'estimation de certaines de ses caractéristiques physiques ou géométriques sont réalisées au sein d'un système dit de "détection de l'environnement du véhicule". Un tel système est composé d'un ou plusieurs capteurs : caméras mono ou stéréo vision, dans les longueurs d'ondes visibles ou infra rouges, radars, lidars, télémètres à balayage... et d'un module de traitement du signal adapté.

Depuis 2004, le constructeur automobile japonais Honda propose, en option, sur son modèle haut de gamme « Legend », le premier système commercial de détection de piétons (stéréo vision avec des caméras dans le domaine infrarouge), qui alerte le conducteur lorsque des piétons sont détectés, la nuit [56]. L'information fournie par un système de détection de piétons seul n'est pas directement exploitable par le conducteur. En ville, un tel système se déclencherait très souvent, le conducteur finirait par s'en laisser et le désactiver. Traiter et filtrer les informations issues de la détection permet de réaliser une prestation automobile sûre, utile et acceptable.

Le suivi à la fois des véhicules et des piétons en utilisant des algorithmes de suivi basés sur la vision a ensuite été abordé dans [59], en fonction de contraintes cinématographiques et visuelles, pour établir une correspondance entre les détections. En effet, Karim et al. [60] ont abordé le problème des conflits Piétons-Véhicules (P-V) en utilisant les techniques de suivi des véhicules et des piétons. Ce travail représente une étape importante dans l'analyse du comportement des piétons en utilisant la vision par ordinateur, mais reste cependant insuffisant car il n'a abordé qu'un seul paramètre de conflit entre les piétons et les véhicules, qui est le temps.

Dans ce même contexte et dans l'objectif de corréler les paramètres des piétons et des véhicules une tentative a été faite par [61] pour étudier le comportement des usagers de la route en utilisant une configuration des caméras normales et un système de vision basé sur un ordinateur personnel (PC), ainsi qu'un système expert développé

spécifiquement pour aider les personnes non expérimentées à effectuer un passage sécurisé. Ce processus a été evolué en étudiant les conflits entre les piétons et les véhicules en tant qu'indicateur d'un accident. Deux caméras ont été utilisées pour filmer le trafic et les piétons. Dans un second temps, un système de vision a été utilisé pour extraire 3317 observations de conflits par des images numériques dans six endroits différents à Irbid-City, en Jordanie. Une base de données comportant des informations sur les piétons, le trafic et la géométrie ainsi que les caractéristiques de la zone de traversée a été développée. Les variables collectées dans cette base de données comprennent : la vitesse des piétons, la vitesse du véhicule, la distance du véhicule par rapport au piéton, le type de véhicule, la géométrie de la route, l'emplacement des conflits, les aménagements urbains, la distance des piétons par rapport au lieu de traversée, l'âge des piétons, le genre des piétons et l'angle de traversée. Des régressions statistiques ont été effectuées pour établir des modèles utiles pour estimer la vitesse des piétons à partir des variables mentionnées précédemment. Un système expert à base de règles de "Si ... alors" et le chaînage vers l'avant et vers l'arrière des règles basées sur les connaissances des experts ainsi que des arbres de décision ont été développés, à l'aide de "VP-Expert Shell", afin d'aider les piétons non expérimentés à prendre des décisions sûres pour effectuer une traversée sécuritaire. Le système a été validé avec des données réelles de piétons traversant dans différents endroits pour les deux cas suivants : traversées sécurisées et non-sécurisées.

Même si le système développé est assez complet et prend en considération les différents paramètres (piétons, trafic et environnement), la majorité de ces derniers tel que l'âge des piétons est collecté soit manuellement soit à partir de la technique de 'videotape'.

1.4.3 Conclusion sur les méthodes d'études du comportement piétonnier

De nombreuses campagnes de mesures ont été réalisées et un certain nombre d'outils développés pour la collecte de données afin de mener une étude des comportements piétonniers. Ces approche sont soit manuelles, semi et/ou automatiques basées sur des observations, des expérimentations en situ, 'videotape', des simulateurs et la réalité projetée et/ou les techniques de vision par ordinateur. Ces approches sont centrées sur le piétons, les véhicules, les conducteurs ou encore les aménagements urbains.

Néanmoins, si l'analyse du comportement piétonnier à partir des techniques manuelles (questionnaires, entretiens, sondages, observations directes et les expérimentations in situ), peu présentent une précision raisonnable. Ces approches sont le plus souvent couteuses en temps et/ou en matériel et génèrent des constatations simples et faibles en détails. La migration vers des techniques plus puissantes et tirant profit des progrès technologiques et des résultats avancés des recherches scientifiques est alors une stratégie encouragée voir même une nécessité.

De nombreuses juridictions installent des caméras vidéo "grand public" aux intersections de rues à des fins de surveillance. De plus, par analogie avec le flux de recherche (ITS) bien établi qui essaie de conférer de "l'intelligence" aux systèmes de transport,

l'utilisation de techniques de vision par ordinateur, considérée comme une tentative d'équiper les systèmes de transport d'un "sens visuel", présente une solution adoptée pour la collecte automatiques de données sur les accidents routiers.

Les techniques de vision par ordinateur reposent sur les capteurs vidéo comme principale source de données. Les capteurs vidéo sont sans doute l'une des méthodes potentiellement les plus puissantes pour la collecte de données des utilisateurs de la route (piétons et véhicules). De plus, les données vidéo sont riches en détails, les appareils d'enregistrement et de traitement sont de moins en moins coûteux mais aussi de plus en plus précis et puissants. Ainsi, grâce à l'application des techniques de vision par ordinateur, des mesures des paramètres automatisés et précises sont possibles de manière robuste et efficace dans le temps. L'analyse peut ensuite être poursuivre en corrélant ces paramètres pour inférer des indicateurs de conflit permettant de caractériser le comportement du piéton ainsi que les potentielles interactions entre les piétons et les véhicules.

1.5 Techniques de détection des conflits du trafic

1.5.1 Conflit du trafic

Un conflit du trafic est un concept très ambiguë dont plusieurs définitions ont été proposées dans la littérature [13, 54]. Selon la définition la plus admise, un conflit de trafic a lieu lorsque "deux ou plusieurs usagers de la route se rapprochent dans l'espace et le temps à tel point qu'une collision est imminente si leurs mouvements restent inchangés" [54, 62].

L'analyse des conflits du trafic dépend de la stratégie ainsi que du type de technique à mettre en œuvre (détection ou prédiction du danger). Une telle stratégie demande un certain nombre d'indicateurs de conflits qui devront être soigneusement définis et calculés.

1.5.2 Les indicateurs de conflit de trafic et de danger lors de la traversée d'une rue

Les indicateurs de conflit objectifs sont des mesures quantitatives calculées à coté des piétons et des véhicules sur un intervalle de temps pour prédire un potentiel point de collision [54]. De nombreuses études dans la littérature ont utilisé ces indicateurs pour quantifier la sécurité des piétons sur la route. Les indicateurs objectifs permettent de dépasser certaines limites de subjectivité des indicateurs traditionnels de conflit basés sur l'observateur. Ils permettent aussi la validation et les comparaisons d'études. Le tableau 1.3 résume quelques indicateurs.

TABLE 1.3 – Les différents indicateurs de conflit de trafic [54]

Indicateur de conflit	limites
Le Temps à l'Accident (TA) [63]	— Ne prend pas en considération le temps de reaction. — Demande que le piéton soit déjà dans une situation de collision potentielle.
Le Temps à Collision (TTC) [64]	— Demande que le piéton soit déjà dans une situation de collision potentielle. — Ne prend pas en considération l'effet de la vitesse. — Ne prend pas en considération la durée de l'interaction.
L'Extension de TTC [65]	— Ne prend pas en considération la durée de l'interaction comme pour le TTC.
Le temps sur le passage piéton [66]	— Ne se base pas sur un mécanisme de collision.
Le Temps de post-collision (PET) [67]	— N'utilise pas des mesures de vitesses et de distances. — Manque de paramètres de mesure de conflit. — Manque d'une définition claire des infractions sur la route. — Capacité limitée à comprendre la sévérité de l'interaction entre les automobilistes et les piétons.
La marge de sécurité [68] (en seconde)	— Manque d'une définition claire des infractions sur la route. — Nécessite une extrapolation des pistes de l'utilisateur de la route.
Décélération au temps de sécurité (DST) [69]	— Nécessite une extrapolation des pistes de l'utilisateur de la route.

L'utilisation de mesures quantitatives de sécurité a été préconisée comme une approche objective pour aborder ces problèmes de conflit du trafic et pour offrir une analyse plus approfondie que s'appuyer uniquement sur les statistiques des accidents. L'une des méthodes les plus développées d'analyse du conflits de trafic est présentée dans [56, 70, 71]. Néanmoins, l'utilisation d'un seul indicateur de conflit du trafic n'est pas assez objectif et ne prend pas en considération les différentes causes possibles

d'accident comme illustré par le tableau 1.3 qui résume les limitations majeurs de ces indicateurs.

Si ces indicateurs des conflits permettent une précision impotente pour quantifier les conflits du trafic, la limitation majeure de leur utilisation pour mesurer le conflit de trafic et estimer le danger qui menace le piéton sur la route est l'incohérence de le faire. En effet, l'utilisation d'un seul indicateur ne suffit pas pour une caractérisation fiable du conflit [54].
Ces constatations appuient une hypothèse de cette thèse selon laquelle les indicateurs de conflit objectifs, mesurant les aspects de sévérité différents et parfois indépendants, devraient être mit en relation pour assurer une meilleure représentation du danger réel et parfois inobservable à l'œil nu des conflits du trafic . Cette hypothèse est expliquée avec plus de détails dans la section suivante ainsi que le chapitre 5.

1.5.3 Les mesures de sécurité et les stratégies d'analyse de conflits du trafic

Les différents usagers de la route, en particulier les piétons et les véhicules sont en interaction quotidienne. Dans un environnement dense tel que les routes des grandes villes, ces interactions sont parfois difficiles à anticiper et peuvent conduire à des conflits graves. Plusieurs types de stratégies d'analyse des conflits P-V ont été mis en place dans l'objectif de la réduction durable du nombre des traumatismes liés à des accidents de la circulation. On distingue quatre classes de ces stratégies. Ces classes sont présentées dans le tableau 1.4. Elles se distinguent notamment par leurs horizons de déclenchement temporel, leurs efficacités et leurs complexités.

Par ailleurs, les techniques d'analyse de conflit du trafic peuvent être divisées en deux catégories. On distingue les mesures de sécurité primaires complexes telles-que le pré-crash piéton et la prévention des chocs piétons et les mesures de sécurité secondaires à savoir la protection "choc piéton" [13]. Alors que l'alerte piéton est considérée comme un "mode de sécurité informative", son intérêt paraît limité [13].

Il est à noter que les différentes stratégies abordées dans le tableau 1.4 sont complémentaires voire même inter-connectées. En fait, une détection de choc n'est faisable et efficace que suite à une étape de prévention et de prédiction d'un choc potentiel. Après la détection et/ou la prévention de choc, une mesure de sécurité doit être prise telle que l'alerte ou les dispositifs de pré-crash comme le freinage d'urgence. De ce fait, un système "complet" de protection des piétons devrait être équipé de toutes ces stratégies abordées dans le tableau 1.4. Ainsi, le déclenchement d'une ou de toutes ces stratégies serait progressif, en fonction de la situation de conflit P-V, des indicateurs de conflit estimées, et de l'impact du danger. De plus, il est recommendé d'avoir un déclenchement adaptatif pour suivre le comportement du conducteur [70, 71].

Un système adaptatif complet de protection des piétons est décrit par la figure 1.13.

Ce système comporte les capteurs de détection et d'estimation de différents éléments caractérisants le conflit P-V à savoir, piétons, véhicules, conducteurs, environnement, objets et obstacles de l'environnement. Ces capteurs sont de type physique implémentés au sein de véhicules. Les capteurs physiques génèrent des signaux bruts qui seront

TABLE 1.4 – Comparaison entre les méthodes de protection des piétons [13]

Méthode	Protection	Pré-crash	Prévention du choc	Alerte
Principe	Agir à partir du premier impact	Agir un peu avant que l'impact devienne inévitable	Agir tant que le choc est encore évitable	Avertir le piéton et/ou le conducteur lorsqu'un risque est détecté
Dispositifs	Passifs : absorption par le capot, les ailes ou le pare-choc. Réactifs : soulèvement du capot (amortir le choc avec la tête du piéton) et airbag piéton	Préparation des systèmes réactifs utilisés en protection "choc" et déclenchement du freinage d'urgence	Pilotage de la trajectoire du véhicule : changement (automatique) de direction et/ou d'allure	Alerte du conducteur et/ou du piéton et modification des trajectoires à leurs initiatives
Avantages	Réduction de la violence du premier impact	Très nette réduction de la violence du choc	Evitement de l'impact	Contrôle par le conducteur et/ou le piéton
Limites	Effet catapulte probable : L'impact augmente avec sol	Impact non évité et complexité	Complexité	Efficacité du temps de réaction du conducteur et/ou du piéton
Décision	Systèmes réactifs : détection de "choc piéton"	Détection des piétons, prédiction des chocs véhicule-piéton		
Points à traiter	Intérêt à démontrer	Faisabilité, législation	Faisabilité	Ergonomie : acceptabilité par le conducteur

ensuite traités et interprétés. Deux types de mesures de sécurité sont à prévoir suite à ces interprétations et à la prédiction des états futurs des véhicules et des piétons :

— Détection de choc P-V : La détection de choc P-V consiste à détecter un piéton ou un obstacle en contact avec la face avant du véhicule. Le système de détection de choc se base particulièrement sur la détection des piétons et peut être plus généralisé et détecté tout obstacles au niveau du pare-choc avant du véhicule et ensuite une classification est effectuée pour distinguer les différents types d'obstacles (Figure 1.13. a). Des mesures de sécurité secondaires seront déclenchées par cette stratégie selon le type d'obstacle détecté.

— Prédiction des chocs P-V : La prédiction des chocs P-V cherche à anticiper le choc avant de son occurrence et prendre des dispositifs de sécurité. Les mesures de sécurité envisagées par ce type de stratégies sont dites "primaires" car il

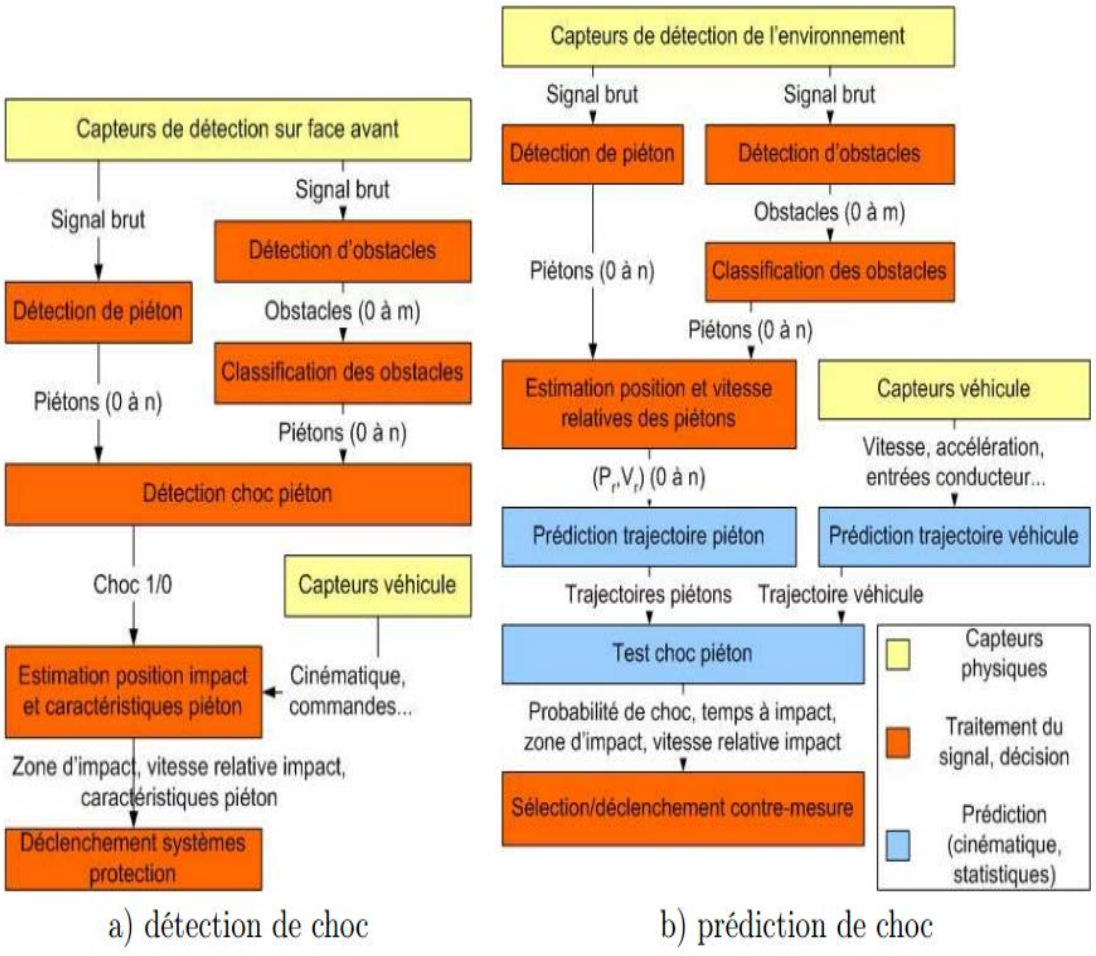


FIGURE 1.13 – Système de protection des piétons [70].

s'agit de prévoir le choc avant qu'il se produit. La prédiction des chocs se base sur l'analyse de trajectoire du piéton à l'instant t et fait des anticipations sur leur position à l'instant $t + x\,temps$ pour prédire si leur trajectoire présente des possibilités de risques (Figure 1.13. b).

Dans l'objectif, de concevoir une stratégie d'analyse des conflits du trafic afin de protéger les usagers de la route, quelques travaux ont déjà été proposés dans la littérature.

Le travail [72] est l'étude pionnier dans ce contexte, et proposait une procédure formelle pour observer les conflits du trafic. Les auteurs ont suivi une ligne de raisonnement semblable à celle de 'Heinrich' en se basant sur l'observation des conflits du trafic pour extraire des données suffisantes permettant de générer des inférences sur les accidents de la route. Ils ont conclu que la réduction des événements de circulation dangereux pourrait entraîner une réduction de la fréquence des collisions routières [72].

Malgré les limites de la définition conceptuelle des conflits du trafic et la procédure de base pour les observations de conflit du trafic dans cette étude, elle peut être considérée comme une étape importante dans le développement des techniques complètes

d'identification des conflits du trafic.

La technique suédoise est parmi les techniques, les plus connues demeure encore utilisée activement pour les évaluations quotidiennes de la sécurité [73]. Cette technique repose sur le Temps d'Accident (TA) entant qu'indicateur de conflit. Le TA est défini comme "le temps restant à partir du moment où l'action évasive est prise jusqu'à ce que la collision ait eu lieu si les usagers de la route avaient continué avec des vitesses et des directions inchangées. Sa valeur peut être calculée en fonction des estimations de la distance au point potentiel de collision et de la vitesse lors de l'action évasive " [73]. La vitesse et le TA sont utilisés pour distinguer entre "conflits sérieux" et "conflits non sérieux".

Un modèle probabiliste pour prédire les accidents de la circulation en utilisant un suivi de véhicules en trois dimensions (3-D) est proposé dans [74]. Les données, y compris les trajectoires des mouvements, sont d'abord obtenues par le suivi du véhicule basé sur un modèle 3D. Un algorithme à base de réseaux de neurones est ensuite appliqué pour apprendre les profils d'activité à partir des trajectoires des échantillons. Enfin, l'activité du véhicule est prédite en localisant et en appariant chaque trajectoire partielle avec les modèles d'activité appris, et la probabilité d'occurrence d'un accident de la circulation est ensuite déterminée. Les expériences montrent l'efficacité des algorithmes proposés, néanmoins l'interaction entre les piétons ainsi que l'évolution au cours du temps des probabilités du conflit de trafic ne sont pas étudiés.

Ce travail est étendu par [75] et un système complet basé sur la vision est mis en place fournissant des résultats expérimentaux sur les données vidéo du monde réel. En fait, cet article présente un cadre probabiliste complet pour l'analyse automatisée de la circulation dans une optique de la sécurité routière. S'appuyant sur les techniques de conflit du trafic et le concept de hiérarchie de sécurité, il fournit des définitions informatiques de la probabilité de collision pour les usagers de la route impliqués dans une interaction. Des mesures d'exposition nouvelles et plus pertinentes ont été dérivées de ce travail et des conflits du trafic ont été détectés sur les bases de données Minnesota et Italy [75].

1.5.4 Conclusion sur les techniques d'études des conflits du trafic et prédiction du danger

Les techniques d'étude des conflits du trafic et la prédiction du danger se basent sur une stratégie d'analyse des conflits, intégrant un ou plusieurs indicateurs pour la conception d'un système de protection des piétons. Ces techniques présentent des résultats satisfaisants et un raisonnement convainquant d'analyse conflit P-V. Néanmoins, il existe toujours des verrous scientifiques et technologiques qui ne sont pas encore beaucoup étudies dans le domaine de la recherche, notamment la quantité et à la qualité des données de collision. Les informations sur les collisions sont souvent non détaillées, non précises, voire même manquantes et incomplètes. Ces données conduisent à des prédictions et identifications des collisions imprécises. L'audit de collision est effectué après l'apparition d'une collision, auquel cas les causes, la localisation spécifique et les aspects comportementaux de l'événement sont soumis à un jugement.

Un jugement reste une constatation avec une part de subjective qui ne peut pas être totalement et exacte, surtout s'il se base sur des données non complètes et/ou manquantes. Ces lacunes laissent la question d'analyse de conflits du trafic et du comportement piétonnier lors de la traversée de la rue très ouverte sur d'autres amélioration, que ce soit au niveau de la qualité et de la quantité de données et indicateurs à collecter et/ou des stratégies et méthodes de corrélations et d'analyse de ces données ou encore d'indicateurs à calculer. Notre contribution à cette problématique sera progressivement développée durant les différents chapitres de cette thèse (chapitre 2, 3 et 4 qui portent sur la quantité et la qualité des données à collecter et le chapitre 5 détaille la stratégie de corrélation et d'analyse des conflits du trafic).

En outre les systèmes de protection des piétons à base de détection et/ou prédiction de choc se basent généralement sur des capteurs spécifiques et complexes implémentés au seins du la véhicule pour offrir une interaction automatisé entre le conducteur et l'environnement. Ces systèmes sont des développés totalement ou partiellement par quelques compagnie de construction de véhicules. Aussi, pour jouir de ces systèmes il faut des dépenses importants pour avoir une voiture équipé de ce système. Ceci renforce l'intérêt potentiel pour l'exploitation des matériels de grande public tel que les caméras pour la mise en œuvre d'un système de protection des piétons et d'analyse des conflits du trafic.

1.6 Synthèse de l'existant et identification des contributions potentielles : Objectifs et hypothèses de travail

En raison des limites de la collecte manuelle des données, une tendance croissante à l'utilisation de systèmes automatiques de collecte de données a pris de l'ampleur dans le domaine de l'ingénierie du transport. En particulier, l'analyse automatisée des vidéo suscite un intérêt croissant. Les capteurs vidéo sont maintenant largement disponibles (des caméras de trafic sont déjà installées sur de nombreuses routes) et sont relativement peu coûteuses.

En outre, les études classiques sur l'accidentologie et l'analyse des risques d'accident de la circulation traitent souvent séparément les usagers de la route, les véhicules et l'environnement [76]. Les accidentologistes ont également tendance à ne s'intéresser qu'à quelques indicateurs de conflit P-V ou à un seul indicateur. Cependant, le contexte de la circulation routière et de la traversée d'une rue est caractérisée par l'interaction de plusieurs paramètres [76] tels que les piétons, les véhicules, les cyclistes, les motocyclistes, les aménagements urbains, les signalisations de la route... Le fait de privilégier certains facteurs par rapport à d'autres peut limiter l'efficacité des efforts de prévention et peut conduire à prendre des mesures de sécurité qui ne protègent pas convenablement les piétons. Pour les piétons, l'environnement dans lequel ils se déplacent et la sécurité sont étroitement liés. D'où l'importance d'étudier le comportement des piétons lors de la traversée de la rue en fonction des variables de l'environnement (trafic et aménagement).

De plus, de nombreuses études d'accidentologie et des bilans de sécurité routière montrent le pourcentage important des personnes âgées dans la mortalité piétonnière des accidents de la route. Cette population, pourtant très fragilisée dans l'environnement urbain, a été beaucoup moins étudiée et presque aucune intervention n'est notée pour prendre des mesures de sécurité spécifiques à cette catégorie d'âge. Toutefois, les analyses des rapports des études d'accidentologie disponibles ont montré des différences entre le comportement des personnes âgées et celui des personnes jeunes adultes lors de la traversée de la rue. Les paramètres les plus discriminants sont 'le temps de prise de décision', 'les marges de sécurité' et 'les opportunités manquées' [27]. En fait, les problèmes de santés dont souffrent les personnes âgés à cause de l'avancée en age et des effets néfastes du vieillissement tels que la diminution de la souplesse articulaire, la diminution de la sensibilité visuelle au mouvement et l'affaiblissement des capacités intentionnelles, conduisent à une distorsion dans la perception du temps et la manipulation des gaps de sécurité qui peuvent être considérées comme un facteur responsable d'erreurs. Ces biais peuvent donc mener les personnes âgées à ne pas percevoir le danger qu'elles encourent lors de la traversée d'une rue.

Ce travail tend à remédier à certaines lacunes antérieures et à dégager des prospections de recherche. Cette thèse traite le développement et le test d'un système automatisé d'analyse vidéo pour contribuer à satisfaire les objectifs variés de la multidisciplinarité de l'analyse comportementale des piétons au profit de l'accidentologie. Les effets du vieillissement, les effets de l'environnement et la prise en compte des différents indicateurs de conflit du trafic se révèlent complémentaires et corrélés pour une analyse objective et détaillée du comportement du piéton lors d'une traversée de la rue. Pour cela, trois objectifs généraux seront abordés puis déclinés en sous objectifs plus spécifiques. Il s'agit de :

1. Collecte automatique de données du trafic et piétonnières :
   - Conception d'un outil automatique de collecte de données à base de techniques de vision par ordinateur à partir de matériels de grand public sans aucune utilisation de capteurs spécifiques.
   - Détection et suivi des usagers de la route dans une scène de circulation routière et classement entre piétons et véhicules.
2. Corrélation des paramètres des piétons et du trafic pour définir les événements et les potentiels indicateurs de conflit P-V.
   - Conception d'un système de mise en relation des paramètres multiples, hétérogènes, pouvant contenir des données manquantes, discrètes,....
   - Modélisation et mise en œuvre des relations entre les variables (vieillissement, aménagement, trafic, piéton).
   - Identification des événements importants dans une séquence vidéo. La définition d'un événement important dans cette étude est "tout événement impliquant un passage piéton et un véhicule conflictuel dans lequel il existe une chaîne imaginable d'événements qui pourrait conduire à une collision entre ces usagers de la route".
   - Proposer des mesures objectives des indicateurs de sévérité pour tous les événements.

3. Prédiction et inférence du type de comportement des piétons en fonction des variables collectées précédemment.
   — Détermination du comportement à risque ou sécuritaire du piéton.
   — Identification de l'adaptation du piéton pour éviter un danger.

1.7 Conclusion

Durant ce chapitre nous avons présenté la mortalité et la mobilité des accidents routières et leurs impacts socio-économiques qui s'illustrent notamment par le grand nombre de décès de piétons et surtout pour les personnes âgées. Ces chiffres représentent une motivation majeur pour de nouvelles études et mise en place de certaines politiques par les états afin de contribuer à cette problématique de santé publique. A l'issu de cette étude panoramique des sciences de l'accidentologie nous notons les difficultés méthodologiques liées à l'étude des traversées de rue en situation réelle et la prise en considération de différents facteurs et enjeux. Ces difficultés ont mis en avant l'analyse des facteurs de risque, l'origine des accidents et l'interaction comportementale entre les usagers de la route ainsi que la prise en considération limitée de l'effet de l'environnement. En fait, des analyses d'accidents sont menées sur la base de scénarios typiques et de méthodologies basiques (manuelles, statistiques, simulateurs...) qui ont parfois des limites et sont éloignées de la variabilité des conditions de terrain. De ce fait, beaucoup d'enjeux et questions soulevés restent ouverts afin d'améliorer la sécurité des piétons.

Ce sont notamment les progrès de la vision par ordinateur qui ont permis et permettent encore aujourd'hui l'avancée de l'ITS et de leurs applications. Ces constations représentent alors les motivations du plan de travail de ce projet de thèse. Les détails de ce plan et des hypothèses étudiées seront présentées durant les chapitres suivants qui constituent les différentes étapes de conception d'un système automatique d'analyse comportementale des piétons âgés lors de la traversée d'une rue.

Chapitre

2 Collecte automatique de données 'microscopiques' de piétons

2.1 Introduction

Le piéton est un élément pivot pour toute étude d'accidentologie et de sécurité routière [77]. Son comportement influence l'état du trafic. Ainsi, la caractérisation du comportement aléatoire du piéton nécessite la collecte d'une base de données énorme et diversifiée. Les données piétonnières à colleter doivent intégrer les paramètres spatio-temporels des piétons durant la traversée de rues. Les opérations manuelles pour l'analyse vidéo semi-automatisée peuvent être laborieuses et limitées en termes de volume de données à analyser par rapport aux méthodes automatisées [78, 79, 54]. L'analyse automatisée de vidéo implique l'utilisation de techniques de vision par ordinateur et peut surmonter certains lacunes associées aux observations manuelles, sur le terrain et à l'analyse manuelle de la vidéo comme par exemple le manque de précision et la difficulté du suivi de réalité des comportements aléatoires.

Afin, de faire face à la demande croissante d'études du comportement des piétons, à la nécessité de tenir compte des changements dans les caractéristiques des piétons et d'améliorer la sécurité des usagers de la voie publique, les techniques automatisées doivent être développées. L'objectif principal de ce chapitre est de documenter le développement et les tests d'un prototype de système à base de vidéo capable de suivre les piétons, d'extraire leurs trajectoires et de les analyser pour collecter les paramètres caractérisant le mieux leurs comportements lors de traversée d'une rue à partir de trafics réels.

Dans ce chapitre, nous commencerons par une présentation générale du processus proposé pour la collecte automatique des données piétonnières à base de traitement bas niveau des séquences vidéo. Une analyse plus détaillée des principales techniques

de détection et de suivi de personnes dans des séquences d'images sera ensuite réalisée. Nous clôturerons cette partie par l'explication et l'évaluation de la méthode proposée pour le suivi multi-objets. La seconde partie de ce chapitre concerne la phase d'analyse des trajectoires des piétons et l'extraction des données microscopiques pour caractériser, ultérieurement, leur comportement.

2.2 Processus proposé : description générale

L'outil automatique de collecte de données de piétons consiste à exploiter les techniques de vision par ordinateur telles que la détection et le suivi pour générer finalement un ensemble de paramètres permettant de caractériser l'interaction entre les piétons et le trafic routier. Cet outil est composé de 3 grandes étapes inter-connectées (Figure 2.1).

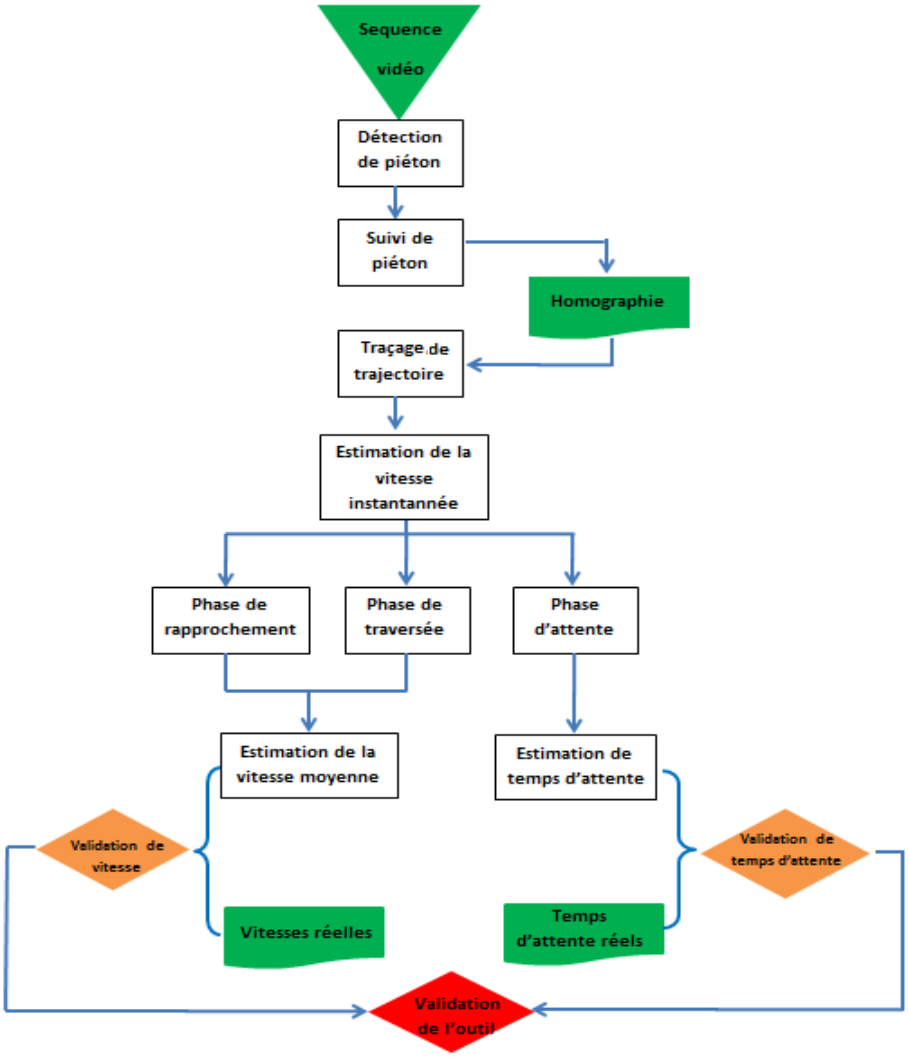


FIGURE 2.1 – Architecture générale de l'outil automatique de collecte de données piétonnières proposé.

— Détection et suivi de piétons.
— Modélisation de la zone de traversée (passage piéton) et identification des trajectoires.
— Collecte de paramètres (vitesse instantanée, phases de traversée, vitesse moyenne, temps d'attente...).

L'enchainement de ces étapes est expliqué par l'algorithme (1) qui a comme information d'entrée une vidéo, séquence de traversée d'une rue capturé au niveau d'un passage piéton, utilisé comme une région d'intérêt. L'algorithme génère comme résultat un ensemble de paramètres microscopiques des piétons calculés par l'analyse spatiotemporelle des trajectoires.

Entrée : vidéo séquence de traversée d'une rue
Sortie : vitesse instantanée, vitesse moyenne par phase de traversée, temps d'attente
1. lancer le processus de suivi de personne(s)
1.1 Detecter le(s) piéton(s) pour chaque image
1.2 Associer l'identité correcte à la personne
1.3 Filtrer la trajectoire du (des) piéton(s)
2. Tracer la trajectoire du (des) piéton(s)
2.1 Calculer la matrice d'homographie "H"
2.2 Déterminer les positions (x, y) du (des) piéton(s) en coordonnées réelles
3. Estimer la vitesse instantanée du (des) piéton(s)
4. Détecter les phases de traversée
5. Estimer X : la vitesse moyenne durant chaque phase
6. Estimer le(s) temps d'attente
7. Mesurer Y : les données de réalité de terrain
8. Calculer les valeurs de R et de $RMSE$

Algorithme 1: Estimation automatique de la vitesse du (des) piéton(s) et des phases de traversée.

2.3 Détection et suivi de personnes

La construction d'une base de données de grande taille, diversifiée et caractérisant le comportement aléatoire du piéton en déplacement peu passer par l'intégration de techniques d'analyse de vidéo et de vision par ordinateur. Cela est rendu possible par les progrès très importants au niveau de la précision des caméras, d'une part et de la puissance de calcul des ordinateurs, d'autre part. En effet, la détection, la quantification et le suivi du piéton durant la traversée se révèle une tâche primordiale. Ces derniers, constituent un domaine de recherche très actif et attire beaucoup d'attention dans la communauté de la vision par ordinateur. En outre, il existe de nombreuses méthodes

dans la littérature pour la détection et le suivi des objets en général mais aussi des personnes en particulier.

2.3.1 Approches existantes pour la détection de personnes

La détection d'objets dans des images et des vidéo est l'objet de beaucoup d'attention dans la communauté de chercheurs en vision par ordinateur et en reconnaissance de formes ces dernières années, du fait des nombreuses applications potentielles. Dans ce qui suit nous présentons une rapide revue de l'état de l'art des techniques de détection et de localisation automatiques d'objets, en accordant une attention particulière à la détection humaine. En fait, la détection de personnes consiste à différencier la personne des autres objets dans la scène. Dans ce contexte, plusieurs travaux ont été proposés dans la littérature. Ils peuvent être divisés en 3 catégories suivant le type de caractéristiques utilisées : i) Approches basées sur la couleur de la peau, ii) Approches basées sur le mouvement et iii) Approches basées sur la forme.

2.3.1.1 Approches basées sur la texture et la couleur de la peau

Ce type d'approche discrimine les personnes et tout autre objet dans la scène par la recherche de l'occurrence que ce soit de la couleur et/ou de la texture de la peau de l'être humain [80, 81]. En effet, la chaleur de la peau humaine est une caractéristique très discriminante de la personne. Cependant, ces caractéristiques ne sont pas toujours disponibles à cause de la variation de point de vue et de l'aspect vestimentaire de la personne.

2.3.1.2 Approches basées sur le mouvement

La façon de se déplacer ou la démarche est une action spécifique et propre à chacun. En effet, la personne se déplace avec une cadence, un mouvement et un rythme qui permet de la distinguer de tous autres objets en mouvement. En se basant sur ces observations, une des approches détecte la personne en exploitant son mouvement global. Pour cela, l'approche proposée dans [82] utilise le flot optique, qui se base sur les modèles de mouvements humains pour une détection robuste, car il est relativement indépendant de l'apparence et des facteurs environnementaux. La technique est basée sur la Machine à Vecteur de Support (SVM) avec un noyau de fonction de base de référence (RBF) formée sur les diagrammes de flux optique pour créer un classifieur humain. D'une autre coté [83] extrait l'objet en mouvement par la soustraction de l'arrière-plan, puis une ellipse est utilisée pour filtrer les objets d'avant plan. En effet, la détection de la personne est faite à base d'une heuristique : la partie supérieure de l'ellipse englobant une personne se déplace légèrement avec une faible fluctuation vers l'avant. D'autres approches utilisent le mouvement local des différentes parties de l'objet pour la discrimination de la personne [83]. Un suivi est d'abord réalisé pour former un squelette de l'objet étudié (un modèle). C'est sur le mouvement des composantes de ce modèle que portera l'étude (suivi des oscillations des jambes et de la variation du pas). La démarche de la personne est périodique [84]. Il s'agit d'une alternance entre deux positions, la position de simple support (la personne a un seul pied d'appuis) et

la position de double support (la personne à ces deux pieds en contact avec le sol). Par conséquent, pour détecter la personne par le mouvement, certaines approches cherchent à caractériser cette périodicité par des vecteurs de mouvement [85].

Ces approches sont dépendantes de l'orientation du sujet : point de vue (en vue frontale, il sera presque impossible de caractériser le mouvement). Aussi, l'utilisation de mouvement nécessite que la détection se fasse sur plusieurs images successives. Généralement, il faut plus d'un cycle de déplacement pour que le système détecte et apprenne la périodicité. Même si les modèles de mouvement sont robustes, face à la variation d'échelle et d'apparence, ce procédé n'est pas approprié pour détecter des humains partiellement occlus.

2.3.1.3 Approches basées sur la forme

L'origine de l'utilisation de la forme pour la détection des objets est les travaux [86, 87, 88], qui ont utilisé un vote binaire de l'appartenance des contours aux classes de l'histogramme dans l'espace log-polaire sans aucune prise en considération de l'orientation des contours [89]. Une amélioration de cette approche est proposée par [90] qui intègre l'orientation sur une seule classe des gradients avec un regroupement circulaire des ensembles de pixels. Un autre type d'approches de forme caractérise la personne par un modèle morphologique englobant les différentes poses prises par le corps humain [90, 91, 92, 93] tels que le modèle squelettique, le modèle de cylindres et le modèle elliptique (Figure 2.2). La localisation de la personne est effectuée par le calcul d'une probabilité de ressemblance entre le modèle recalculé et le modèle prédéfini [91].

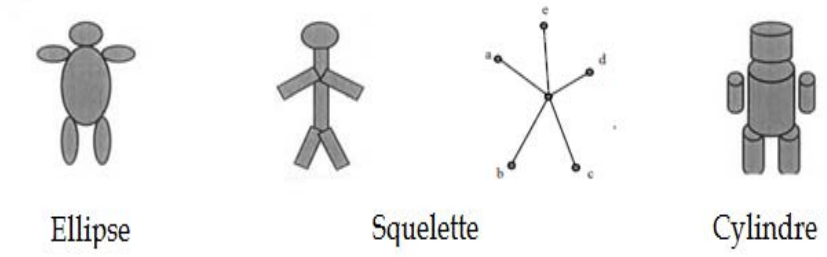


FIGURE 2.2 – Modèles morphologiques de la personne.

La forme et l'apparence de la personne sont très sensibles à la variation de poses et de point de vue de la caméra. La détection de la personne est une tâche très complexe vue les difficultés qui caractérisent l'analyse et le traitement d'images en générale et la caractérisation d'un objet non rigide et déformable tel qu'une personne.

En effet, parmi les problématiques les plus présentes dans ce domaine nous pouvons citer :

— La variation de point de vue, en effet la moindre variation de l'orientation ou de la position de la caméra peut changer totalement l'apparence de l'objet cible.
— La variation de l'apparence, car la plupart des classes d'objets naturels ont de grandes variations intra-classe. Par exemple, pour les humains, l'apparence et la pose changent considérablement entre les images de la même classe et les différences de vêtements rendent la variation plus intense. Un détecteur robuste doit essayer d'obtenir l'indépendance de ces variations.

— La couleur de l'objet et l'éclairage général peuvent varier considérablement. Par exemple, la lumière directe du soleil et les ombres pendant la journée (éclairage artificiel ou faible la nuit). Bien que les modèles d'invariance de couleur et d'illumination aient fait des progrès significatifs, ils sont encore loin d'être des solutions efficaces.
— La grande variabilité de l'arrière-plan et son influence directe sur la qualité de la détection, surtout dans un environnement extérieure très variable comme des arbres exposés au vent par exemple.

Prenant en considération toutes les difficultés citées précédemment, plusieurs approches utilisent soit la cascade et la fusion de descripteurs, soit des classifieurs pour l'implémentation d'un détecteur de personnes plus performant. En effet, [94] fusionnent les paramètres de mouvement et de forme pour une bonne discrimination entre la personne et le véhicule. Ils supposent comme postulats qu'une personne et un objet dont la hauteur est supérieure ou égale à $2 \times \ largeur$, mais aussi que sa concentration ($\frac{perimetre^2}{aire}$) et sa minceur ($\frac{projection\ verticale}{projection\ horizontale\ du\ squelette}$) sont plus importantes que celles du véhicule (Figure 2.3) [94].

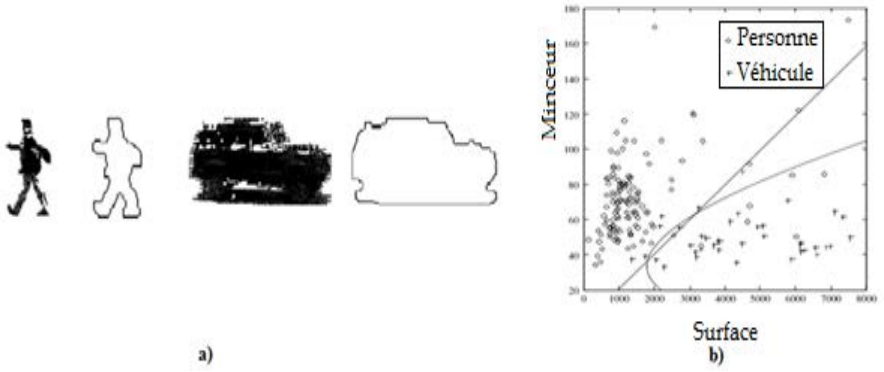


FIGURE 2.3 – Illustration de la discrimination entre une personne et un véhicule [94].

La fusion de descripteurs est introduite par [95]. Ce travail fusionne le mouvement (image de différence) et l'apparence (différence d'intensité) avec Adaboost. D'autres approches adoptent une cascade de classifieurs en se basant sur l'hypothèse qu'une série de classifieurs peut être plus efficace qu'un seul [95, 96, 97].

2.3.2 Approches existantes pour le suivi de personnes

Pour suivre une personne, la majorité des travaux utilisent les méthodes statistiques telles que les modèles de mouvement, comme le filtre de Kalman [98] et le filtre de particules [99]. Ces filtres estiment récursivement l'état actuel, basé sur l'état passé et effectuent une phase de correction des prédictions sur la base des positions réelles. En fait, le filtre de particules surmonte la limitation de mouvement linéaire du filtre de Kalman. Mais dans le cas de suivi multi-objets, le nombre de particules requises augmente considérablement et devient difficile à utiliser dans des situations pratiques.

En outre, les modèles de mouvement ne peuvent pas gérer le chevauchement des trajectoires et la confusion d'identité car ils n'ont aucune information pour la discri-

mination intra-classe. Pour ces raisons, certaines approches proposent de fusionner les modèles de mouvement et d'apparence. En fait, le suivi par un modèle d'apparence consiste à trouver une mesure de similarité entre les noyaux successifs des frames (i) et $(i+1)$. La similarité est calculée en se basant sur l'une des caractéristiques suivantes : la couleur [100], la texture [101] et/ou la forme [102]. Ces mesures de similarité 'dopent' le modèle d'apparence utilisé. Nous pouvons citer, à titre d'exemple, la distance Euclidienne et la distance de Bhattacharyya [103].

Le modèle d'apparence le plus populaire utilisé dans la littérature est l'histogramme de couleurs. Dans ce contexte l'approche proposée dans [104] utilise le filtre de Kalman pour prédire la position des piétons et l'histogramme de couleurs 4D, présenté dans l'espace de couleurs 'Lab', pour assurer le couplage d'identité. Puisque l'histogramme ne représente que la répartition des couleurs, deux piétons portant des vêtements de couleur(s) semblable(s) ont une probabilité importante d'être confondus. Dans [105], les auteurs proposent de combiner des histogrammes de couleur définis dans l'espace 'HSV (Hue Saturation Value)' et un filtre à particules. Étant donné que l'espace de couleurs 'HSV' est significativement plus puissant, en termes de représentation des couleurs, par rapport aux autres, les performances du tracker sont améliorées. L'inconvénient de cette méthode est le coût en temps de calcul important du filtre à particules.

La solution adoptée par les travaux récents [104, 106, 107] est d'associer les détections durant la séquence vidéo. En fait, dans [106], les auteurs proposent un modèle de trajectoire continue basé sur des B-splines cubiques. Pour cela, ils doivent connaître les positions de départ et de fin. De plus, pour associer les détections inter-frames, cette méthode utilise uniquement la distance entre des positions consécutives qui reste un inconvénient majeur, en particulier lors du chevauchement des trajectoires et des changements de direction. Cependant, dans [107], les auteurs basent leur méthode sur les fondements de la théorie de l'estimation bayésienne. Ils ont seulement la visibilité sur l'objet courant et non sur les autres objets.

2.3.3 Approche proposée : suivi multiple basé sur la maximisation de similarité de spatiogramme

A l'issu de ce rapide état de l'art de la détection et du suivi de personnes, les approches sont classées selon les techniques utilisées en 3 catégories : (i) Approches basées sur la forme, qui cherchent à caractériser la personne par un modèle morphologique englobant les différentes poses prises par le corps humain, (ii) Approches basées sur un modèle de mouvement, exploitant la périodicité du déplacement des personnes, et (iii) Approches basées sur la textures et la couleurs, caractérisant la personne par l'occurrence de textures et de couleurs de la peau. Suite à une analyse comparative entre ces différentes approches (cf partie 2.3.1), nous avons opté pour l'utilisation du descripteur de HOG (Histogramme of Oriented Gradian) par blocs de 2×2 cellules pour caractériser une personne, avec une étape de classification par SVM. Cette approche est basée sur l'apparence et la forme pour tenir compte de la nature très articulée du corps humain. Nous avons aussi proposé une méthode d'interpolation inter-frames permettant d'améliorer l'étape de détection. Cette interpolation est relative à la trajectoire ainsi qu'à la vitesse de déplacement du piéton. Une fois que la personne est localisée, nous

cherchons à la suivre à l'approche de la traversée d'une rue et durant la traversée. En fait, pour suivre un objet en mouvement, on peut distinguer deux grandes catégories d'approches : (i) des modèles de mouvement tels que les filtres de Kalman et les filtres particulaires et (ii) des approches de ré-identification basées sur l'apparence à savoir les histogrammes de couleur, les spatiograms et la combinaison de textures et de couleurs (cf partie 2.3.2). Tenant compte de notre contexte de travail nous avons opté pour le suivi par association de détections avec un module de ré-identification. Comme on est dans un contexte de suivi multiple où les personnes peuvent se chevaucher et être perdues ou invisibles pour quelques images, nous proposons d'utiliser un modèle d'apparence, tel que le spatiogramme, pour le calcul de similarité et la ré-identification des personnes.

2.3.3.1 Détection continue : frame par frame

2.3.3.1.1 Histogrammes de Gradient Orienté (HOG)

L'idée générale du descripteur HOG consiste à caractériser l'objet cible à la fois par l'apparence et la forme. Une telle représentation répond convenablement à la spécificité de la personne, qui est un objet très articulé, dont la forme varie considérablement à cause du mouvement et de l'inclinaison de la tête, des bras et des jambes, surtout lors du déplacement. En effet, une simple caractérisation par un modèle morphologique fixe peut, ne pas détecter la personne dans leurs positions multiples. Les HOG par blocs se basent sur l'hypothèse que la morphologie locale de la personne peut rester semblable, même si la morphologie globale est modifiée. Par la suite, le principe de HOG est de calculer un vecteur caractéristique global, par la fusion de plusieurs HOGs calculés localement sur des cellules de pixels, puis de les regrouper et de les normaliser par blocs de cellules contiguës. Les étapes de calculs du descripteur HOG [89] sont illustrées par la figure 2.4 et détaillées comme suit :

1. La première étape consiste à réduire l'influence des effets d'illumination et les variations d'éclairage. Pour cela, une égalisation globale des images est appliquée par l'intermédiaire de la compression gamma (power law) en calculant la racine carrée de chaque canal de couleur.
2. La deuxième étape vise à produire un encodage qui est sensible au contenu d'image local tout en restant résistant aux petits changements de pose ou d'apparence. La méthode adoptée regroupe les informations d'orientation de gradient local. En premier lieu, la fenêtre de recherche (de taille $[64 \times 128]$) est divisée en petites régions spatiales de 3×3 pixels, appelées "cellules". Pour chaque cellule, nous accumulons un histogramme $1D$ local des orientations de gradients de tous les pixels de la cellule. Chaque histogramme d'orientation divise la plage d'angles de gradients en un nombre fixe de cases prédéterminées. Deux cas sont ensuite distingués, le cas de $[0, 180]$ et le cas de $[0, 360]$. Dans le cas de la présence de personne, la plage d'angle testée et estimée dans [89] à $[0, 180]$. Les grandeurs de gradients des pixels dans la cellule sont utilisées pour voter dans l'histogramme d'orientation suivant le nombre de classes (bin) utilisés ($9\,bins$). Le masque $[-1\,0\,1]$ est utilisé pour le calcul de gradient.

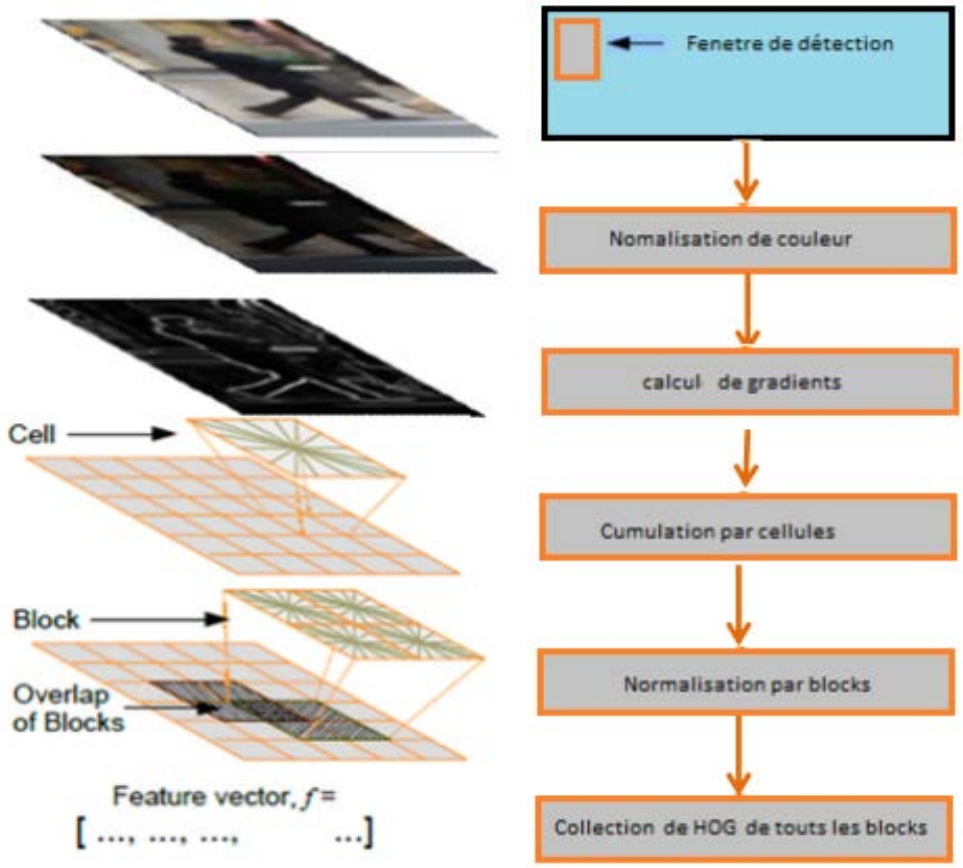


FIGURE 2.4 – Illustration des étapes d'extraction des Histogrammes de Gradient Orientés : cas de la détection d'une personne [89].

3. La troisième étape est une étape de normalisation pour une meilleure correction des effets de luminance et de changements d'éclairage, ainsi que des contrastes de bord. En fait, les cellules locales déjà générées sont regroupées en blocs de 2×2 cellules chevauchantes. Par conséquent, une seule cellule peut appartenir à plusieurs blocs. Cela peut créer une sorte de redondance, mais les expérimentations de [89] ont montré que cette méthode améliore considérablement les performances. La phase de normalisation est assurée par le $L2_{Hys}$ [89].
4. L'étape finale consiste à regrouper les descripteurs HOGs de tous les blocs couvrant la totalité de la fenêtre de recherche en un seul vecteur caractéristique. Ce dernier sera passé au classifieur SVM pour déterminer si la fenêtre contient ou non une personne.

2.3.3.1.2 Interpolation

Le HOG est robuste vis-à-vis de la variation de luminosité, de conditions d'éclairage, de contrastes et du fait qu'il contourne convenablement la forme complexe et variable du corps humain par une représentation de l'apparence et de la forme locale de l'objet via la distribution locale des intensités de gradients ou de directions des contours. Cependant,

l'occultation partielle et/ou totale reste l'inconvénient majeur de ce descripteur. Pour pallier à cet inconvénient, nous proposons d'interpoler les détections manquantes pour un ou plusieurs frames. En fait, nous cherchons la position de piéton(s) perdu(s) par un processus d'interpolation basé sur des hypothèses tenant compte de la vitesse moyenne et le déplacement du piéton (un pas mesure environ $0,65\,m$, il avance de $2,05\,pas.s^{-1}$ pour une vitesse moyenne de $1,33\,m.s^{-1}$ et la distance avec le prédécesseur est d'au moins 1 m [108]). Ainsi, un piéton fait approximativement un déplacement moyen de $0,053m$ par frame. Par conséquent, nous modélisons le processus d'interpolation comme une translation, avec un vecteur $\vec{u}$, des dernières détections comme le montre la figure 2.5.

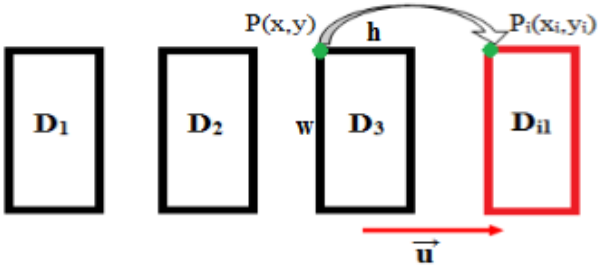


FIGURE 2.5 – Interpolation des détections manquantes.

En fait, la boite englobant le piéton se caractérise par un quadruplé (x, y, w, h) avec (x, y) les coordonnées du point de départ, alors que $'w'$ et $'h'$ sont respectivement la largeur et la hauteur de la boîte englobante. La translation consiste à trouver l'image $Pi(x_i, y_i)$ du point $P(x, y)$ dans la même direction et en utilisant un vecteur $\vec{u}$, ayant comme norme la distance moyenne calculée par rapport à la vitesse moyenne, et en gardant les mêmes dimensions w et h.

Pour illustrer les performances du détecteur de personnes adopté, des évaluations ont été faites sur des images collectées pour des personnes en train de traverser une rue. Ces images incluent des difficultés à plusieurs niveaux à savoir, l'éclairage, le point de vue, la distance des personnes par rapport à la caméra ainsi que le nombre de personnes disponibles dans la scène. La figure 2.6 présente des exemples de détections réalisées suite à l'application de processus proposé : HOG + Interpolation.

Les performances du processus proposé (HOG + Interpolation) ont été testées en termes de détection de personnes vis-à-vis de plusieurs contraintes : variation de luminosité (personne dans des zones claires (Figure 2.6 b, c et d) et d'autres sombres (Figure 2.6 a), variations de point de vue (vues frontale, fronto-parallèle (Figure 2.6 c), de face (Figure 2.6 b), de dos (Figure 2.6 d et b), variation de distances par rapport à la caméra (loin, proche (Figure 2.6 c et d) et nombre de personnes (Figure 2.6).

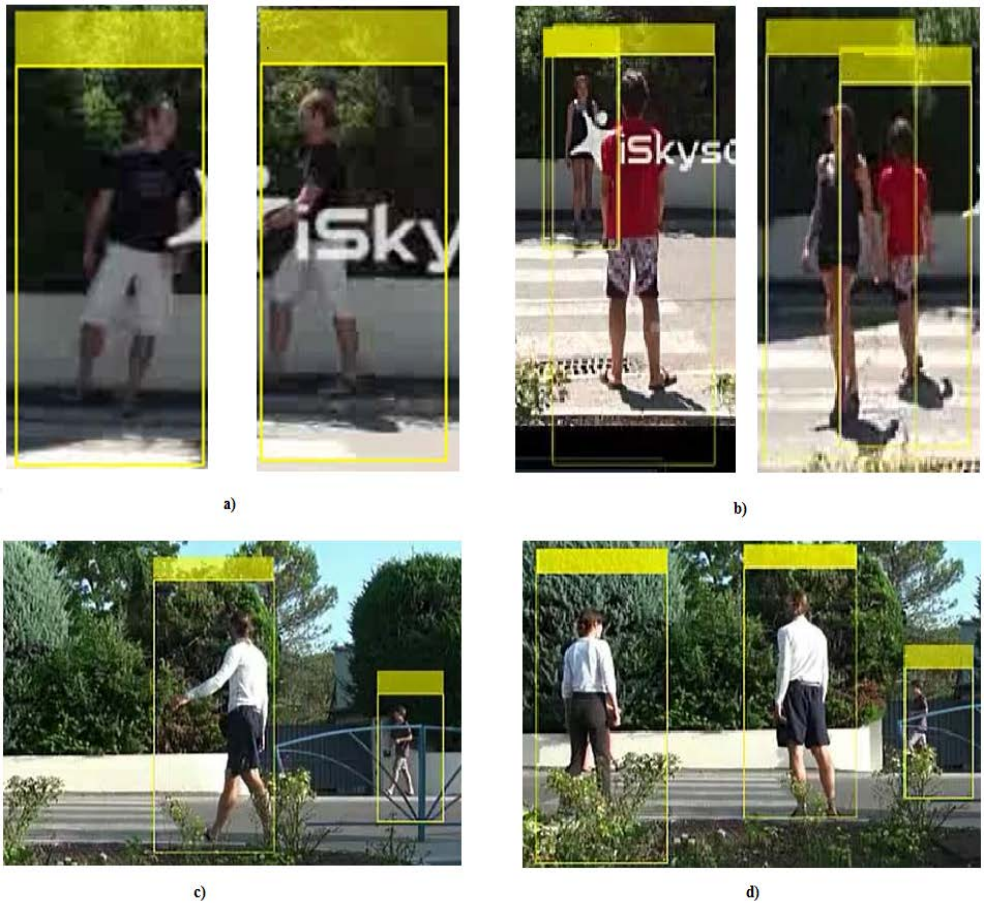


FIGURE 2.6 – Illustrations de détections à partir du processus HOG + interpolation.

2.3.3.2 Suivi de personnes par l'association de détections avec utilisation d'une métrique de similarité de spatiogramme

2.3.3.2.1 Associations de détections

Parmi les problèmes les plus critiques pour le suivi de plusieurs objets dans une scène, on retrouve le chevauchement des objets et la ré-identification de l'objet après sa perte, soit à cause de l'occultation inter-objets ou par un obstacle dans la scène, soit à cause d'intersection de trajectoires et de changement de direction (problème de 'split' et 'merging' [109]). Dans le contexte de traversée d'une rue, on rencontre fréquemment ce genre de problèmes puisque l'on est dans un environnement extérieur avec l'existence de plusieurs obstacles (voiture, arbre, bâtiments) ainsi que d'un flux de piétons pouvant se chevaucher en traversant la rue dans les deux directions.

Par conséquent, notre stratégie de suivi se base sur l'association de détection avec un module d'association et de ré-identification tel que le spatiogramme. En effet, les détections déjà faites dans l'étape précédente sont organisées sous forme d'une arborescence (Figure 2.7). Chaque frame de la séquence vidéo représente un niveau dans l'arborescence. Chaque détection est, initialement, inter-connectée avec toutes les détections dans le niveau suivant par des arcs ayant comme coût la valeur de similarité

entre eux. Ensuite, la détermination de la trajectoire du piéton est faite en trouvant le chemin qui maximise la similarité entre les détections dans les différents niveaux de l'arborescence. Comme métrique de similarité nous avons choisi la distance entre les spatiogrammes [110] des objets détectés.

Dans le processus d'association, chaque détection au niveau $'i'$ devrait être associée à une et une seule détection au niveau $'i+1'$. Ainsi, une nouvelle personne entrant dans la scène peut être simplement détectée et intégrée dans le système de suivi, étant donné qu'elle n'est pas similaire à toutes les autres détections précédentes, mais qu'elle est similaire à la suivante. Cependant, dans le cas d'une détection erronée, elle sera supprimée du niveau suivant, car elle n'aura pas de similarité dans la suite de la séquence.

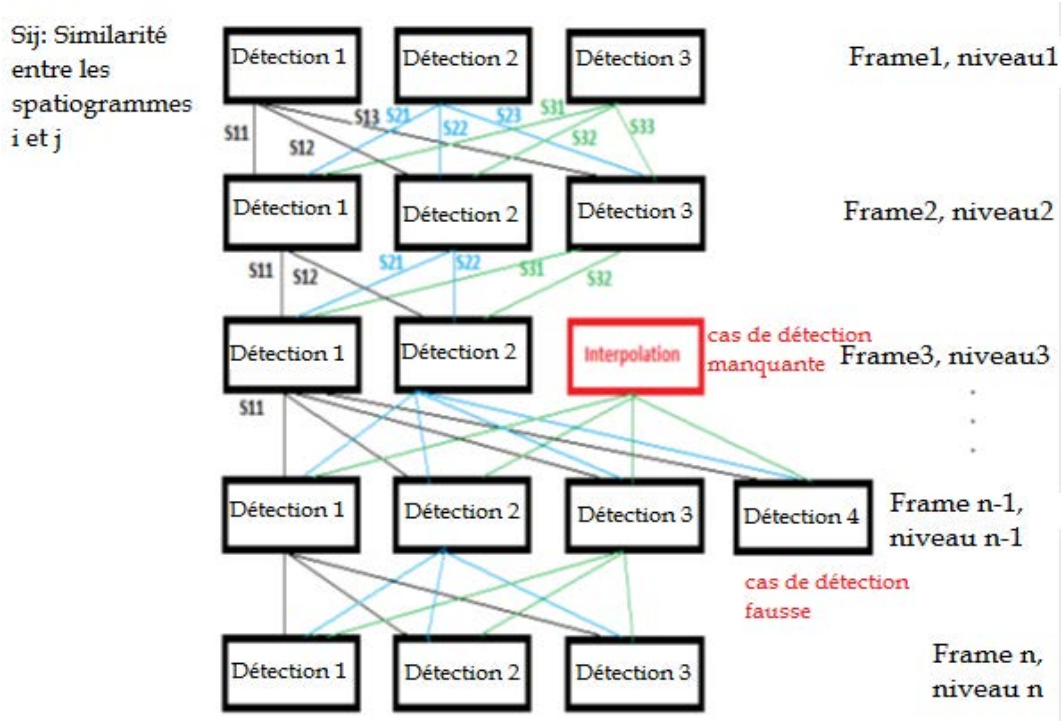


FIGURE 2.7 – Association des détections [111].

2.3.3.2.2 Calcul de spatiogramme et mesure de similarité

Selon une étude comparative entre les descripteurs de couleurs, de textures et de formes pour l'identification de plusieurs personnes [110], le spatiogramme présente les meilleures performances. En effet, le spatiogramme représente une extension de l'histogramme. Il contient l'information de la distribution des couleurs et leurs répartitions spatiales. Par conséquent, nous proposons d'utiliser le spatiogramme comme modèle d'apparence pour vérifier la similarité entre les détections. En particulier, nous utilisons celui de Conaire [112] vu son pouvoir représentatif montré dans [110]. Pour cela, une phase de quantification est tout d'abord effectuée sur l'image originale pour réduire la variation de couleur. Deuxièmement, un histogramme de couleur (noté $'\eta'_b$)

est calculé, qui correspond à l'attribution d'un nombre de pixels à chaque intensité de couleur. L'algorithme calcul alors, pour chaque intensité existante dans l'image quantifiée, sa répartition spatiale représentée par la moyenne $'\mu'$ et la covariance $'\sigma'$ de la position spatiale de tous les pixels qui appartiennent à la même classe d'intensité de l'histogramme tel que formalisé dans l'équation 2.1.

$$\eta_b = \sum_{i=1}^{N} \delta_{ib} \tag{2.1}$$

$$\mu b = \frac{1}{\sum_{j=1}^{N} \delta_{jb}} \sum_{i=1}^{N} X_i^T \delta_{ib}$$

$$\sigma_{ib} = \frac{1}{\sum_{j=1}^{N} \delta_{jb}} \sum_{i=1}^{N} (X_i - \mu_b)(X_i - \mu_b)^T \delta_{ib}$$

Avec $\delta_{ib} = 1$ si le pixel i appartient à la classe d'intensité $'b'$, 0 si non et $X_i = (x_i, y_i)$ est la position spatiale de chaque pixel. Pour comparer deux spatiogrammes $S(\eta, \mu, \sigma)$ et $S'(\eta', \mu', \sigma')$ contenant chacun B classes, une mesure de similarité est calculée comme formalisé à l'équation 2.2.

$$similarite = \sum_{b=1}^{B} \psi_b \sqrt{\eta_b \eta_b'} \tag{2.2}$$

Avec

$$\psi_b = \eta exp(\frac{-1}{2}(\mu_b - \mu_b')\sigma_b'^{-1}(\mu_b - \mu_b'))$$

$$\sigma_b'^{-1} = \sigma_b^{-1} + (\sigma')_b^{-1}$$

Où $'\psi_b'$ représente la mesure de similarité spatiale et η est un facteur de normalisation gaussienne.

2.3.4 Evaluation des performances de l'approche proposée

2.3.4.1 Présentation des bases de données

Pour tester les performances du processus proposé pour le suivi multiple de personnes, nous présentons une première étude expérimentale, menée sur deux bases de données, avec des conditions différentes. La première est une base d'images intérieures alors que la seconde est une base d'images dans un environnement extérieure très varié :

— **CAVIAR [113] :** Context Aware Vision using Image-based Active Recognition. Il s'agit d'une large base de données d'images dans un centre commercial au Portugal intégrant plusieurs scénarios avec mono ou multi personnes marchant dans le couloir du centre. Nous avons utilisé le scénario de trois personnes, marchant ensemble, en vue frontal à la caméra et qui croisent un autre personnage se déplaçant dans la direction inverse. Les vidéos utilisées sont réalisées à 25 images/seconde avec 384×288 pixels par image.

— **CWV :** CrossWalk Valenciennes (*Annexes, Scénarios de test 1*). Cette deuxième base de données est celle réalisée dans le cadre de cette thèse. Nous l'avons construite progressivement devant la difficulté à trouver une base de données de traversée d'une rue avec les différentes conditions et scénarios possibles qui peuvent être observées en situation écologique. Nous envisageons d'héberger cette base sur le Net et de la rendre accessible aux publics intéressés par la sécurité routière et l'analyse comportementale de piétons. Ces séquences vidéo sont réalisées à 25 images/seconde avec 720×576 pixels par image.
Dans CWV, deux sens de scénarios de suivi multiple de personnes ont été utilisés : Le premier contient 2 piétons traversant le passage piétons dans la même direction et le deuxième contient 3 piétons avec 2 dans la même direction et le troisième dans la direction opposée. Notre base de données contient des situations d'occlusion volontairement difficiles, où des personnes ont été occultées par des véhicules ou par d'autres personnes. Les images sont issues de vidéo capturées dans un environnement extérieur avec des changements importants d'éclairage.

2.3.4.2 Protocole de tests

Par ces expérimentations, nous voulons déterminer les performances du processus de suivi, en termes d'association de la même identité à la bonne personne durant toute la séquence vidéo. Un numéro unique (1, 2, 3...) est associe initialement pour chaque personne, dès la première apparition dans la scène. Le résultat est dit correct, si le numéro associé à la personne reste le même tout au long de la séquence vidéo.

2.3.4.3 Résultats quantitatifs

Un processus idéal de suivi des objets multiples devrait, à tout moment, trouver le bon nombre d'objets présents dans la vidéo et estimer la position de chaque objet avec précision. Il doit également conserver une identité constante de chaque objet au cours du temps : chaque objet doit être affecté à une trajectoire unique qui reste constante tout au long de la séquence [114]. Ainsi, pour une évaluation quantitative, nous utiliserons les métriques fréquemment utilisées dans la littérature et proposées par [114], telles que :

— **La fiabilité de suivi des objets multiple (Multi-Object Tracking Accuracy : MOTA**) Par cette métrique nous voulons quantifier l'erreur moyenne. La valeur de MOTA se calcule par l'équation 2.3.

$$MOTA = 1 - \frac{\sum_{i,t}(f_{pt}, m_t, mme_t)}{\sum_t g_t} \tag{2.3}$$

Avec des réponses négatives-positives $'f'_{pt}$ (tout autre objet qui est considéré comme personne par erreur), des négatives-négatives $'m'_t$ (toute personne non détectée) et des erreurs d'association d'identité $'mme'_t$ (toute position associée à la mauvaise trajectoire) sur toutes les détections $'g'_t$ au frame $'t'$.

— **La fiabilité de détection des objets multiple (Multi-Object Detection Accuracy : MODA)** Par cette métrique nous voulons évaluer l'effet de la

technique d'interpolation sur les performances de détection. En effet, la valeur de MODA est calculée en fonction des réponses négatives-positives $'f'_{pt}$ (tout autre objet qui est considéré comme personne par erreur) et les négatives-négatives $'m'_t$ (toute personne non détectée) pour chaque frame $'t'$. La valeur de MODA se calcule par l'équation 2.4.

$$MODA = 1 - \frac{\sum_{i,t}(f_{pt}, m_t)}{\sum_t g_t} \tag{2.4}$$

Avec des réponses négatives-positives $'f'_{pt}$ (toute autre objet qui est considéré comme une personne par erreur) et les négatives-négatives $'m'_t$ (toute personne non détectée) pour chaque frame $'t'$.

Les valeurs de MODA et MOTA calculées sur les deux bases de tests sont résumées dans le tableau 2.1.

TABLE 2.1 – Résultats quantitatifs du processus proposé pour le suivi de personnes en termes de MODA et MOTA [111].

	MODA (%)	MOTA (%)
CAVIAR	73,74	55,33
CWV(scénario 1)	91,67	86,41
CWV(scénario 2)	92,06	85,56

Au regard du tableau 2.1, nous constatons en premier lieu des valeurs importantes du $MODA$. En effet, nous avons obtenu une précision de détection de $73,74\%$, $91,67$ % et $92,06$ % respectivement pour la base CAVIAR et les deux scénarios de notre base de données. Ces valeurs montrent que notre processus d'interpolation basé sur la vitesse moyenne, contourne le problème de détections perdues à cause de l'occultation totale ou partielle.

Les performances de l'association sont illustrées par les valeurs importantes de $MOTA$. En fait, dans une séquence de 114 images (scénario 1 de notre base de données), nous avons une valeur de $MOTA$ de $86,41$ %, avec un nombre d'associations erronées qui est de 6 sur 114. Cela traduit la robustesse de notre processus de couplage basé sur la maximisation de la similarité des spatiogrammes de détections successives.

2.3.4.4 Résultats qualitatifs

Pour la visualisation des performances du processus proposé pour la détection de personnes, des extraits des bases d'évaluation CAVIAR (Figure 2.8 a), CWV séquence 1 (Figure 2.8 b) et 2 (Figure 2.8 c) sont présentés. Ces extraits sont des images respectivement du début, milieu et fin des séquences pour visualiser l'évolution du suivi et l'association de la bonne identité à la bonne personne. Ces extraits montrent que les identités '1, 2 et 3' sont toujours associées à la même personne durant les différents instants de la séquence vidéo.



FIGURE 2.8 – Illustration de résultats de détection de personnes : a) Base CAVIAR, b) Base CWV séquence 1 et c) Base CWV séquence 2 [111].

Malgré les performances du spatiogramme en termes d'identification et de réidentification du bon objet cible durant plusieurs frames montrés, tant par la littérature que par les expérimentations présentées précédemment, ce descripteur peut générer des réponses négatives en nombre réduits (6 sur 114). En effet, à cause de la confusion d'identités, quelques positions de la première trajectoire (en bleu) peuvent être associées par erreur à la trajectoire 2 (en rouge) comme illustré par la figure 2.9 a).

2.3.5 Correction des trajectoires par filtrage

Pour résoudre ce problème nous proposons une étape de raffinement de trajectoires. Ce raffinement se base sur l'hypothèse qu'une personne à pieds ne peut pas subitement changer sa position plus que les conditions physique et morphologique d'un sujet ne lui permettent de le faire. Par conséquent, la distance entre deux positions successives de la même personne ne peut pas dépasser un seuil déterminé en fonction de sa vitesse moyenne de déplacement. Pour garantir que la bonne position est associée à la bonne trajectoire, nous proposons de calculer la distance entre cette position et les différentes trajectoires existantes. Ensuite, deux types de vérifications sont présentés comme formalisé par l'algorithme 2.

Entrée : Positions $P(0, \ldots k)$ sur l'image, trajectoires disponibles $t(0, \ldots n)$, seuil $'S'$
Sortie : Trajectoire relative à cette position
mind : **entier** ;
mind $\leftarrow$ 100 ;
Si ($t = 0$ et $p > 0$) **Alors**
  $T_p \leftarrow 1$;
Sinon
  Pour (i$\leftarrow$ 1 ; i <= k ; i++) **faire**
    Pour (j$\leftarrow$1 ; j <= n ; j++) **faire**
      Calculer la distance $'d'$ entre la position i et la trajectoire j
      Si ($d <= S$) **Alors**
        Si ($d < mind$) **Alors**
          mind $\leftarrow$ d
          $T_p \leftarrow j$
        Fin Si
      Sinon
        $T_p \leftarrow j$+1 ; *[cas d'un nouveau piéton dans la scène]*
      Fin Si
    Fin Pour
  Fin Pour
  Retourner T_p ;
Fin Si

Algorithme 2: Filtrage des trajectoires.

Pour une évaluation des performances de notre proposition de filtrage de trajectoires, nous avons utilisé une métrique fréquemment utilisée dans la littérature et proposées par [114], $MOTP$.

— **La précision de suivi des objets multiple (The Multiple Object Tracking Precision : MOTP)** Par cette métrique, nous quantifions la précision d'association de la bonne identité à la bonne personne. La valeur de MOTP se calcule par l'équation 2.5.

$$MOTP = \frac{\sum_{i,t} d_{it}}{\sum_t c_t} \tag{2.5}$$

Avec d_{it} la distance entre les positions réelles et celles estimées de la trajectoire $'i'$ à l'instant $'t'$ et c_t le nombre d'associations faites pour le frame $'t'$.

Les valeurs de MOTP calculées avant et après filtrage des trajectoires sont présentées par le tableau 2.2.

TABLE 2.2 – Évaluation des performances du filtrage des trajectoires.

	MOTP (cm)	
	Avant	Après
CWV(scénario 1)	29,89	10
CWV(scénario 2)	31,4	14,34

Au regard du tableau 2.2, nous constatons une amélioration importante des valeurs de $MOTP$ qui sont divisées par 3 entre avant et après le raffinement des trajectoires du scénario 1 et par 2 pour le scénario 2. Cette diminution de $19,89\,cm$ de la distance moyenne entres les trajectoires réelles et celles estimées montre l'importance de cette étape de raffinement des trajectoires. En effet, cette étape améliore considérablement l'estimation de la trajectoire et corrige certaines fausses détections comme présenté à la figure 2.9 b). Notre solution est basée sur le calcul de la distance euclidienne entre la position du piéton au niveau $'i'$ et sa position au niveau $'i+1'$. Ainsi, la détection sera associée à la trajectoire la plus proche.

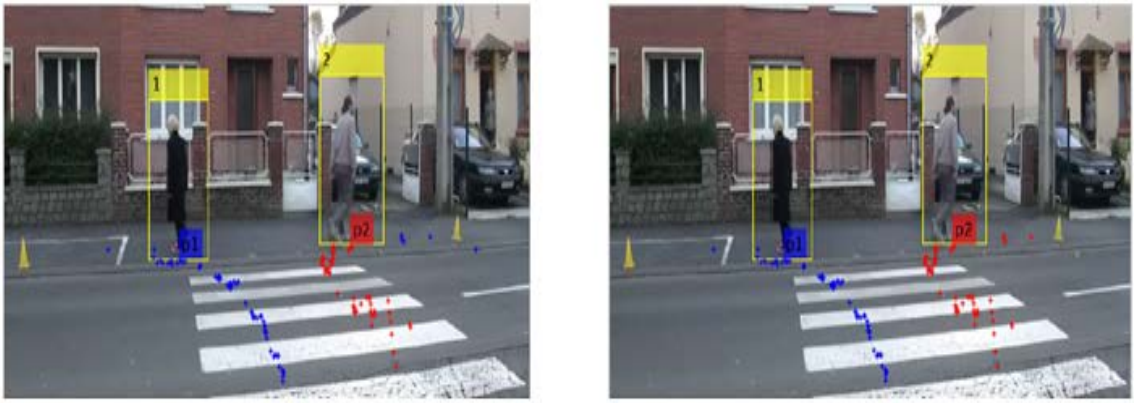


FIGURE 2.9 – Résultats du filtrage des trajectoires : a) Avant filtrage, et b) Après filtrage [111].

2.4 Modélisation de la zone de traversée (passage piéton) et détection de la position réelle des piétons

L'analyse comportementale des piétons lors de la traversée de la rue à partir de traitements de vidéos nécessite la collecte de variables réelles telles que la vitesse des

piétons en $m.s^{-1}$. Par conséquent, on cherche à calculer la position des piétons en cordonnées réelles et non celle en pixels. Ces coordonnées sont alors directement interprétables et ont une signification directe dans ce contexte. Ainsi, la calibration des caméras est une étape obligatoire pour assurer le passage de 'coordonnées images' aux 'coordonnées réelles' (cartésiennes). Pour calibrer les caméras et déterminer les cordonnées des positions des piétons dans le repère cartésien nous avons utilisé une méthode à base d'homographie [115]. Cette méthode se base sur le fait d'établir une correspondance entre les points de l'image et un modèle de la zone d'intérêt construit avec les coordonnées réelles de la scène. Dans le contexte de la traversée d'une rue, nous avons choisi le passage piéton avec les deux trottoirs comme modèle de mappage, puisqu'il est la zone de traversée ordinaire et dispose de suffisamment de points caractéristiques pour cela.

2.4.1 Modèle de passage piéton

Le modèle de passage piéton est construit par la réalisation de mesures réelles dans la zone d'intérêt, là où les séquences ont été filmées (CWV). Le passage piéton dans cette zone se compose de 5 bandes et de deux trottoirs tels que présenté à la figure 2.10.

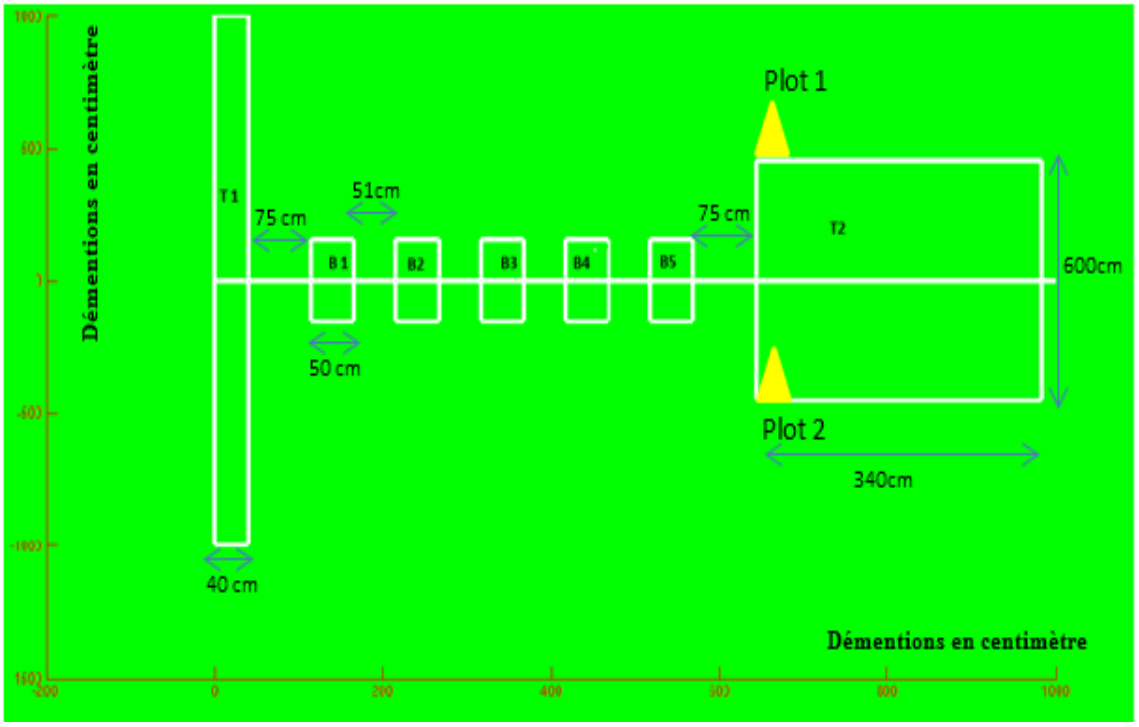


FIGURE 2.10 – Le modèle de calibration au niveau du passage piéton.

2.4.2 Homographie

L'homographie plane transforme une relation linéaire non singulière entre deux plans projectifs. Dans notre cas, la transformée d'homographie définit une application de correspondances linéaires de points entre le modèle plan de passage piéton (position 2D, réalité de terrain) et le plan d'image (position de pixel) suivant l'équation 2.6.

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix} \times \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} \tag{2.6}$$

Pour calculer la matrice d'homographie $'H'$ il est nécessaire de récupérer des coordonnées (X_i, Y_i) de quelques positions sur le modèle et leurs correspondants (x_i, y_i) tels que présenté à la figure 2.11. L'étape suivante est alors de résoudre le système de l'équation 2.7.

$$\begin{pmatrix} 0 & 0 & 0 & -X_1 & -Y_1 & -1 & y_1X_1 & y_1X_1 & y_1 \\ X_1 & Y_1 & 1 & 0 & 0 & 0 & -x_1X_1 & -x_1Y_1 & -x_1 \\ 0 & 0 & 0 & -X_2 & -Y_2 & -1 & y_2X_2 & y_2X_2 & y_2 \\ X_2 & Y_2 & 1 & 0 & 0 & 0 & -x_2X_2 & -x_2Y_2 & -x_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & -X_n & -Y_n & -1 & y_nX_n & y_nX_n & y_n \\ X_n & Y_n & 1 & 0 & 0 & 0 & -x_nX_n & -x_nY_n & -x_n \end{pmatrix} \begin{pmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \\ h_{33} \end{pmatrix} = \theta_9 \tag{2.7}$$

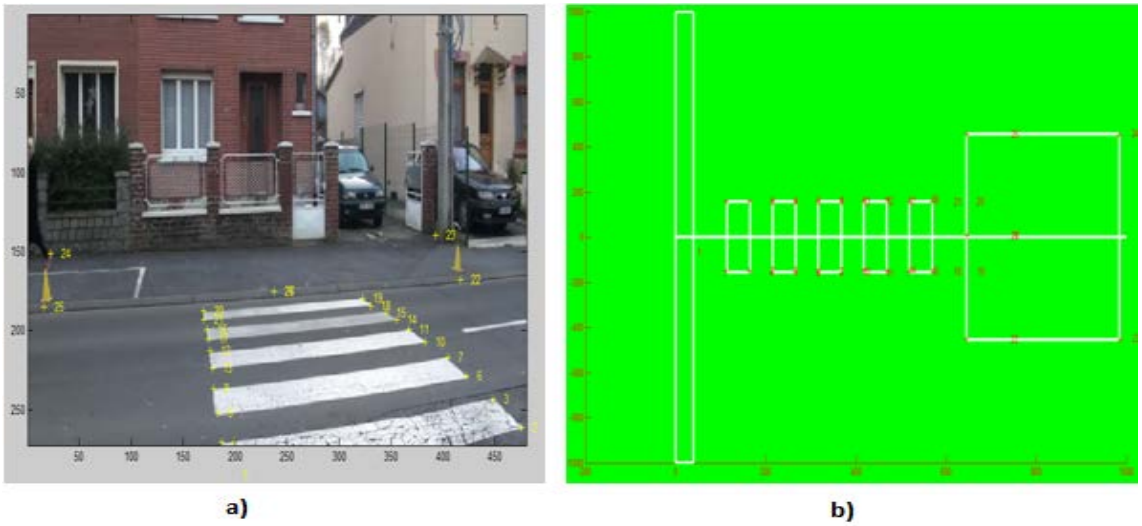


FIGURE 2.11 – Points de correspondances entre le plan image et le modèle de passage piéton.

Une évaluation de la robustesse de cette technique a été effectuée à partir du calcul de positions et de trajectoires de différents niveaux de complexités. Trois types de

trajectoires sont ainsi présentées à la figure 2.12. Ces trajectoires sont soit simples et linières (uniquement la phase de traversée) soit composées (phases de rapprochement et de traversée). Elles incluent aussi une multitude de variations de directions.

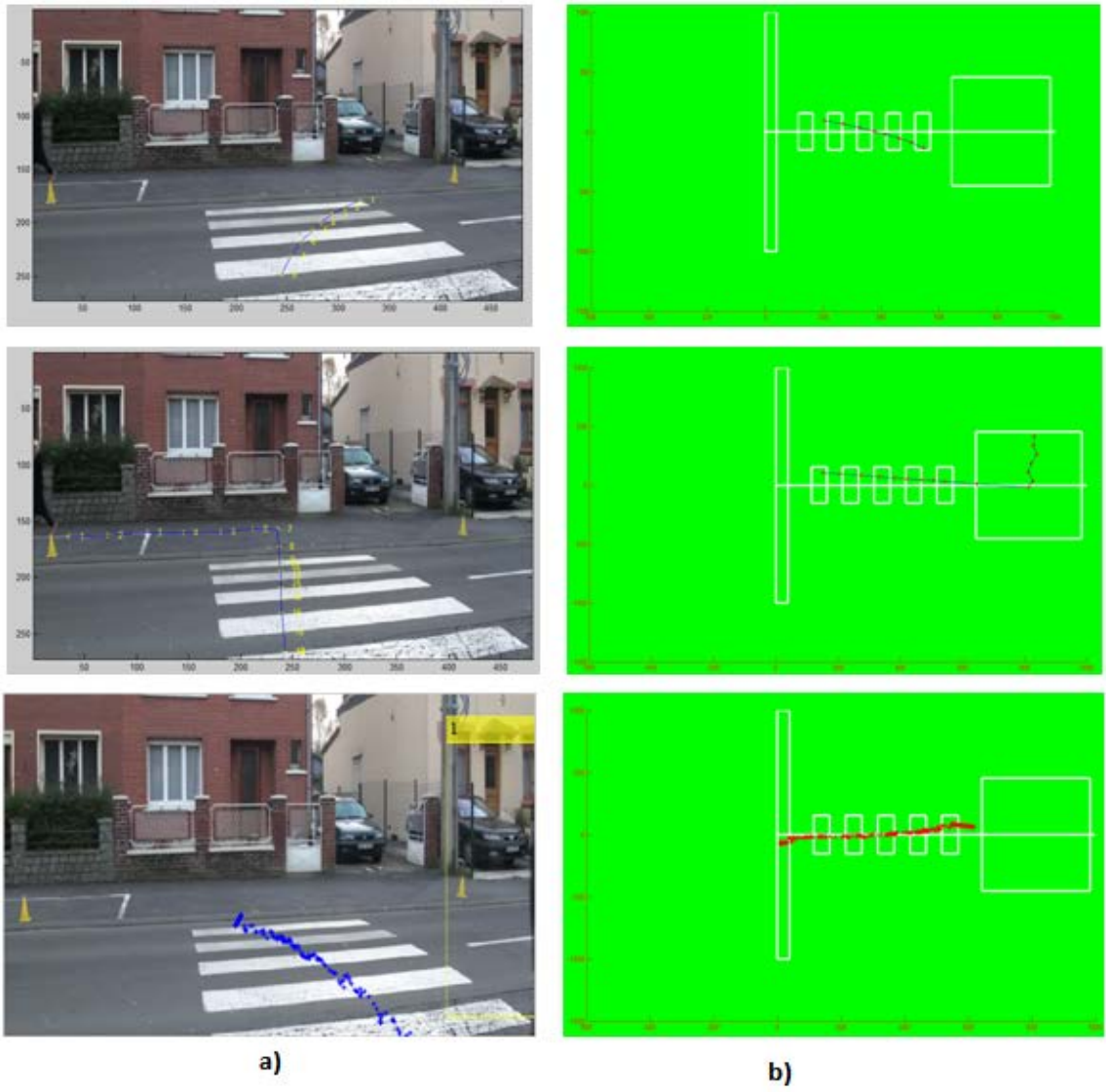


FIGURE 2.12 – Exemples de transformations de trajectoires des coordonnées images en coordonnées réelles : a) Trajectoires sur plan image, b) Trajectoires projetées sur le modèle de passage piéton.

2.5 Données collectées pour la caractérisation du comportement du piéton

2.5.1 Statistiques sur la cinématique du piéton

La cinématique du piéton est un facteur clé pour toute analyse de traversée d'une rue. En effet, l'analyse de la vitesse des piétons a été l'objet de plusieurs études et suivant plusieurs modalités. Les résultats présentés dans le tableau 2.3 sont issus de travaux de recherche récents, réalisés sur des piétons en situations réelles [116, 117].

TABLE 2.3 – Données caractéristiques pour une traversée de passage piéton avec un feu tricolore (moyennes ± écarts types) [117].

Attention du piéton	Temps de réaction (s)	Accélération (g)	Vitesse ($m.s^{-1}$)
Normal	0,8 ± 0,15	0,14 ± 0,07	1,33 ± 0,24
Anticipe	0,77 ± 0,75	0,16 ± 0,13	1,4 ± 0,25
Distrait	1,87 ± 1	0,14 ± 0,07	1,39 ± 0,25
Moyenne		0,14 ± 0,09	1,36 ± 0,24

Ces valeurs numériques donnent de bonnes indications sur les ordres de grandeur de vitesses et d'accélérations, mais la vitesse d'un piéton reste une grandeur variable puisqu'elle dépend selon [108] de plusieurs paramétres à savoir :

- La raison du déplacement, comme le montre le tableau 2.4.
- Le sexe du piéton, comme le montre le tableau 2.5., les hommes marchent plus vite que les femmes.
- L'âge du piéton.
- La température.
- L'heure de la journée.
- La surface sur laquelle il se déplace, des infrastructures (escaliers, tapis roulants...).

TABLE 2.4 – Vitesse des piétons en fonction de leur activité [108].

Activité	Vitesse moyenne ($m.s^{-1}$)
Changement de mode de déplacement	1,49
Faisant des courses	1,16
Loisir et détente	1,10

TABLE 2.5 – Vitesses moyennes respectives des hommes et des femmes, population allemande [108].

Population	Vitesse moyenne ($m.s^{-1}$)	écart type
Ensemble	1,34	0,26
Hommes	1,41	-
Femmes	1,27	-

2.5.2 Déplacements des piétons

En moyenne, la trajectoire d'un piéton qui traverse la route sur un passage piéton sera plutôt perpendiculaire à la route, alors qu'un piéton qui traverse hors d'un passage piéton aura tendance à arrondir les angles et à prendre la diagonale, comme le montre la figure 2.13.

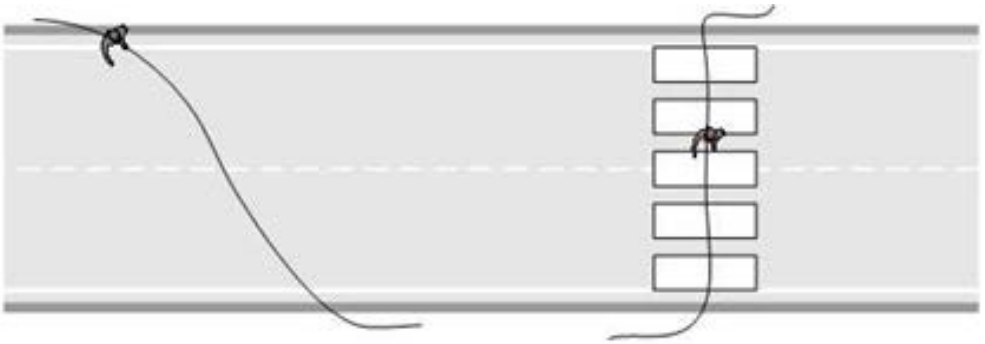


FIGURE 2.13 – Trajectoires d'un piéton traversant une route, hors (à gauche) et sur (à droite) un passage piéton [13].

Les différentes phases d'une traversée de route sont typiquement [13] :

1. "Démarrage" en descendant du trottoir ou au début d'une nouvelle voie, avec augmentation de la vitesse, jusqu'à atteinte de la vitesse voulue.
2. Maintien de la vitesse de croisière.
3. Éventuellement, avant de traverser une nouvelle voie empruntée par des véhicules : accélération, ralentissement, ou changement de direction doux.
4. "Fin de traversée" : léger ralentissement, changement de direction.

Un piéton se déplace en adoptant une des trois allures suivantes : marche, course lente (jogging en anglais) ou course rapide (running en anglais) [13].

Pour piloter son déplacement, le piéton a la possibilité d'agir sur les éléments suivants :

— Lorsque le piéton marche "normalement", vers l'avant (déplacement longitudinal), il joue essentiellement sur : la longueur des pas, la cadence et/ou la direction suivie.
— D'autres déplacements sont également possibles : les pas sur le côté, l'arrêt ou la marche arrière mais ils restent rares.

A partir de ces observations et de l'analyse accidentologique quatre paramètres seront collectés pour caractériser le comportement des piétons lors de la traversée de la rue :

— Vitesse instantanée.
— Vitesse moyenne.
— Phases de traversée.
— Temps d'attente.

2.5.3 Estimation des vitesses instantanées et moyennes des piétons

Peu d'outils automatiques sans instrimentation du sujet sont déjà implémentés pour l'estimation de la vitesse des piétons, en particulier en utilisant les techniques de vision par ordinateur, comme le rappel [15].

En exploitant le processus de détection et de suivi de personnes présenté précédemment, la vitesse instantanée est calculée comme le rapport de la distance parcourue par le temps. La distance parcourue est la différence entre la position piétonne dans le frame $'i'$ et le frame $'i + x'$ (x étant égale à 25, la fréquence d'acquisition de la caméra, si l'on veut le calculer toutes les secondes par exemple).

La vitesse moyenne est un paramètre très important pour la caractérisation du comportement général du piéton [118, 119]. Ainsi, pour être plus fin dans l'analyse, nous déterminerons la vitesse moyenne du piéton durant les phases de mouvement telles que la phase de rapprochement et la phase de traversée. Nous proposons pour cela deux méthodes pour le calcul de la distance parcourue.
Pour les deux méthodes, le temps du mouvement se calcul comme étant le rapport entre le nombre de frames (n) par séquence et 25 : $T = n/25$.

— Méthode 1 (M1) : La distance est estimée comme la différence entre le premier et le dernier point de la trajectoire, désignés respectivement par leur position sur le modèle de passage piéton, $(x_1; y_1)$ et $(x_n; y_n)$ comme indiqué dans l'équation 2.8. Par conséquent, avec cette méthode on détermine la vitesse moyenne indépendamment des mouvements microscopiques du piéton.

$$D = \sqrt{(x_1 - x_n)^2 + (y_1 - y_n)^2} \tag{2.8}$$

— Méthode 2 (M2) : La distance est égale à la somme des distances entre chaque binôme de points successifs des $'n'$ positions de la trajectoire (Équation 2.9).

$$D = \sum_{i=1}^{n-1} \sqrt{(x_i - x_{i+1})^2 + (y_i - y_{i+1})^2} \tag{2.9}$$

La vitesse moyenne sera alors calculée comme étant : $V = \frac{D}{T}$.

2.5.4 Estimation des phases de traversée

Le comportement du piéton durant la traversée de la rue est influencé par le niveau de vigilance juste avant la phase de traversée [120]. L'exécution d'une traversée sécuritaire est par conséquent fortement reliée au comportement des piétons dans la phase qui précède la descente du trottoir, classiquement une phase d'attente. La traversée sécuritaire est également liée à d'autres paramètres tels que la vitesse de rapprochement, l'existence de gap(s) sécuritaire(s), etc.

Bien que ces paramètres soient importants et influencent le comportement du piéton, ils ne sont pas actuellement abordés dans les études antérieures [60]. Cet état de fait a motivé à approfondir cette approche pour compléter avantageusement les données

de la littérature qui n'analysent les phases de traversée qu'académiquement.

Par conséquent, la détermination d'une méthode automatique pour traiter la détection des phases de traversée est très utile pour les études visant à améliorer la sécurité des piétons.
Dans notre travail, deux méthodes sont proposées afin de détecter les phases de traversée de la rue à partir de :

— L'analyse des positions spatiales des piétons.
— L'analyse de la vitesse instantanée.

2.5.5 Estimation du temps d'attente

En temps normal, lors de ses déplacements, que ce soit dans un bâtiment ou à l'extérieur, le piéton fait continuellement appel à ses sens, et en particulier à la vue, pour s'orienter et adapter sa trajectoire [121]. Un piéton surveille son environnement de façon permanente et "naturelle" ou automatique, afin d'adapter son comportement et réagir convenablement face aux circonstances. En effet, dans un contexte de traversée d'une rue, pour bien surveiller l'état du trafic et repérer les gaps sécuritaires, le piéton effectue une petite phase d'attente au bord du passage piéton avant la traversée.

Dans ce travail, le temps d'attente $'T'_w$ avant de traverser la rue est également automatiquement estimé en divisant le nombre de frames là où le piéton est au bord du passage ($n_{waiting}$) par 25 (Équation 2.10).

$$T_w = \frac{n_{waiting}}{25} \tag{2.10}$$

2.5.6 Evaluation de la précision des données calculées

2.5.6.1 Etude de cas et base d'évaluation

Afin de mieux expliquer l'estimation des données piétonnières et illustrer leurs fiabilités, une première étude quantitative est présentée. Elle a été effectuée sur notre base de données de traversée de la rue CWV [111] (*cf partie 2.3.4.1*). Les séquences vidéo ont été filmées dans un lieu public à Aulnoy lez Valenciennes - France. Nous avons utilisé 5 scénarios de piétons traversant une rue au niveau du passage piéton. Ces scénarios intègrent une variété de vitesses (allant de $0,6$ à $1,2\,m.s^{-1}$), de directions (vue de face ou vue de dos) et de temps d'attente (0 à $3\,s$). La base de données utilisée dans ce travail contient 32 séquences. Elle contient des situations d'occlusion volontairement difficiles, où les piétons sont occultés par des voitures et/ou d'autres piétons. Les séquences vidéos ont été filmées dans des espaces extérieurs intégrant des changements importants d'éclairage causés par les conditions météorologiques variables (Nord de la France : région à luminosité moyenne à faible).

Durant ces séquences les piétons devaient marcher le long du trottoir avec une vitesse normale, lente ou rapide, effectuer une phase d'attente aléatoire et traverser la rue, avec une des trois vitesses présentées ci-dessus. Dans chaque phase, les vitesses des piétons ont été estimées manuellement et ensuite automatiquement en utilisant le système proposé. Chaque limite de phase est détectée en fonction de la variation de

vitesse, puis validée à partir des données de terrain collectées manuellement par des techniciens.

2.5.6.2 Résultats expérimentaux

2.5.6.2.1 Evaluation de la précision d'estimation des phases de traversée

Les dimensions des passages piétons sont connues précisément puisque mesurées manuellement sur la zone de travail, là où la base de données est acquise (5 m de longueur et 3 m de largeur). Chaque position dans cet intervalle sera considérée comme faisant partie de la phase de traversée. Cependant, le piéton peut attendre en retrait d'un mètre avant le passage pour surveiller l'état de la route : phase d'attente. La phase de rapprochement est au-delà de cette zone (zone d'attente) d'un mètre à une distance non fixe. La figure 2.14 illustre la trajectoire d'un piéton après avoir été divisée en trois phases en fonction des positions spatiales :

— Phase de rapprochement.
— Phase d'attente.
— Phase de traversée.

Elle montre que la projection des positions spatiales brutes du piéton sur le modèle de passage piéton permet de déterminer les phases de traversée de la rue si le piéton suit les aménagements urbains présents tel que le passage piéton.

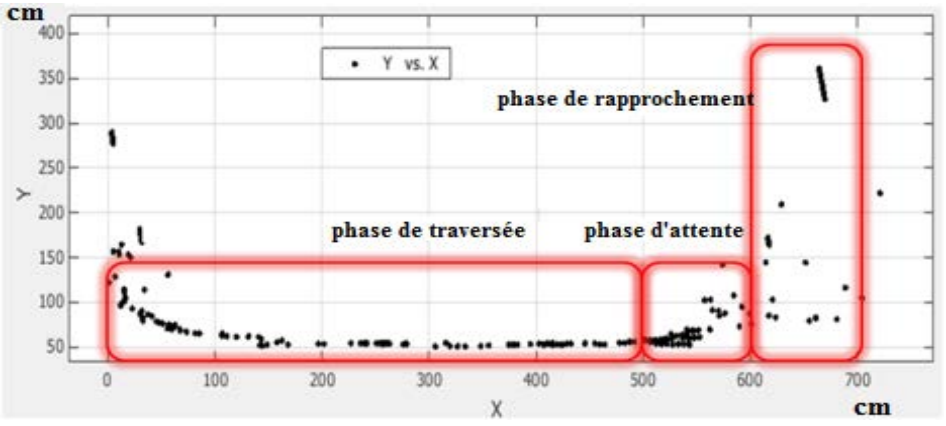


FIGURE 2.14 – Estimation des phases de traversée par l'analyse des positions spatiales du piéton.

Habituellement, tous les piétons se déplacent selon des trajectoires plus ou moins aléatoires et ne respectent pas l'ensemble des aménagements urbains spécifiques déjà implémentés tel que le trottoir. Ainsi, les phases d'approche et d'attente peuvent ne pas être sur le trottoir. Il ne suffit donc pas d'utiliser la méthode d'analyse des positions spatiales des piétons pour détecter les phases de traversée de tous les piétons.

Contrairement à la première méthode, on propose donc ici une méthode 'intelligente' non sensible à la position des piétons et à la non-linéarité des trajectoires. Elle est basée principalement sur notre système automatique de suivi des piétons, en utilisant l'analyse des trajectoires et des vitesses des piétons qui sont calculées chaque seconde

et utilisées pour mettre en évidence les limites des phases de traversée. Pour cela, nous utilisons la variation de la vitesse instantanée, en supposant que les phases de rapprochement et de croisement sont habituellement séparées par une vitesse réduite, pouvant atteindre $'0'$, pendant la phase d'attente. Pour présenter les performances de cette méthode et illustrer le potentiel de variation de vitesse pour détecter les limites de phases de traversée, 3 scénarios différents, pour 3 piétons, de notre base de données 'CWV' ont été utilisés comme illustré à la figure 2.15. Les scénarios utilisés sont les suivants :

— Scénario 1 : Les piétons ont une vitesse normale pendant les phases de rapprochement et de traversée sur le passage piéton (Figure 2.15 a).
— Scénario 2 : Les piétons ont une phase de rapprochement avec une vitesse rapide et une phase de traversée avec une vitesse lente sur le passage piéton (Figure 2.15 b).
— Scénario 3 : Les piétons ont une vitesse de marche lente le long du trottoir et une phase de traversée rapide sur le passage piéton (Figure 2.15 c)).

La méthode proposée permet de visualiser facilement la diminution de la vitesse lors de la phase d'attente, même si elle est de l'ordre de 2 secondes. Aussi, la variation de vitesse montre le potentiel de notre outil automatique pour détecter les frontières de chaque phase malgré les variations de vitesses.

2.5.6.2.2 Évaluation de la précision de l'estimation de la vitesse moyenne

Il est important pour un système automatique d'avoir des données de référence pour évaluer les résultats mesurés. Les données de référence, utilisées à ce stade pour évaluer la précision d'estimation de la vitesse moyenne, ont été obtenues à partir d'observations recueillies par des techniciens. La vitesse moyenne estimée a été comparée à ces données pour illustrer la précision de l'outil. La comparaison a été obtenue en termes d'erreur moyenne ($RMSE$) et de coefficient de corrélation (R) qui ont été calculés avec les équations 2.11 et 2.12.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(X_i - X_n)^2}{n}} \tag{2.11}$$

$$R = \frac{\sum_{i=1}^{n}(X_i - \bar{X}) \times (Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X}) \times \sum_{i=1}^{n}(Y_i - \bar{Y})}} \tag{2.12}$$

Où X et Y sont respectivement les vecteurs de réalité de terrain et ceux mesurés et $\bar{X}$ et $\bar{Y}$ sont respectivement les moyennes de $'n'$ valeurs de réalité de terrain et mesurés.

Le tableau 2.6 résume les résultats de calculs de la $RMSE$ et de R entre les valeurs de la vitesse moyenne calculée et celle réelle en utilisant les deux méthodes $'M1'$ et $'M2'$ (*cf partie 2.5.3*). La figure 2.16 illustre la corrélation entre ces valeurs de vitesses.

D'après le tableau 2.6, nous constatons la valeur réduite de l'erreur moyenne, de l'ordre de $0,07\ m.s^{-1}$, soit $4,66\%$ de la vitesse moyenne de marche, donnée par M1, entre la trajectoire générée automatiquement basée sur notre 'tracker' et la valeur de réalité de terrain. Cette indice montre que notre outil automatique présente de bonnes performances en matière de détection et de suivi des piétons puisqu'il génère une trajectoire assez bien corrélée à celle de la réalité du terrain. Ces performances sont illustrées

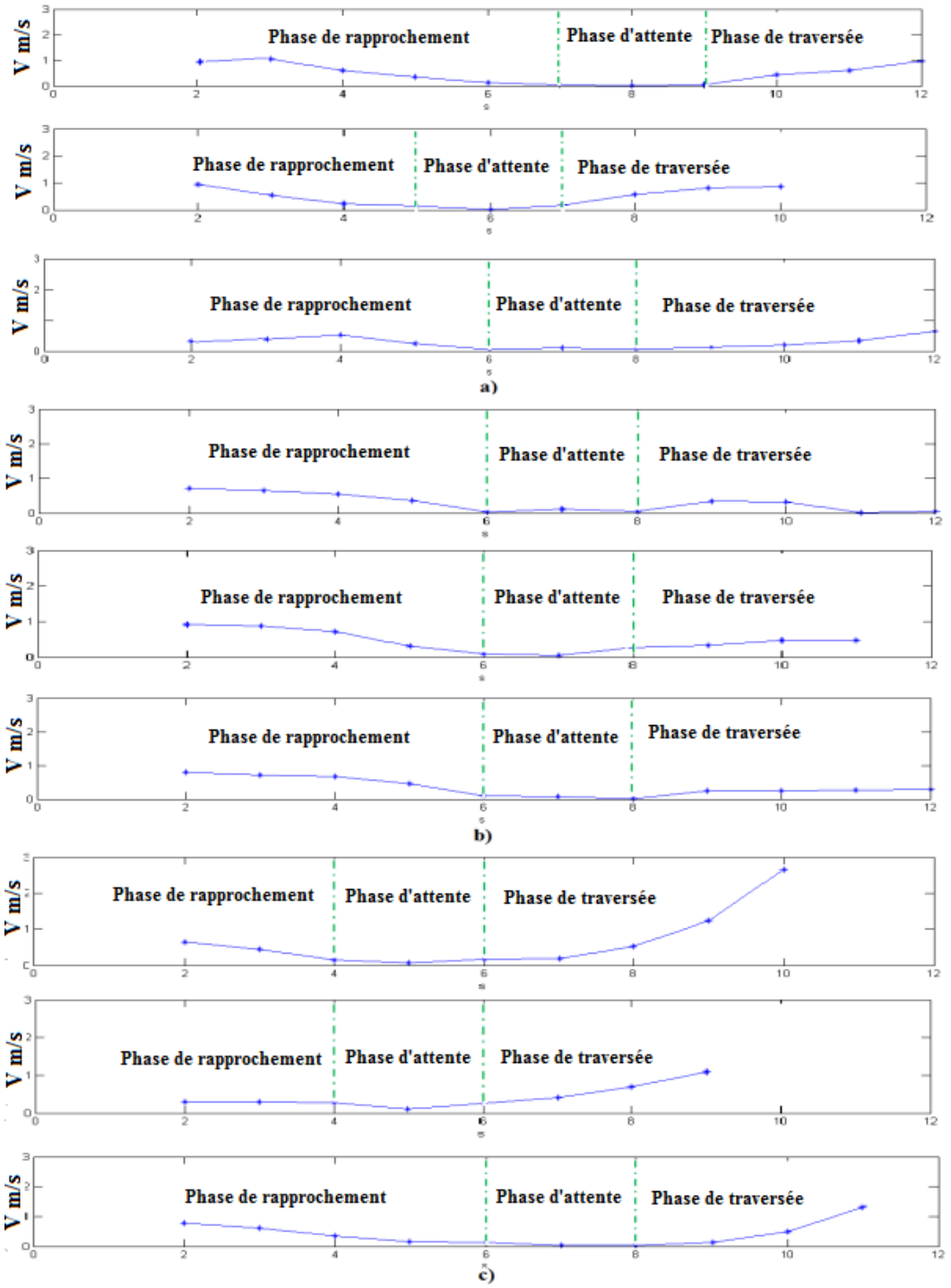


FIGURE 2.15 – Analyse de la variation de la vitesse instantanée pour la détection des phases de traversée de la rue.

également par la valeur, raisonnablement réduite, de l'erreur moyenne obtenue à partir de (M2) ($0,21\ m.s^{-1}$) étant donné que M2 est basée sur chaque position de la trajectoire du piéton. De plus, même lorsque la vitesse est estimée sur des intervalles de temps très limités en capturant les petits mouvements du piéton (chaque $\frac{1}{25}s$ le temps

TABLE 2.6 – Résultats de validation de la vitesse moyenne en utilisant les deux méthodes proposées pour le calcul des distances.

	Phase de rapprochement		Phase de traversée	
	M1	M2	M1	M2
RMSE	0,4	0,95	0,07	0,21
R^2	0,89	0,77	0,98	0,79

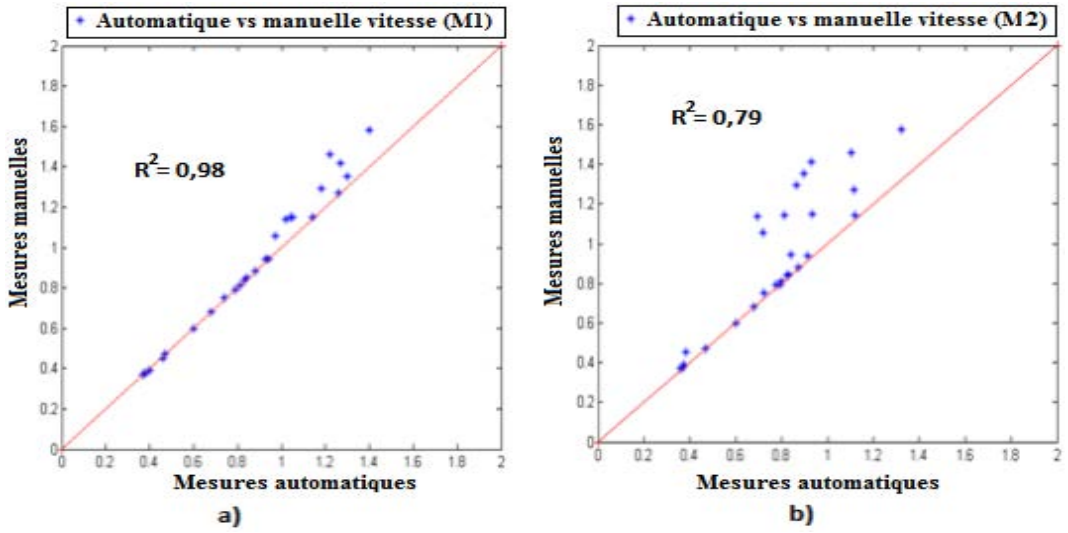


FIGURE 2.16 – Illustration des corrélations entre les vitesses moyennes estimées et les vitesses de réalités de terrain, a) Méthode M1, b) Méthode M2.

séparant deux positions successives), les valeurs de $RMSE$ et de R^2 avec M2 montrent une bonne corrélation entre les vitesses de réalité de terrain et estimées. Ces résultats montrent que le 'tracker' proposé rapporte correctement les données de référence en tenant compte d'un haut degré de détail (chaque $\frac{1}{25}s$).

En outre, l'erreur moyenne faible de l'ordre de $0,4 m.s^{-1}$ lors de la phase de rapprochement, malgré la variation significative du point de vue, étant donné que les piétons se déplacent latéralement par rapport à la caméra pendant la phase de rapprochement, montre que notre outil automatique est peu sensible à la variation de point de vue. Aussi, un coefficient de corrélation de 0, 89 montre la capacité de notre outil à suivre correctement le mouvement des piétons et la variation de vitesse quelle que soit la trajectoire empruntée.

De plus, à partir de la figure 2.16 (a et b), on peut remarquer que les vitesses de références et estimées sont fortement corrélées lorsque la vitesse est inférieure à $1\ m.s^{-1}$ (la vitesse humaine moyenne [15]) ; Alors qu'au-delà de cette valeur, la corrélation se dégrade. Cette déviation est plus importante avec M2 lorsque la vitesse du piéton est estimée en s'appuyant sur tous les points de la trajectoire. Ce phénomène peut être expliqué en se basant sur l'hypothèse que la distance parcourue entre deux images successives de la séquence vidéo est plus petite et se calcule comme suivant (Équation

2.13) :

$$d = \frac{V_p}{25} \tag{2.13}$$

Où V_p est la vitesse moyenne du piéton et 25 la fréquence d'acquisition de la caméra. Par conséquent, lorsque la vitesse des piétons est moins importante, la distance parcourue est plus petite. Plus les points sont proches plus l'imprécision influencera le calcul de la distance, à l'inverse, plus les points de trajectoire sont parfaitement localisés plus le comportement du piéton est correctement suivi. En conséquence, la vitesse moyenne sera estimée plus précisément. En effet, les erreurs résiduelles peuvent être attribuées à l'imprécision dans le calibrage de la caméra et les irrégularités dans l'estimation de toutes les positions de piéton dues au bruit dans la détection et le suivi.

2.5.6.2.3 Évaluation de la précision de l'estimation du temps d'attente

Le temps d'attente estimé, doit être également validé par rapport aux données de réalité de terrain, prises comme données de référence. Les valeurs de $RMSE$ et de R^2 sont également calculées pour estimer la précision de la collecte de cette variable du piéton à base de traitement bas niveau des séquences vidéo. Le tableau 2.7 résume ces valeurs.

TABLE 2.7 – Précision d'estimation du temps d'attente en fonction du R^2 et de la $RMSE$.

	Temps d'attente
RMSE	1,00
R^2	0,68

Étant donné que l'estimation du temps d'attente est difficile, le piéton n'étant généralement pas totalement arrêté et immobile mais ayant tendance à bouger les bras, la tête ou les pieds lors de la phase d'attente, se baser uniquement sur le suivi du mouvement ne peut pas donner une bonne solution. La méthode de projection proposée pour passer des coordonnées d'image à celles réelles en utilisant le processus d'homographie avec la modélisation du sol présente une solution qui remédie à ce problème (tableau 2.6). Cette méthode permet de déterminer la position des piétons indépendamment de petits mouvements des bras. Par conséquent, notre outil automatique peut parfaitement, estimer la durée de la phase d'attente et localiser le piéton sur le trottoir avant de traverser la rue. Cette affirmation est illustrée par la valeur de $RMSE$ faible entre les temps d'attente de réalité de terrain et calculés, de l'ordre de $1s$ pour 32 séquences vidéos, avec ou sans période d'attente.

2.5.6.3 Synthèse et étude comparative

Une étude comparative entre notre approche et celle de [60], en termes de coefficient de corrélation (R^2) et de RMSE utilisant la même méthode M1, est présentée dans le tableau 2.8.

TABLE 2.8 – Comparaison de la précision de l'estimation de la vitesse moyenne des piétons en utilisant les techniques de vision par ordinateur (suivi de piétons et analyse de trajectoires).

	L'outil proposé(M1)	[60]
RMSE	0,070	0,041
R^2	0,98	0,92

Bien que notre approche analyse le comportement des piétons en suivant un long chemin avec une variation de direction et que le travail précédent [60] repose sur des conditions expérimentales restreintes (petite distance, trajectoire droite et avec un seul point de vue), une étude comparative montre que l'estimation de la vitesse produite par notre outil est plus précise. En fait, elle permet d'obtenir une valeur de coefficient de corrélation de Pearson plus importante de 0,06.

2.6 Conclusion

La plupart des études sur la sécurité routière et l'analyse du comportement précis des piétons lors de la traversée d'une rue dans la littérature, ne font pas appel à une analyse automatisée de la vidéo pour la collecte de données sur les piétons.

Dans cette étude, la mise en place d'un outil automatique de collecte de données du piétons à l'aide de techniques de vision par ordinateur ainsi que l'identification et la collecte des paramètres qui décrivent le mieux le comportement des piétons lors de la traversée d'une rue ont été nos principaux objectifs.

Un système a été développé et testé sur des données vidéo réelles recueillies près de l'Université de Valenciennes et du Hainaut Cambrésis : UVHC – Valenciennes – France, dans une rue péri-urbaine ayant un trafic relativement dense et un climat très changeant, alternant soleil et nuages.

L'outil automatique proposé, basé sur la détection et le suivi des piétons par des techniques de vision par ordinateur, montre de bonnes performances comparé à la littérature. Les valeurs des erreurs moyennes et des coefficients de corrélations calculés entre les vitesses moyennes déterminées manuellement et celles estimées le montrent également. En utilisant les variations de vitesse, nous avons ensuite proposé une stratégie pour détecter les phases de traversée d'une rue qui donne aussi de bons résultats.

Gagner en précision et en finesse dans les paramètres recueillies permettra notamment de mieux apprécier les effets du vieillissement sur le comportement du piétons qui se dégrade de manière lente mais certaine avec l'avancée en âge. Ainsi, les paramètres (vitesse, temps d'attente..) sensibles au vieillissement pourront être étudiés de façon plus efficace et à grande échelle. En effet, la prise en compte du paramètre 'âge' pour l'analyse du comportement du piéton lors de la traversée de la rue est très importante, dans une perspective de santé publique, au regard du nombre majoritaire d'accident de la circulation qui concernent les personnes âgées. Cette hypothèse est l'objectif du chapitre suivant.

Chapitre

3 Estimation automatique de la classe d'âge du piéton à partir de la démarche

3.1 Introduction

Le vieillissement affecte progressivement le comportement du piéton lors de la traversée de rues, sa capacité à surveiller l'état de la circulation ainsi qu'à, manipuler les gaps sécuritaires [120].

En effet, les piétons âgés, que nous définirons comme un sujet ayant plus de 65 ans, représentent 52% des tués en 2015 en France [17] alors qu'ils ne représentent que 18,8% de la population. De plus, la part des plus de 65 ans dans la mortalité piétonne est passée de $37,2\%$ en 2011 à 52% en 2015, soit $14,8\%$ de décès de plus chez les seniors de 65 ans et plus. S'intéresser à l'âge comme paramètre piétonnier pour la reconnaissance de scénario à risque de traversée de la rue s'avère ainsi un élément clef dans le cadre de la sécurité routière.

Dans ce chapitre, nous allons tout d'abord présenter le processus général de reconnaissance automatique de l'âge, ensuite nous introduirons les principales approches qui existent dans la littérature. Nous présenterons dans un troisième temps le descripteur proposé pour caractériser l'age. Nous clôturerons par une étude expérimentale évaluant les performances de ce descripteur en le comparant aux descripteurs existants.

3.2 Processus proposé : description générale

La phase d'estimation de l'âge des piétons consiste à prédire automatiquement s'ils sont âgés ou non. Dans ce travail nous avons opté pour l'utilisation de la démarche

comme un indicateur de vieillissement, vu notre contexte de travail qui est une traversée de rue où le piéton se situe à une distance raisonnablement grande par rapport aux caméras. En fait, la reconnaissance de la classe d'âge consiste à quantifier les effets du vieillissement qui touchent la démarche et/ou la posture du piéton.

En effet, à partir d'une séquence vidéo d'une personne en marche, nous appliquons en premier lieu des pré-traitements qui servent à extraire et nettoyer la silhouette de la personne. Ensuite, étant donnée la périodicité de la marche humaine, une étape de détection du cycle de marche est réalisée afin de synchroniser les traitements pour les différents descripteurs. L'étape suivante consiste à extraire les caractéristiques décrivant le mieux les effets du vieillissement et générer un modèle de données. La phase suivante est une phase d'apprentissage qui permet à l'algorithme de classification d'apprendre le modèle des données générées. Une classification est enfin réalisée pour juger l'appartenance d'une personne à une des deux classes, à savoir "jeune" ou âgée. La figure 3.1 illustre le processus de reconnaissance de la classe d'âge.

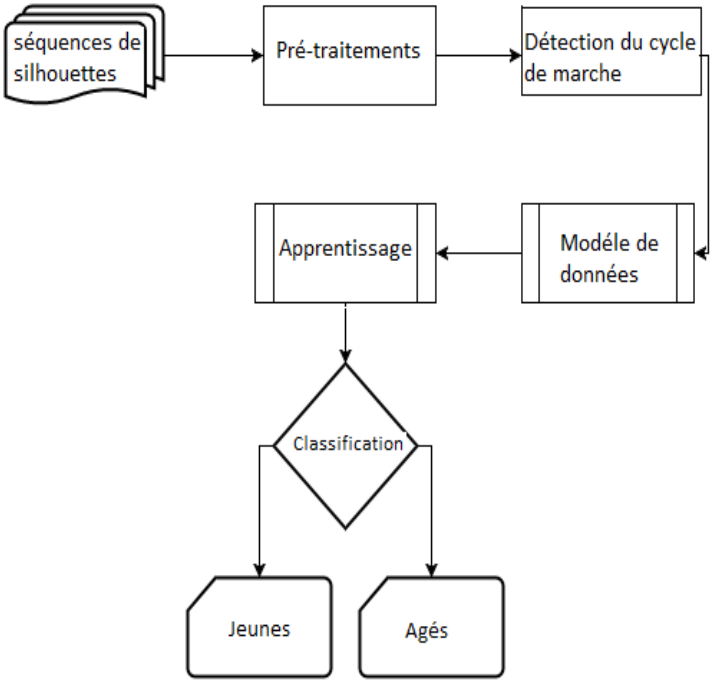


FIGURE 3.1 – Processus proposé pour la reconnaissance de la classe d'âge.

3.3 Les approches existantes pour la reconnaissance de l'âge

L'effet de l'âge sur l'être humain se distingue facilement sur différents aspects, à savoir les changements qui touchent la texture et la forme du visage, la taille et la forme du corps, la vitesse de déplacement de l'individu etc... En effet, l'estimation de l'âge précis, qui est un problème de régression [122, 123, 124] et la reconnaissance de la classe d'âge, (jeune ou âgé) qui est un problème de classification [125, 126], peuvent

être faites par l'analyse des informations que ces critères détiennent. Il existe plusieurs approches dans la littérature qui ont abordé le problème de la reconnaissance de la classe d'âge. Ces approches peuvent être classées en deux catégories : i) les cranio-faciales, qui traitent seulement les changements qui se produisent sur le visage ou bien le crâne, ii) les approches par la démarche, qui se focalisent sur l'analyse de la variation de la silhouette et de la posture de la personne lors de son déplacement.

3.3.1 Les approches cranio-faciales

L'impact du vieillissement qui touche l'être humain peut notamment être observé sur son visage à travers l'apparition des rides, le changement de la couleur de la peau par l'apparition des tâches noires et/ou brunes et le blanchissement des cheveux [123]. En effet, l'apparition des rides est due au mouvement perpétuel des muscles du visage et au relachement de la peau. La direction de ces déformations est toujours perpendiculaire à la direction du mouvement [123]. En outre, l'apparition des tâches noires sur la peau du visage a pour origine l'endommagement des mélanocytes (les cellules productrices de la mélanine) à cause de l'exposition de ces derniers aux rayons ultraviolets issus du soleil, qui excite le taux de production de la mélanine [123]. Le crâne est touché par un ensemble des changements morphologies affectant sa taille et ses proportions. Plus précisément, l'étude [127] montre que la tête des enfants possède une taille exagérée rapporté aux membres et au tronc par rapport à celle des adultes, comme illustré à la figure 3.2.



FIGURE 3.2 – Evolution de la morphologie du crâne chez l'être humain avec l'âge [127] : Le crâne qui se trouve au centre représente celui d'un enfant, celui qui se trouve à l'extérieur représente le crâne d'un homme mature.

Ainsi, les effets du vieillissement apparaissent via des changements sur la forme générale du visage ou bien sur des endroits précis par des modifications affectant l'apparence ainsi la qualité de la peau. Par conséquent, les approches cranio-faciales qui existent dans la littérature peuvent être classées en trois catégories :

1. Les approches basées sur la forme, qui exploitent l'évolution morphologique des dimensions cranio-faciale et discriminent les différentes classes d'âge, soit à base de mesures de distances entre les composants du visage, (yeux, nez, bouche...) [128, 129], soit à partir d'un modèle de formes constitue à partir des points d'intérêts tel que l'ASM (Active Shape Model) [130].
2. Les approches basées sur la texture [131] qui exploitent la non homogénéité de la texture du visage chez les sujets âgés. Elles se focalisent sur les changements

qui touchent la texture de la peau dans des endroits précis sur le visage tels que : le front, les yeux, les joues ou encore l'entourage de la bouche, comme l'apparition des rides et la variation de couleur du visage. Ces caractéristiques ont été modélisées à partir de quelques descripteurs d'apparence tels que les ondelettes de Gabor [132], le LBP (Local binary patterns) [133], le LPQ (Local Phase Quantization) [134] et le HOG [133].

3. Les approches hybrides [130], qui combinent les informations de la texture et celles de la forme afin d'avoir un seul modèle de données riche en informations et plus performant. A titre d'exemple de ces approches, on peut citer le processus abordé par [135] qui détecte les points d'intérêts sur l'image du visage par l'extraction des caractéristiques globales (forme) en se basant sur le modèle d'apparence active AAM (Active Appearance Model). Ensuite, le LPQ a été appliqué pour extraire des vecteurs caractéristiques correspondants aux basses fréquences (texture). Dans un autre travail [130], l'extraction de la caractéristique "form" se réalise en se basant sur le modèle ASM qui sert à détecter la forme du visage par un ensemble de 68 points. Ensuite, une application de la transformation de radon a été réalisée sur l'image résultante d'ASM pour extraire les caractéristiques de texture qui apparaissent sur le visage d'un individu mature. La triangulation de Delauney a aussi été utilisée pour caractériser la forme du visage [136] et extraire les régions qui localisent les yeux, le nez et la bouche sur le visage. La texture et les basses fréquences (rides) sont quantifiées à partir des transformés de Hilbert et le LBP.

L'estimation automatique de l'âge de la personne est souvent faite à partir de l'analyse du visage, jugé comme la partie de corps dont les effets du vieillissement sont les plus visibles. En effet, les approches cranio-faciales représentent une majorité des travaux déjà abordés dans cette thématique de recherche. En revanche, en s'appuyant sur l'étude des approches cranio-faciales, déjà présentées, nous déduisons que l'estimation de l'âge à partir de l'analyse des images du visage nécessite la présence de données suffisamment adéquates en termes de niveau de définition et de précision. Le visage doit être assez clair et rapproché afin que les vecteurs caractéristiques puissent être extraits. Du fait de la distance assez importante du piéton par rapport au matériel d'acquisition, ainsi que de l'intervention non coopérative des piétons, l'obtention de données convenables à partir du visage pour l'estimation précise de l'âge dans un milieu extérieur, à savoir la rue, n'est pas toujours garantie. Cette limitation nous a encouragé à rechercher d'autres pistes assurant la résolution de ce problème à savoir le recours à la démarche humaine comme étant une méthode non intrusive pour l'estimation de l'âge.

3.3.2 Les approches utilisant la démarche

La démarche de l'homme évolue avec l'âge pour atteindre sa maturité vers l'âge de 7 ans, une fois que l'enfant arrive à marcher de manière stable et régulière. Cependant, après l'âge de 60 ans, la performance de la marche diminue rapidement [137]. Cette diminution est due notamment à la réduction de la force des principaux muscles moteurs des jambes [126] à savoir les muscles soléaires et gastrocnémien. La vélocité, ainsi que la longueur du pas, varient significativement d'une classe d'âge à l'autre [126, 137,

138]. Par conséquent, la démarche constitue un élément clef pour la détection de la classe d'âge, garantissant un apport prometteur [126] lorsque les approches intrusives semblent être non efficaces, voire impossibles.

Bien que les caractéristiques du vieillissement soient facilement détectables sur la posture et la démarche des personnes âgées, la démarche n'est pas encore très utilisée comme classificateur d'âge [126, 125] comme c'est le cas pour la reconnaissance de genre [139, 140, 141, 142, 143] ou l'identification des personnes [144, 145, 146].

Une revue de la littérature montre que peu d'approches ont abordé le problème de l'estimation automatique de l'âge en fonction des caractéristiques de la démarche. Plusieurs raisons peuvent être avancées. Parmi elles, l'analyse de la démarche et de la posture des piétons en situation, à partir d'images vidéo classiques, est une thématique de recherche assez récente [125]. L'absence de bases de données intégrant des personnes en situation de locomotion avec différentes catégories d'âge constitue également un frein à ce type de recherches.

Cependant, les approches de reconnaissance de l'âge par la démarche peuvent être classées selon deux catégories, à savoir les approches basées sur la forme de la silhouette, et celles basées sur son contour. La première catégorie consiste à étudier les déformations qui touchent la forme de la silhouette lors du déplacement. Par contre, les approches utilisant le contour visent à étudier les variations de distances entre les points du contour et le centroïde associé.

3.3.2.1 Les approches basées sur la forme de la silhouette

Plusieurs approches dans la littérature se basent sur la forme de la silhouette pour l'identification de l'âge ou de la classe d'âge d'un individu. Pour cela, il existe plusieurs techniques qui servent à extraire les informations de l'âge soit en utilisant l'image moyenne (Gait Enregie Image : GEI) d'une séquence d'images de silhouettes, soit par la représentation fréquentielle de la silhouette.

Ainsi, Makihara *et al.* [125] ont proposé une analyse de la marche de genre et d'âge pour étudier les différences de démarche entre les classes d'âge et les sexes, à savoir les jeunes, les adultes, les personnes âgées, les hommes et les femmes. Ils ont représenté la silhouette dans le domaine fréquentiel (Figure 3.3) pour extraire les caractéristiques de la démarche et ont fourni diverses expériences de classification.

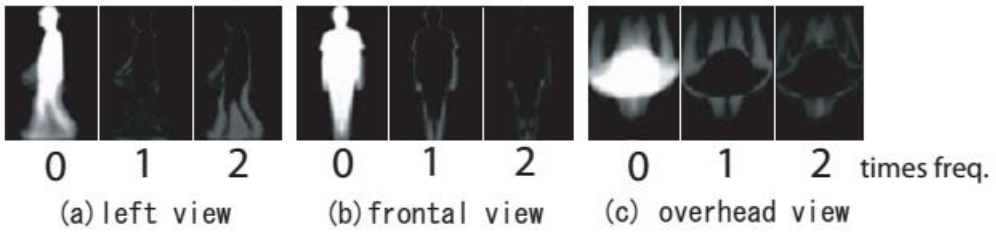


FIGURE 3.3 – Représentation fréquentielle de la GEI [125].
Les valeurs entre 0 et 2 représentent la transformation discrète de Fourier (TFD) appliquée zero, une et deux fois respectivement.

Cette étude présente une analyse importante de l'unicité de la démarche pour chaque classe afin d'obtenir un aperçu des différences de démarche entre les sexes et entre les classes d'âge. Avec le vieillissement, des changements importants dans les paramètres de la démarche sont rapportés. Par rapport aux jeunes, les plus âgés ont une plus faible vitesse de marche, la longueur des pas et de la foulée sont aussi réduits. De même, le temps de la phase de simple support est considérablement réduit chez les personnes âgées au profit de la période de double appui.

L'approche introduite dans [122] met en œuvre un algorithme basé sur une régression Gaussienne (GPR) pour l'estimation de l'âge. Le processus de régression gaussienne est constitué de deux étapes. La première étape consiste à dériver la distribution gaussienne de la variable dépendante $'f'$ à partir du régresseur $'x'$ (les caractéristiques de la démarche chez l'individu) et le paramètre de régression $'\theta'$ (paramètre explicatif). Ensuite, la distribution gaussienne de l'âge observé $'y'$ est dérivée à partir de la variable dépendante $'f'$ et du bruit d'observation σ. La figure 3.4 illustre les étapes du processus de régression gaussienne utilisé dans [122].

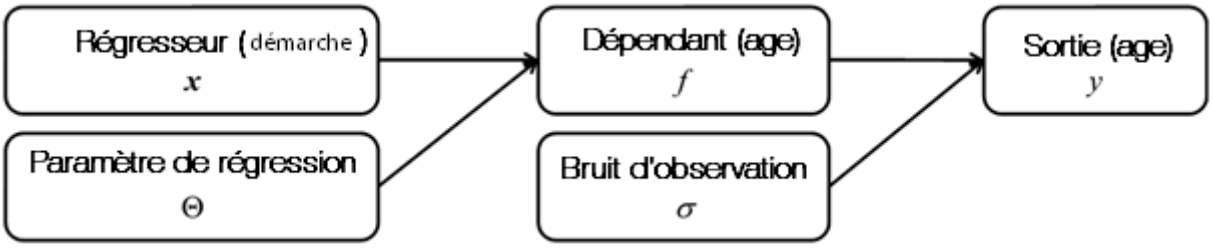


FIGURE 3.4 – Diagramme du processus de régression gaussienne [122].

La phase d'apprentissage pour le GPR est assurée par l'ensemble D qui contient deux sous-ensembles à savoir : l'ensemble $'x'$ des paramètres de la démarche de N personnes et l'ensemble $'y'$ des âges réels correspondants.

$$D = [x, y] : \ ensemble\ d'apprentissage.$$
$$x = [x_1, x_N] : \ ensemble\ de\ la\ démarche.$$
$$y = [y_1, y_N] : \ ensemble\ d'\hat{a}ges\ réels.$$

Lors de la phase de test, le GPR reçoit un nouvel ensemble x^* contenant les paramètres de la démarche des N personnes inconnues, pour estimer la distribution de leurs âges approximatifs.

Les régresseurs ou les paramètres de la démarche utilisés dans [122] sont : L'image moyenne de la silhouette (GEI), la représentation fréquentielle de la GEI (FREQ [122]) et la période de la démarche (GP), c'est-à-dire le temps nécessaire pour qu'un individu accomplisse un cycle de marche complet.

Le recours à plusieurs descripteurs dans [122] a permis d'explorer leurs performances en termes d'estimation de l'âge par le biais de l'erreur absolue moyenne MAE et du

score cumulatif S(l) donnés par l'équation 3.1

$$MAE = \frac{1}{N^t} \sum_{i=1}^{N^t} |\hat{y}_i^t - y_i^t| \tag{3.1}$$

$$S(l) = \frac{N^t(l)}{N^t}$$

Avec N^t qui désigne le nombre d'échantillons de test, y_i^t l'âge réel, $\hat{y}_i^t$ l'âge estimé et l le nombre d'années d'erreur.

En fait, le GEI et sa représentation fréquentielle (FREQ) ont été les plus performants en obtenant les plus petites erreurs moyennes, en l'occurrence 8 et 8,2 années respectivement, ainsi que les meilleurs scores cumulatifs. Le GEI arrive en deuxième place avec une différence très négligeable par rapport au FREQ (qui n'est qu'une représentation fréquentielle de GEI). Le GEI conserve les informations statiques et dynamiques de la démarche dans une représentation bidimensionnelle (2D). Par contre, l'utilisation du FREQ (illustré dans la figure 3.3, selon [125]) a pour objectif de conserver les informations statistiques et dynamiques apportées par le GEI avec l'avantage d'avoir un format plus compact que celui du GEI en matière de taille du vecteur caractéristique. En fait, bien que le GEI ait été largement utilisé dans la littérature pour l'analyse de la démarche et l'estimation de l'âge, il a quelques inconvénients tel que la manipulation des données de la silhouette entière, sans se concentrer sur les régions les plus discriminantes. Cette limitation conduit le GEI à être un descripteur lourd qui encapsule des données très détaillées (inutiles dans certains cas) et qui dimine la qualité des traits discriminants.

L'apport de [122] est important en matière d'estimation de l'âge ainsi que par l'élaboration d'une base de données assez large et qui contient de nombreux individus, jeunes et âgés avec une multitude de prises de vue. Cependant, l'environnement où les données ont été collectées reste contrôlé et assez différent de notre contexte de recherche qui s'intéresse aux données prises dans un environnement extérieur.

Les travaux présentés précédemment montrent que les caractéristiques biométriques douces, telles que l'âge, peuvent également être dérivées des modèles de démarche. En effet, les auteurs dans [147] utilisent un filtre de Gabor amélioré et la maximisation de l'information mutuelle pour extraire des caractéristiques à faibles dimensions. Les ondelettes de Gabor sont appliquées pour l'extraction de caractéristiques, qui décomposent la silhouette en orientations et échelles locales.

L'étude expérimentale, menée avec une phase d'apprentissage supervisée et la classification à base de modèle de Markov caché (HMM), montre des résultats très encourageants, même s'ils sont réalisés sur un petit ensemble de données, (sept personnes pour chaque groupe d'âge). En effet, le nombre d'échantillons de test réduit et l'utilisation des HMM restent l'inconvénient majeur de cette étude.

3.3.2.2 Les approches basées sur le contour

L'utilisation du contour de la silhouette dans un contexte d'estimation automatique de la classe d'âge consiste à étudier les variations de la forme du contour à travers les coordonnées des points du contour au cours d'un intervalle de temps donné. L'approche proposée dans [126] consiste à modéliser le cycle de marche par un HMM avec les distances entre les points du contour et le centroïde associé comme descripteur appelé FED (Feature to Exemplar Distance).

Afin de réduire la dimension du vecteur caractéristique, chaque cycle de marche est subdivisé en 5 segments égaux. Le vecteur moyen de distance est calculé pour chaque segment. Le vecteur caractéristique utilisé (FED) [148] représente donc la distance euclidienne entre le vecteur de chaque image et le vecteur moyen de chaque segment.

L'utilisation du contour et du vecteur FED a permis de réduire la taille du vecteur caractéristique comparé à celui du GEI, ainsi que l'obtention d'une plus grande précision de classification (taux des échantillons correctement classés) soit 83,33 % par rapport à celle obtenue en utilisant la silhouette (76,24 %). En revanche, le corpus de travail de cette approche est très limité, avec seulement 14 sujets. De plus, une seule prise de vue (fronto-parallèle) a été prise en considération, ce qui représente un environnement restreint pour notre contexte de travail.

Par ailleurs, le HMM est connu pour sa nécessité d'une phase d'apprentissage hors ligne pour la construction des modèles qui est couteuse en matière de temps de calcul en comparaison avec les autres techniques de classification. La matrice de transition est également connue pour sa taille importante nécessitant un temps de calcul conséquent qui augmente en fonction de N.

Certains paramètres géométriques de bases tels que la fréquence de la foulée et le rapport tête/corps sont introduits dans [149] afin de discriminer l'adulte de l'enfant. Ces paramètres sont calculés en se basant sur l'image contour de la silhouette de la personne afin de localiser d'abord les points d'intérêts. Ces points sont la tête, le cou, les hanches, les genoux et les chevilles.

Cette étude est évaluée sur une petite partie de la base de données OU-ISIR (20 personnes) à l'aide d'un classifieur linéaire (SVM). Par conséquent, le bon résultat obtenu peut être influencé par le petit nombre de sujets utilisés pour l'expérimentation. Il est ainsi, conjecturé que les performances pourraient être dégrader avec l'utilisation de la base OU-ISIR complète [149].

En se basant sur cette étude de l'existant, nous pouvons conclure que la reconnaissance de l'âge est une thématique qui fait partie de la biométrie douce. Elle cherche à reconnaître l'âge d'une personne à partir de certaines caractéristiques. Sachant que le visage est parmi les parties du corps les plus visiblement touchées par le vieillissement, la plupart des approches l'utilisent pour la reconnaissance de l'âge. Cependant, malgré les avantages concrets de ces méthodes, la reconnaissance demande une certaine coopération de la personne étudiée, ainsi qu'un environnement bien contrôlé, afin d'assurer la pertinence des résultats. Ceci n'est pas le cas pour notre sujet d'étude (traversée de rues) où le sujet se trouve à une distance plus ou moins importante (longueur du

passage piéton de $\simeq 6\,m$) par rapport au matériel d'acquisition. De plus, le piéton est filmé sans qu'il soit prévenu, par conséquent il ne sera pas conscient de la procédure d'analyse et il est fortement probable que son visage ne soit pas clairement visible. Pour ces raisons, la démarche et l'analyse de la posture constituent une solution plus adéquate pour la reconnaissance des sujet âgés des personnes distantes, par rapport au matériel d'acquisition.

Par conséquent, nous proposons un nouveau descripteur pour l'estimation automatique de la classe d'age des piétons. Cette proposition s'appuie sur des recherches médicales et biomécaniques [150, 151] qui confirment que les oscillations des bras, l'inclinaison de la tête, la courbure de la posture et la longueur de la foulée sont parmi les caractéristiques du vieillissement les plus remarquables lors de la marche des personnes âgées. Le descripteur proposé représente, alors, les variations spatio-temporelles de la démarche et tire profit de ces caractéristiques sous hypothèse d'une meilleure identification des personnes âgées.

3.4 Approche proposée pour la reconnaissance de la classe d'âge

3.4.1 Processus de pré-traitement

L'extraction des caractéristiques de l'âge nécessite une phase de préparation des données constituées d'un ensemble d'opérations qui servent à convertir les séquences vidéo d'une personne en images binaires de silhouettes non bruitées. Deux phases s'avèrent nécessaires à savoir l'extraction de la personne et le nettoyage des silhouettes. La figure 3.5 illustre les différentes étapes de pré-traitement appliquées sur l'ensemble des silhouettes.

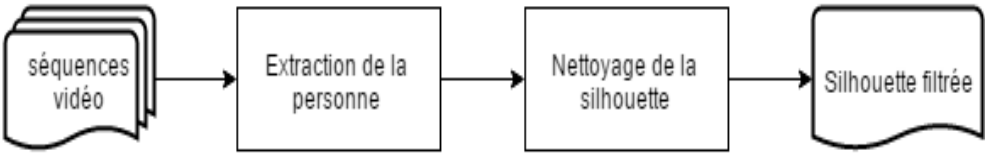


FIGURE 3.5 – Diagramme du processus de pré-traitement proposé.

L'étape d'extraction de la personne a déjà été présentée et détaillée dans le chapitre 2 (section 2.3.3.1). Elle est basée sur la technique de l'HOG par bloc comme descripteur avec le SVM comme classifieur. Après avoir localiser la personne, une soustraction de l'arrière-plan est utilisée pour récupérer la silhouette de la personne sous la forme d'une image binaire.

L'image binaire obtenue contient la silhouette de la personne mais aussi d'autres régions de taille réduites, comme illustré à la figure 3.6 a. Quelques trous sont aussi présents dans la silhouette. Le nettoyage de celle-ci consiste à 'éliminer' ces régions de tailles réduites et à 'boucher' les trous par l'application d'opérateurs morphologiques. En se basant sur une étude expérimentale, les opérations morphologiques que nous avons adoptées pour le nettoyage de la silhouette sont une érosion suivie d'une fermeture

(dilatation suivie d'une érosion) avec un élément structurant de type carré de taille (3×3). Un exemple de silhouette nettoyée est représenté à la figure 3.6 b.



FIGURE 3.6 – Exemple de silhouette nettoyée par l'application des opérateurs morphologiques : a) Silhouette bruitée, b) Silhouette nettoyée.

3.4.2 Détection des cycles de marche

Nous ne considérerons que les informations d'un cycle de marche par personne pour étudier les caractéristiques de l'âge, afin de synchroniser les traitements pour les différents descripteurs.
La définition d'un cycle de marche varie d'une approche à l'autre dans la littérature. Par exemple, dans [125] le cycle de marche est vu comme une double alternance d'attaques de talons droit et gauche respectivement, tandis que pour [152] il est vu comme un intervalle de temps qui sépare deux instants d'attaque de talon du pied initial, c'est à dire deux pas. La figure 3.7 illustre ce concept.

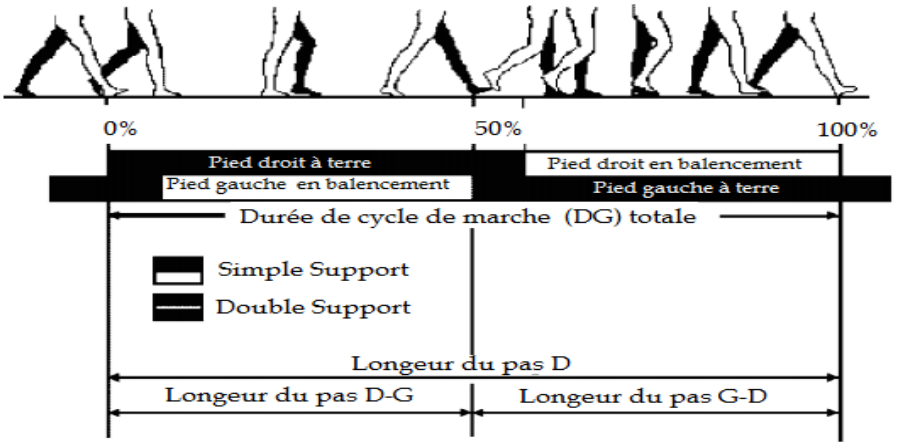


FIGURE 3.7 – Diagramme du cycle de marche [152].

La détection des cycles de marche consiste à identifier les images des silhouettes qui constituent un cycle fermé en se basant sur la variation de la distance entre les chevilles droite et gauche.

La marche est un processus répétitif qui se traduit par l'avancement des deux pieds en avant d'une façon alternée et qui comporte deux phases à savoir, la phase de double support et celle de simple support. Afin de détecter le cycle de marche, nous nous somme basés sur la variation du pas qui est faite en utilisant la distance entre les deux chevilles. Les positions des deux chevilles sont calculées dans les deux phases de la marche en se basant le processus proposé dans [153]. Le résultat d'application de ce processus sur une silhouette est illustré par la figure 3.8.

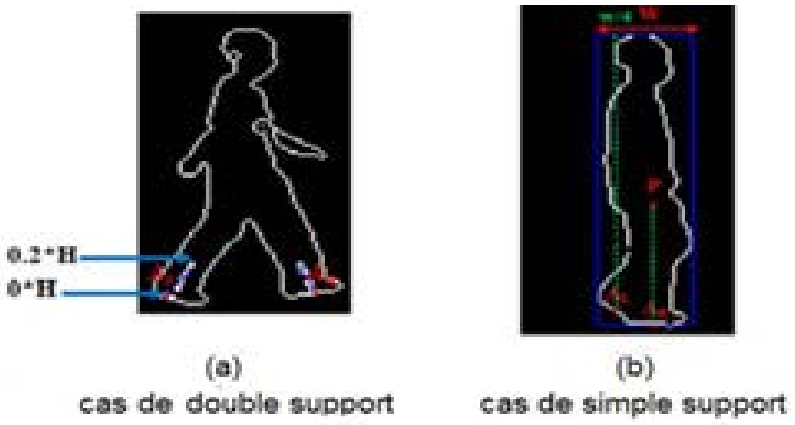


FIGURE 3.8 – Détection des positions des chevilles [153].

En se basant sur la courbe de variation du pas présentée à la figure 3.9, tout en prenant la position de double support (pas maximal) comme un point de départ, le cycle de marche est identifié par la succession de trois maxima du pas. Par conséquent, les images qui seront prises en considération durant les analyses à venir sont celles d'un cycle de marche, ce qui a pour objectif d'assurer que tous les traitements soient synchrones et cohérents.

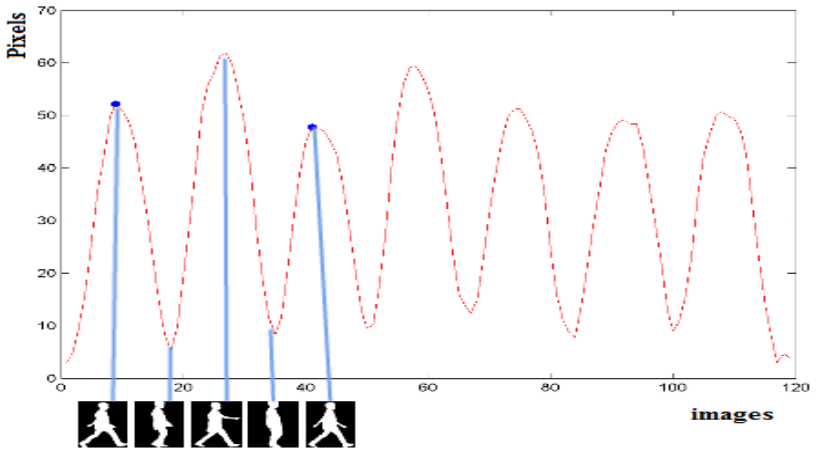


FIGURE 3.9 – Variation de la longueur du pas et détection d'un cycle.

3.4.3 Descripteur proposé : modèle de la silhouette

En plus des paramètres évoqués précédemment, notamment spatio-temporels, l'analyse des caractéristiques de la marche des personnes âgées, révèle également qu'elles ont tendance à davantage bouger les bras que les sujets plus jeunes afin de garder l'équilibre, que le dos est souvent faiblement ou fortement courbé [122, 125], mais aussi que la vitesse de déplacement se dégrade avec l'âge et que la longueur des pas diminue [126, 137, 138].

En se basant sur ces conclusions, nous proposons un nouveau descripteur qui englobe à la fois les variations spatiales mais aussi temporelles affectant la silhouette de la personne lors de la marche. Les caractéristiques citées précédemment sont visiblement détectables par l'analyse de la séquence des silhouettes, soit en largeur soit en hauteur. Notre descripteur consiste alors à analyser les variations de la morphologie de la silhouette, et ce, à travers une projection longitudinale (largeur de l'image) et une autre transversale (hauteur de l'image) appelées respectivement **SLP** (Projection Longitudinale de la Silhouette) et **STP** (Projection Transversale de la Silhouette).

3.4.3.1 Projection transversale de la silhouette

La projection transversale (STP) consiste à étudier la distribution transversale des pixels de la silhouette, ligne par ligne, en comptant le nombre de pixels blancs comme l'illustre la figure 3.10.

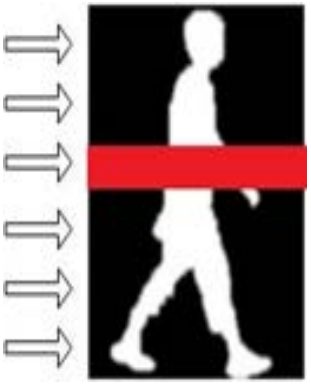


FIGURE 3.10 – Projection transversale de la silhouette.

Le vecteur caractéristique STP d'une image de silhouette I est donné par l'équation 3.2 :

$$STP_i = \sum_{j=1}^{N} I(i,j) \tag{3.2}$$

Avec i qui désigne le numéro de ligne de l'image variant de 1 à M et N désigne le nombre de colonnes de l'image.

Sachant que la période d'un cycle de marche varie d'une personne à l'autre, le vecteur caractéristique de la projection transversale pour un cycle de marche complet, STP_{Cycle} est, la moyenne des vecteurs STP de chaque image de la silhouette donné par

l'équation 3.3.

$$STP_{Cycle} = \frac{1}{T}\sum_{t=1}^{T} STP_t \tag{3.3}$$

Où $'T'$ représente le nombre d'images dans un cycle de marche.
Un exemple de la variation transversale de la silhouette est présenté à la figure 3.11.

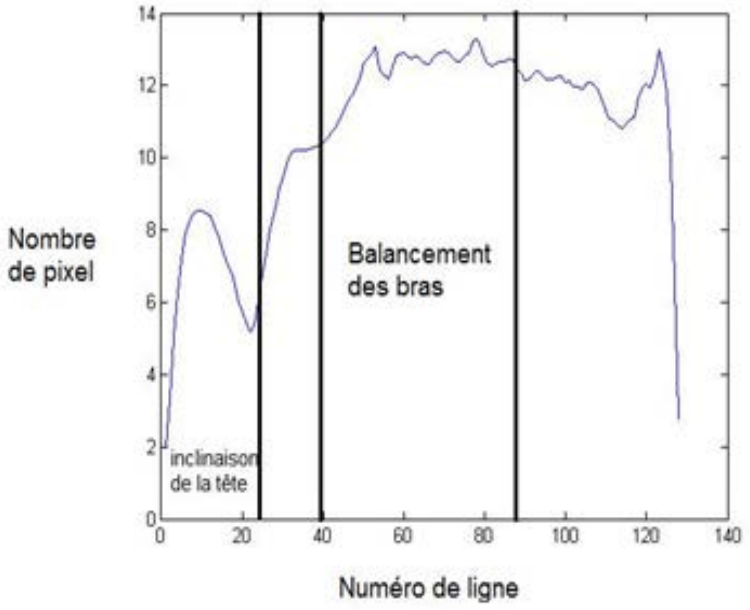


FIGURE 3.11 – Variation transversale d'une silhouette durant un cycle de marche.

Le vecteur caractéristique STP_{Cycle} représente un modèle statistique qui décrit les variations spatio-temporelles transversales de la forme de la silhouette pour chaque personne. Ces caractéristiques servent à dévoiler l'existence d'un ensemble de traits discriminants les sujets jeunes des âgées, à savoir l'inclinaison de la tête et le balancement des bras.

3.4.3.2 Projection longitudinale de la silhouette

La projection longitudinale (**SLP**) consiste à étudier la distribution longitudinale des pixels de la silhouette via le parcours de l'image de la silhouette, colonne par colonne, afin de compter le nombre des pixels blancs dans chaque colonne comme représenté à la figure 3.12.

Le vecteur caractéristique de la projection longitudinale SLP d'une image de silhouette est donné par l'équation 3.4.

$$SLP_j = \sum_{i=1}^{M} I(i,j) \tag{3.4}$$

Avec j qui désigne le numéro de colonne de l'image variant de 1 à N et M désigne le nombre de lignes de l'image.

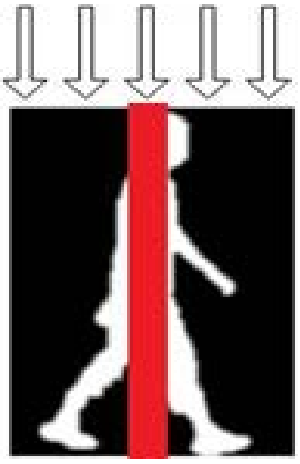


FIGURE 3.12 – Projection longitudinale de la silhouette.

Le vecteur caractéristique SLP_{Cycle} d'un cycle de marche complet est ainsi obtenu par la moyenne des vecteurs SLP de chaque image du cycle (équation 3.5).

$$SLP_{Cycle} = \frac{1}{T}\sum_{t=1}^{T} SLP_t \tag{3.5}$$

Où $'T'$ désigne le nombre d'images d'un cycle de marche.

Un exemple de la variation longitudinale de la silhouette est présenté à la figure 3.13.

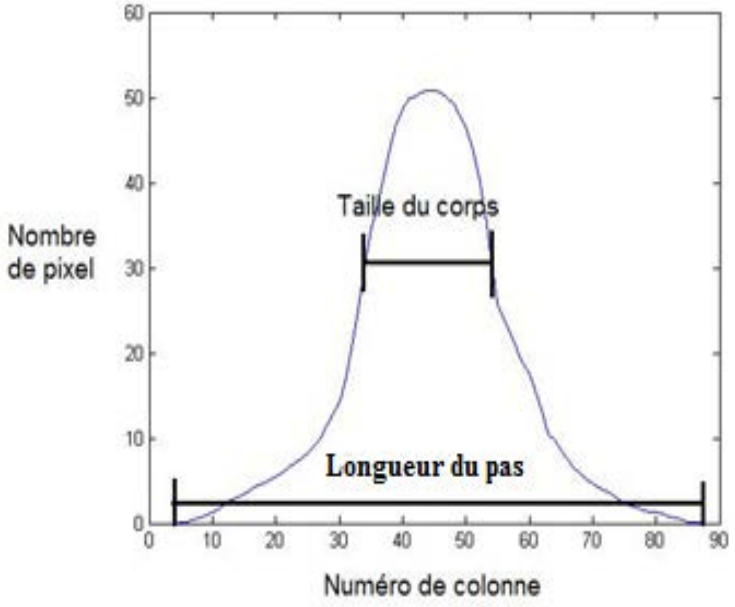


FIGURE 3.13 – Variation longitudinale d'une silhouette durant un cycle de marche.

La construction du vecteur caractéristique de projection verticale d'un cycle de marche complet permet, ainsi, de modéliser la variation de la configuration verticale de

la silhouette tout au long d'un cycle de marche. Les caractéristiques les plus recherchées par la projection longitudinale sont, la longueur du pas et la taille du corps.

3.4.3.3 Conception du modèle de la silhouette

L'analyse longitudinale de la silhouette a permis de décrire la longueur du pas et la taille du corps (Figure 3.13) alors que les autres paramètres de la démarche à savoir le balancement des bras, la courbure du dos et l'inclinaison de la tête sont visibles sur la projection transversale de la silhouette (voir figure 3.11).

Afin de bien cerner et suivre les variations de tous ces paramètres, nous avons opté pour concevoir un modèle de la silhouette englobant à la fois la projection transversale et celle longitudinale. Ce modèle consiste à concaténer ces deux projections comme illustré à la figure 3.14.

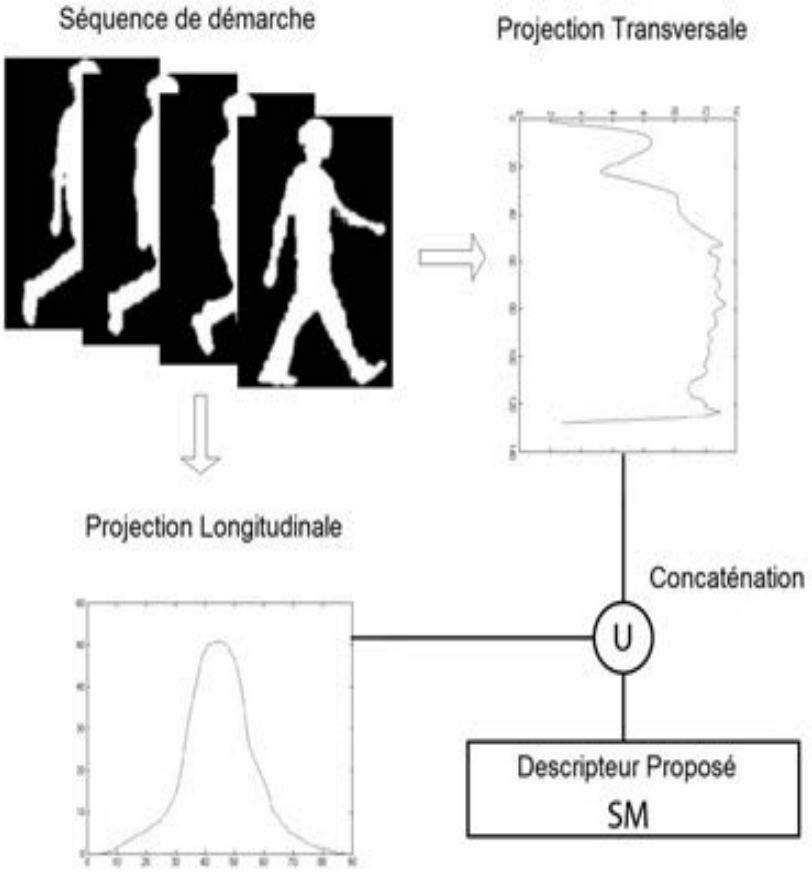


FIGURE 3.14 – Modèle de la silhouette.

3.5 Évaluation des performances

3.5.1 Résultats qualitatifs : Une analyse morphologique de la démarche

3.5.1.1 Les vecteurs moyens de SLP et STP

Pour illustrer les variations morphologiques significatives des paramètres de la démarche (balancement des bras, hauteur de la tête, longueur de la foulée ...) entre les personnes âgées et les jeunes, une comparaison entre les variations des mesures en pixels des silhouettes longitudinales et transversales, SLP_{Cycle} et STP_{Cycle} est réalisée.

La superposition des courbes de projection longitudinale de deux personnes, jeune et âgée, illustrée par la figure 3.15, montre une différence en longueur de pas et en taille du corps

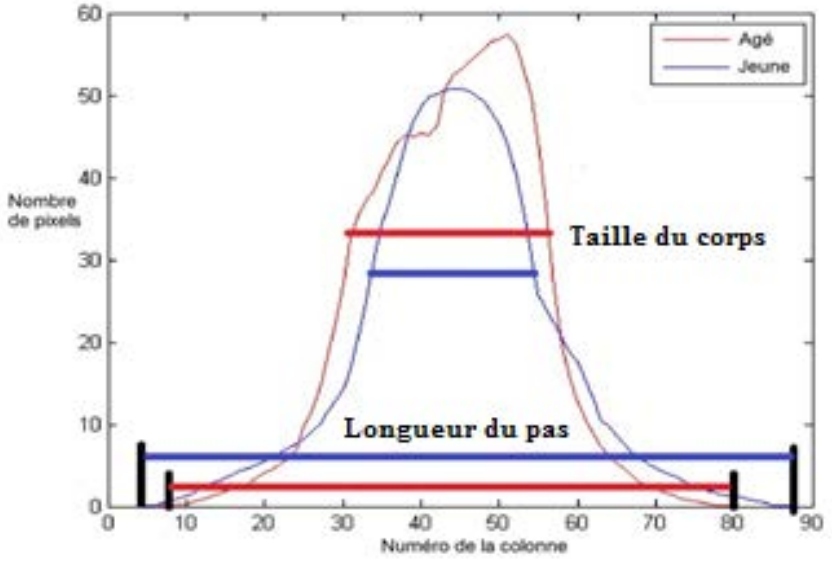


FIGURE 3.15 – SLP d'un sujet jeune (bleu) et d'un âgé (rouge).

De même, la superposition des courbes de projection transversale de deux personnes, jeune et âgée illustrée par la figure 3.16, montre une différence pour la densité du balancement des bras et l'inclinaison de la tête.

L'observation de la figure 3.15, montre que la taille du corps de la personne âgée est plus importante que celle du sujet jeune. Cependant, les jeunes ont une foulée plus grande que les personnes âgées. Les personnes âgées ont un balancement du bras plus fréquent. Aussi, seules les personnes âgées ont la tête inclinée et une posture plus large, qui peuvent être facilement identifiables sur la figure 3.16.

3.5.1.2 Les variances de SLP et STP

La variance est utilisée pour mettre en évidence la variation dynamique des paramètres de la démarche et illustrer les différences significatives entre le modèle de

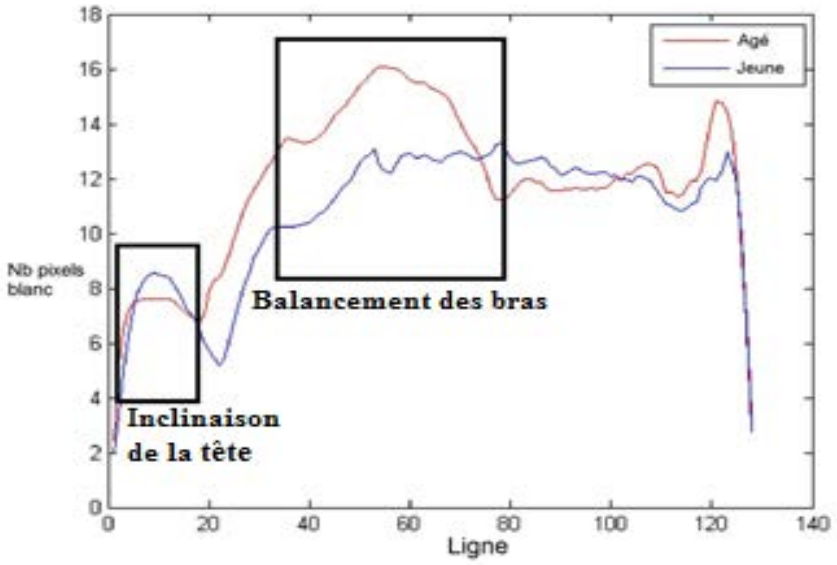


FIGURE 3.16 – STP d'un sujet jeune (bleu) et d'une âgé (rouge).

silhouette de la personne âgée et celui du jeune. Les variations temporelles de SLP_j et STP_j sont calculées, durant un cycle de marche de N frames, respectivement pour SLP_{Cycle} et STP_{Cycle} comme le formalisent les équations 3.6 et 3.7.

$$Var_{SLP} = \frac{1}{N}\sum_{j=1}^{M}(SLP_j - SLP_{Cycle})^2 \tag{3.6}$$

$$Var_{STP} = \frac{1}{N}\sum_{i=1}^{N}(STP_i - STP_{Cycle})^2 \tag{3.7}$$

La superposition des courbes de variances (Var_{SLP} et Var_{STP}) de deux personnes, jeune et âgée est illustrée aux figures 3.17 et 3.18.

L'observation de ces deux figures 3.17 et 3.18, montre que les différences des valeurs de variances des jeunes et âgés les plus importantes sont obtenues dans les régions d'intérêts pour les courbes SLP et STP : i) tête, ii) tronc et iii) bras.

Les valeurs importantes des écarts, représentés par des pics sur les courbes de variance, montrent que les piétons âgés ont davantage de mouvements irréguliers que les jeunes. En effet, les sujets âgés adaptent leurs façons de se déplacer pour une démarche plus sûre et plus stable. Ils ont tendance à balancer leurs bras davantage que les jeunes afin de conserver l'équilibre tout en marchant [125]. Le balancement plus important de la tête et la vitesse réduite sont également deux indicateurs du vieillissement qui caractérisent les personnes âgées.

De plus, la personne âgée a fréquemment une peur de chute élevée qui conduit à des mouvements plus fréquents, mais de petites tailles afin de maximiser l'équilibre, contrairement aux sujets jeunes qui se déplacent avec une foulée plus grande et maintiennent plus dynamiquement leur équilibre.

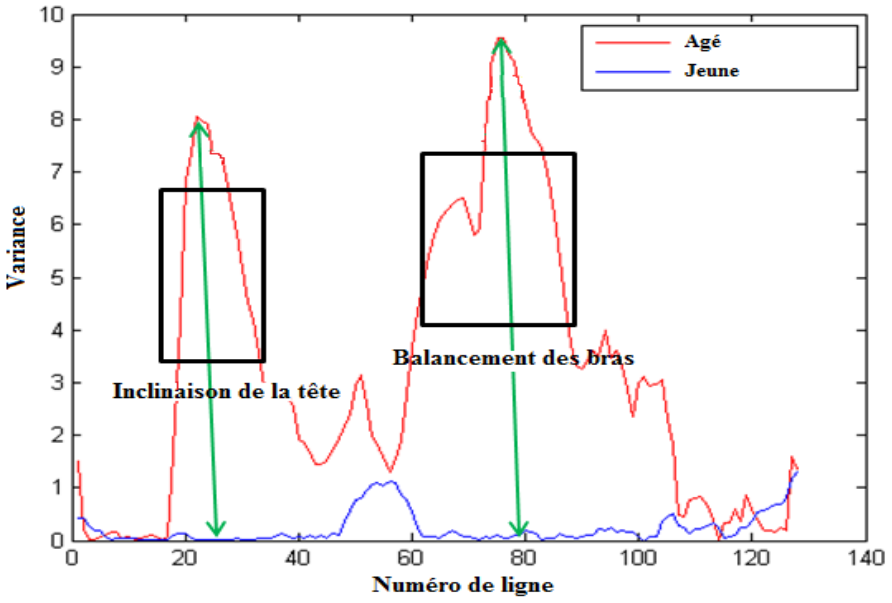


FIGURE 3.17 – Superposition des courbes de variances de SLP de la silhouette d'un sujet jeune (bleu) et d'un âgé (rouge).

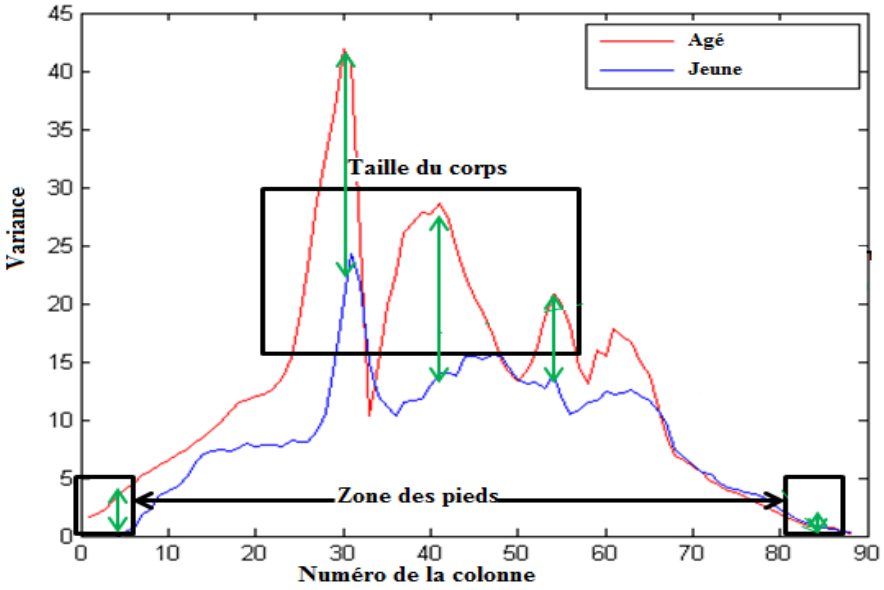


FIGURE 3.18 – Superposition des courbes de variances de STP de la silhouette d'un sujet jeune (bleu) et d'un sujet âgé (rouge).

3.5.1.3 Le modèle de Frieze de SLP et STP

Soit une séquence d'images de la silhouette $b(t) \equiv b(x, y, t)$ indexées spatialement par la localisation des pixels (x, y) et temporellement par le temps t. Le modèle $2D$ de SLP est formé en empilant des projections de colonnes pour former un modèle spatio-temporel SLP. Cependant, le modèle STP peut également être construit en empilant

les projections de lignes. Puisque la démarche humaine est périodique par rapport au temps, ici le temps t est limité à un cycle de marche. Un modèle bidimensionnel qui se répète le long d'une dimension est appelé un modèle de Frieze dans la littérature mathématique et géométrique. Les modèles de Frieze de SLP et STP d'un sujet jeune et d'un âgé sont illustrés à la figure 3.19.

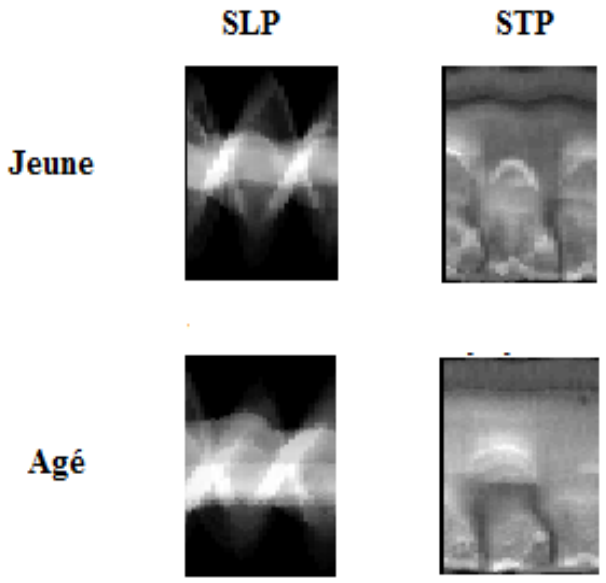


FIGURE 3.19 – Modèles de Frieze de SLP et STP des sujets jeunes et âgés.

La construction de modèles de Frieze de SLP et STP permet de visualiser la fréquence et l'intensité des mouvements des membres des piétons. En effet, en se référant à la figure 3.19, nous observons graphiquement une différence entre le modèle de Frieze STP et SLP d'un sujet jeune et celui d'un âgé. En effet, la personne âgée a une fréquence d'oscillation des bras plus importante mais avec une intensité plus faible (la surface de couleur gris clair est plus grande, mais l'amplitude de ces impulsions reste inférieure à celle du sujet jeune qui effectue un mouvement important mais moins fréquent). Cette observation peut s'expliquer par le fait que la personne âgée bouge et mentalise ou contrôle davantage ses mouvements car elle se focalise sur son équilibre. Par conséquent, sa vitesse diminue et sa cadence augmente contrairement à la personne jeune qui se déplace plus aisément avec une foulée plus grande et plus fluide.

L'analyse visuelle des projections des silhouettes transversale et longitudinale des personnes âgées et des jeunes tend à montrer l'existence d'une discrimination possible entre les classes d'âge. Toutes les observations précédemment mentionnées, déduites de l'analyse morphologique de la démarche et d'un modèle de la silhouette (SM), seront confirmés par une évaluation quantitative des performances du descripteur proposé.

3.5.2 Résultats quantitatifs

L'analyse des performances des descripteurs est assurée par l'intermédiaire d'une étude expérimentale sur une base de données populaire intégrant des exemples très

variés, tout en calculant plusieurs métriques et critères d'évaluation. Une étude comparative entre les différents descripteurs (existants et celui proposé) est également réalisée.

3.5.2.1 Base utilisée (corpus de travail)

Afin de tester les performances du descripteur proposé, nous avons choisi la base OU-ISIR [154] qui contient des séquences de marche d'une population très large en comparaison avec d'autres bases telles que USF-NIST [155], CASIA-C [156] et SOTON [157]. En effet, cette base contient 4007 personnes dont 2135 sont des hommes et 1872 sont des femmes. L'âge des personnes dans cette base varie de 1 à 94 ans. La figure 3.20 représente un échantillon de la base OU-ISIR qui montre la variation de l'âge et du sexe dans la population utilisée.



FIGURE 3.20 – Échantillon de données de la base OU-ISIR [154].

La base OU-ISIR permet d'analyser l'impact du vieillissement sur la démarche humaine en tenant compte de la grande variabilité des séquences proposées.

3.5.2.2 Protocole de test

Nous avons utilisé, comme échantillon de test, une population contenant 476 personnes dont 238 sont des sujets jeunes et 238 sont des âgés. Afin d'éviter le problème de sur-apprentissage d'une classe par rapport à l'autre, nous avons décidé de maintenir un équilibre entre la population des sujets jeunes et âgés.

Puisqu'il s'agit d'une bi-classification à savoir jeunes et âgés, nous avons sélectionné, à partir des données de chaque classe, 60% comme échantillons d'apprentissage et 40% comme échantillons de test. Ainsi, la portion d'apprentissage compte 284 personnes où 142 personnes ont été sélectionnées à partir de chaque classe jeunes et âges. Par ailleurs, la portion de test contient 192 personnes où 96 personnes ont été prises dans chaque classe.

Afin d'examiner la robustesse des différents descripteurs face à la variation des données, nous avons fait varier les échantillons d'apprentissage et de test. En effet, nous avons défini 5 expériences où les échantillons d'apprentissage et de test sont pris aléatoirement. En tant que méthode de classification, nous avons utilisé le SVM. Le SVM est une technique d'apprentissage supervisée pour résoudre les problèmes de classification surtout lorsqu'il s'agit exactement d'une classification binaire. Pour la tâche de classification, le SVM sépare les caractéristiques de chaque classe en trouvant le meilleur

hyperplan qui maximise la marge entre ces deux classes. Des nombreux noyaux (Linear, RBF, polynomial) peuvent être utilisés afin de résoudre le problème. Après avoir testé différents noyaux de SVM, nos données semblent pertinentes et linéairement séparables. En effet, nous avons utilisé la fonction linéaire du noyau (W.x), pour discriminer les personnes âgées et les jeunes.

3.5.2.3 Critères de performances

Au cours de ce travail, nous avons eu recours au rappel, à la précision et au taux de reconnaissance (TR), comme métriques afin d'illustrer les performances des descripteurs et leur capacité à discriminer entre les deux classes d'âge avec un minimum d'erreur. En effet, le rappel est une mesure de performance qui décrit l'aptitude d'un système à retourner des réponses correctes d'une classe donnée par rapport aux nombre des réponses retournées. La précision a pour objectif de décrire le potentiel d'un système à retourner des réponses exactes d'une classe donnée parmi la population de test de cette classe.

En outre, pour mettre en évidence les performances des descripteurs indépendamment du nombre de sujets dans chaque classe (nombre de sujets âgés et le nombre de sujets jeunes dans la base de données) et donner une évaluation graphique et plus expressive, la courbe ROC est adressée pour chaque descripteur et la valeur de la zone sous la courbe (AUC) est également calculée.

Sachant que :
— TY désigne le nombre de sujets jeunes correctement classés.
— FY désigne le nombre de sujets jeunes mal classés.
— TO désigne le nombre de sujets âgés correctement classés.
— FO désigne le nombre de sujets âgés mal classés.
Le rappel de chaque classe d'âge est donné par les équations 3.8 et 3.9.

$$Rappel_{jeunes} = \frac{TY}{TY + FO} \tag{3.8}$$

$$Rappel_{\hat{a}g\acute{e}s} = \frac{TO}{TO + FY} \tag{3.9}$$

De même, la précision de chaque classe d'âge est donnée par les équations 3.10 et 3.11.

$$Prècision_{jeunes} = \frac{TY}{TY + FY} \tag{3.10}$$

$$Prècision_{\hat{a}g\acute{e}s} = \frac{TO}{TO + FO} \tag{3.11}$$

Alors que le taux de reconnaissance est donné par l'équation 3.12.

$$TR = \frac{TY + TO}{TY + FY + TO + FO} \tag{3.12}$$

3.5.2.4 Performances du descripteur proposé

Les résultats expérimentaux du descripteur proposé (SM) en termes de *rappel*, *prècision* et *TR*, pour les cinq expériences (A, B, C, D et E), sont présentés dans le tableau 3.1. La courbe ROC est représentée à la figure 3.21.

TABLE 3.1 – Performances du descripteur proposé (SM).

Expérience	$Rappel_{Jeunes}$	$Rappel_{\hat{a}g\acute{e}s}$	$Prècision_{jeunes}$	$Prècision_{\hat{a}g\acute{e}s}$	TR
A	80	73,83	70,83	82,29	76,56
B	71,28	70,41	69,79	71,87	70,83
C	69,05	64,81	60,41	72,91	66,67
D	80,72	73,39	69,79	83,33	76,56
E	76,07	90,67	92,7	70,83	81,77
Moyenne	75,42	74,62	72,7	76,24	74,47

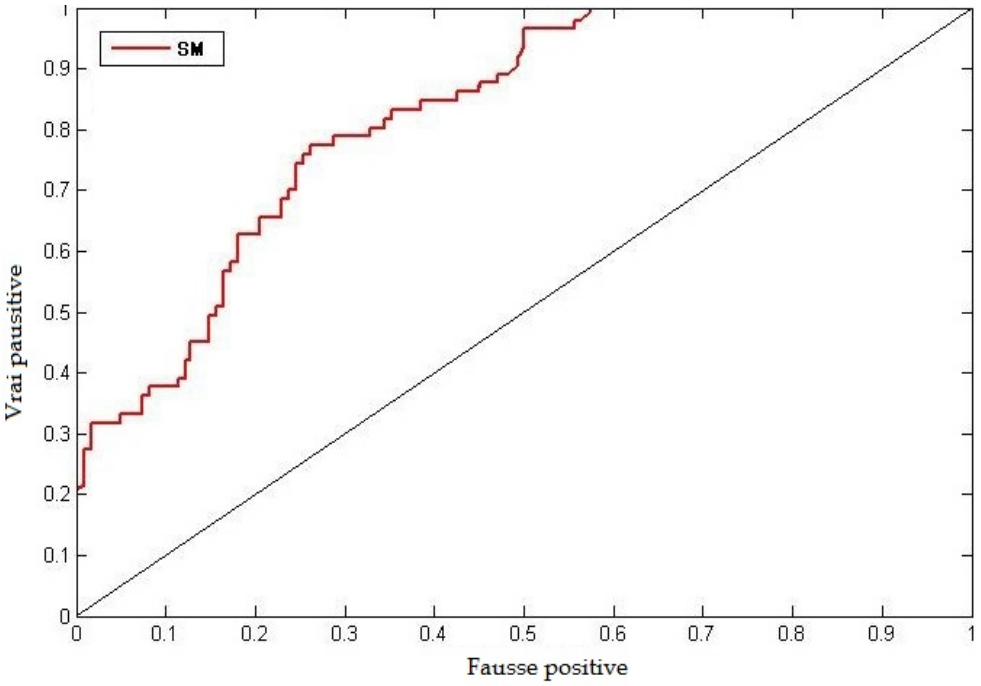


FIGURE 3.21 – La courbe ROC du descripteur proposé (SM).

D'après le tableau 3.1, nous constatons que le taux de reconnaissance du descripteur proposé a atteint une valeur moyenne de 74,47%. En termes de précision, nous observons une légère différence entre les précisions des deux classes (Jeunes et Âgés) qui sont égales à 72,7% et 76,24% respectivement. Les précisions des deux classes signifient que le SM associe convenablement les données à leurs classes exactes dans pratiquement 3 cas sur 4. En ce qui concerne les rappels, le $rappel_{jeunes}$ et $Rappel_{\hat{a}g\acute{e}s}$ sont presque égaux ce qui indique que le modèle de la silhouette (SM) discrimine bien entre les deux catégories d'âge.

En outre, la figure 3.21 montre que l'AUC sous la courbe ROC du descripteur proposé est de l'ordre de 0,83. Cela signifie que la probabilité de choisir correctement et de façon aléatoire un sujet âgé du groupe des âgés est de l'ordre 80%.

3.5.2.5 Robustesse du descripteur proposé vis à vis du genre

Nous avons également effectué un test supplémentaire pour mettre en évidence un éventuelle effet de genre sur la reconnaissance des groupes d'âge en fonction des caractéristiques de la démarche. Le test est effectué sur un sous-ensemble de la base OU-ISIR avec des sujets mâles et des sujets femelles séparément. Les résultats expérimentaux sont résumés dans le tableau 3.2.

TABLE 3.2 – Performances de descripteur proposé 'SM' vis-à-vis du genre

Expérience	$Rappel_{Jeunes}$	$Rappel_{\widehat{âgés}}$	$Prècision_{jeunes}$	$Prècision_{\widehat{âgés}}$	TR
Hommes	80,66	75,66	75	91,66	83,33
Femmes	52	43,67	67,21	49,45	58,33

Les résultats présentés dans le tableau 3.2 montrent que le descripteur fonctionne nettement mieux pour les hommes comparé aux femmes. Les bonnes performances sont enregistrées pour les sujets masculins avec un taux de reconnaissance égale à 83,33% comparativement à 58,33% pour les sujets féminins. Ces résultats peuvent justifier que le genre est parmi les principaux facteurs qui provoquent des erreurs dans la reconnaissance de l'âge en fonction de la démarche. En fait, la démarche et la posture de la femme peuvent être affectées par plusieurs facteurs autres que le vieillissement, comme la grossesse. Ainsi, par exemple les femmes enceintes peuvent avoir une grande taille, une petite foulée et des bras oscillants comme étant une personne âgée.

3.5.3 Étude comparative avec les approches existantes

Dans l'objectif de comparer les performances du descripteur proposé (SM) par rapport à d'autres approches de la littérature, nous avons choisi deux types de descripteurs :
i) Le GEI le plus performant descripteur basé sur la silhouette [122], ii) le FED, le meilleur descripteur basée sur le contour. Comme notre descripteur et ceux existants (GEI et FED) sont testés sur différentes bases en utilisant différents classifieurs et métriques d'évaluation, nous avons opté pour les re-implémenter et les tester sur la base OU-ISIR afin de fournir une étude comparative fiable.

Le GEI est une représentation spatio-temporelle d'une séquence d'images de silhouettes durant un cycle de marche [144], calculée en appliquant la formule de l'équation 3.13.

$$G(x,y) = \frac{1}{N}\sum_{t=1}^{N} I_t(x,y) \tag{3.13}$$

Où x et y désignent les coordonnées spatiales de l'image de la silhouette I_t retrouvées à l'instant t dans la séquence d'images. Le nombre d'images dans un cycle de marche est égale à N.

GEI est donc une image qui présente deux parties à savoir, une partie statique et une autre dynamique comme schématisé à la figure 3.22. La partie statique définie les parties du corps qui ne subissent aucun changement tout au long du cycle de marche, c'est à dire les parties du corps où aucun mouvement ne se produit. Elle est décrite par des pixels de couleur blanche. En ce qui concerne la deuxième partie, c'est-à-dire la partie dynamique, illustrée par une région floue dans la figure 3.22, elle définit les parties du corps qui présentent un mouvement durant le cycle de marche.



FIGURE 3.22 – Image moyenne d'une silhouette durant un cycle de marche : La lettre $'A'$ désigne la partie statique alors que la lettre $'B'$ représente la partie dynamique.

Les résultats expérimentaux du descripteur GEI présentés dans le tableau 3.3 montrent que le taux de reconnaissance atteint une valeur de 68,43 %. La précision pour les sujets âgés est plus grande que celle des jeunes avec une différence de 3,55 %, ce qui tend à montrer que GEI reconnaît mieux les sujets âgés que les jeunes. Par contre, les rappels des deux classes sont presque égaux, ce qui tend à montrer que ce descripteur maintient une classification équilibrée.

TABLE 3.3 – Performances du descripteur (GEI)

Expérience	$Rappel_{Jeunes}$	$Rappel_{âgés}$	$Prècision_{jeunes}$	$Prècision_{âgés}$	TR
A	69,38	70,21	70,83	68,75	69,79
B	64,58	64,58	64,54	64,58	64,58
C	70,24	65,74	61,45	73,95	67,71
D	74,16	70,87	68,7	76,04	72,4
E	67,71	67,71	67,71	67,71	67,71
Moyenne	69,21	67,82	66,65	70,2	68,43

Le descripteur FED consiste à calculer la distance euclidienne entre le centroïde de la silhouette et 60 points de son contour, associé durant un cycle de marche. Ensuite, chaque cycle de marche est subdivisé en cinq parties pour calculer le vecteur de distance moyen de chaque partie. Enfin, une distance euclidienne entre chaque vecteur de

distance d'un contour et les vecteurs de distances moyens sont calculées comme proposé dans l'étude [126].

Ces distances sont déterminées telles que deux points successifs soient séparés par un angle de 6 degrés tout en suivant le sens contraire des aiguilles d'une montre et en choisissant comme point initial celui qui intersecte avec l'axe horizontal et qui passe par le centroïde (Figure 3.23).

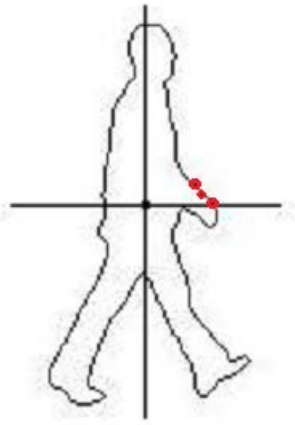


FIGURE 3.23 – Sélection des points de contour [126] : Le point qui se trouve au centre du contour désigne le centroïde.

Ainsi, le vecteur multidimensionnel des distances respectives résultant est exprimé comme suit dans l'équation 3.14.

$$C_{ij} = [D_1, D_2, D_3, ..., D_{60}] \tag{3.14}$$

Avec C_{ij} qui désigne le $i^{ème}$ contour de la $j^{ème}$ personne, D_{60} qui représente la $60^{ème}$ distance euclidienne entre le centroïde et le $60^{ème}$ point du contour.

Afin de réduire la dimension du vecteur caractéristique, chaque cycle de marche est subdivisé en 5 segments égaux. Le vecteur moyen e de C_{ij} est calculé pour chaque segment. Le vecteur caractéristique utilisé (FED) [148] représente ainsi la distance euclidienne entre le vecteur $'C'$ et le vecteur moyen $'e'_N$ des vecteurs de distances de l'étape en question, formalisé par l'équation 3.15.

$$FED_{ij} = [d(C_{ij}, e_1), ..., d(d(C_{ij}, e_N))] \tag{3.15}$$

Avec le *FED* qui désigne le vecteur des distances modèles de la $i^{ème}$ étape de la $j^{ème}$ personne et d la distance euclidienne.

Les résultats expérimentaux de ce descripteur, présentés au tableau 3.4, montrent que le taux de reconnaissance a atteint la valeur de 63,95%. Les précisions et les rappels des deux classes sont égales, ce qui tend à montrer que le FED reconnaît bien les deux catégories d'âge d'une façon équilibrée.

Le tableau 3.5 résume les performances de descripteur proposé *SM* et de deux existants (*GEI et FED*). De même, les courbes ROC de ces trois descripteurs comparés sont superposées à la figure 3.25.

TABLE 3.4 – Performances du descripteur (FED).

Expérience	$Rappel_{Jeunes}$	$Rappel_{âgés}$	$Prècision_{jeunes}$	$Prècision_{âgés}$	TR
A	62,65	59,63	54,16	67,7	60,94
B	62,62	65,88	69,79	58,33	64,07
C	59,63	62,65	67,7	54,16	60,94
D	70,93	66,98	63,54	73,95	68,74
E	65,26	64,95	64,58	65,62	65,1
Moyenne	64,21	64,01	63,94	63,95	63,95

TABLE 3.5 – Performances des différents descripteurs : étude comparative.

Expérience	$Rappel_{Jeunes}$	$Rappel_{âgés}$	$Prècision_{jeunes}$	$Prècision_{âgés}$	TR
FED	64,21	64,01	63,94	63,95	63,95
GEI	69,21	67,82	66,65	70,2	68,43
SM	75,42	74,62	72,70	76,24	74,47

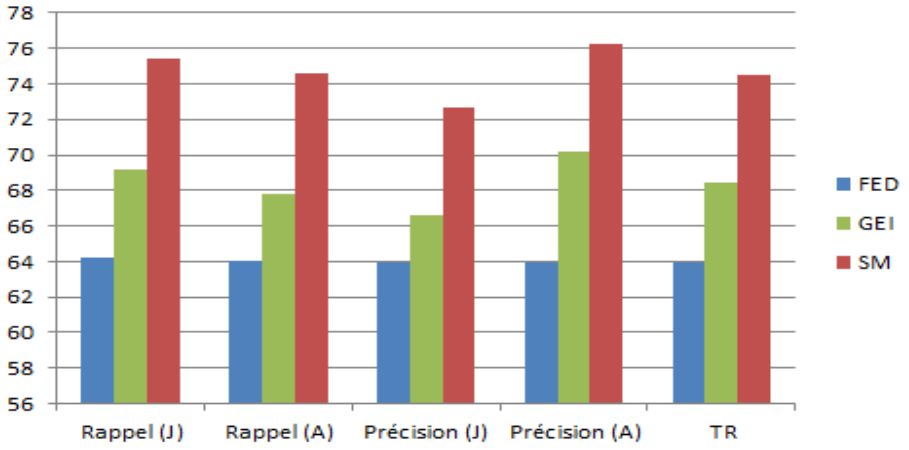


FIGURE 3.24 – Comparison des performances du descripteur proposé (SM) par rapport à ceux existants (GEI et FED) en termes des 'rappels' et 'précisions' pour les deux classes jeune (J) et âgé (A).

En se basant sur le tableau 3.5 et la figure 3.24 qui récapitulent les performances moyennes des différents descripteurs existants (GEI et FED) et celui proposé SM, nous pouvons ainsi déduire que :

— Le descripteur SM montre les résultats les plus satisfaisants comparé aux deux autres descripteurs, et ce notamment, au regard du TR moyen qui est égale à 74,47%. Ces performances élevées s'expliquent par la puissance du SM à représenter spécifiquement les caractéristiques les plus affectées par le vieillissement de la démarche humaine, à savoir, la longueur du pas et la taille du corps, facilement détectables par la projection longitudinale de la silhouette, mais aussi le

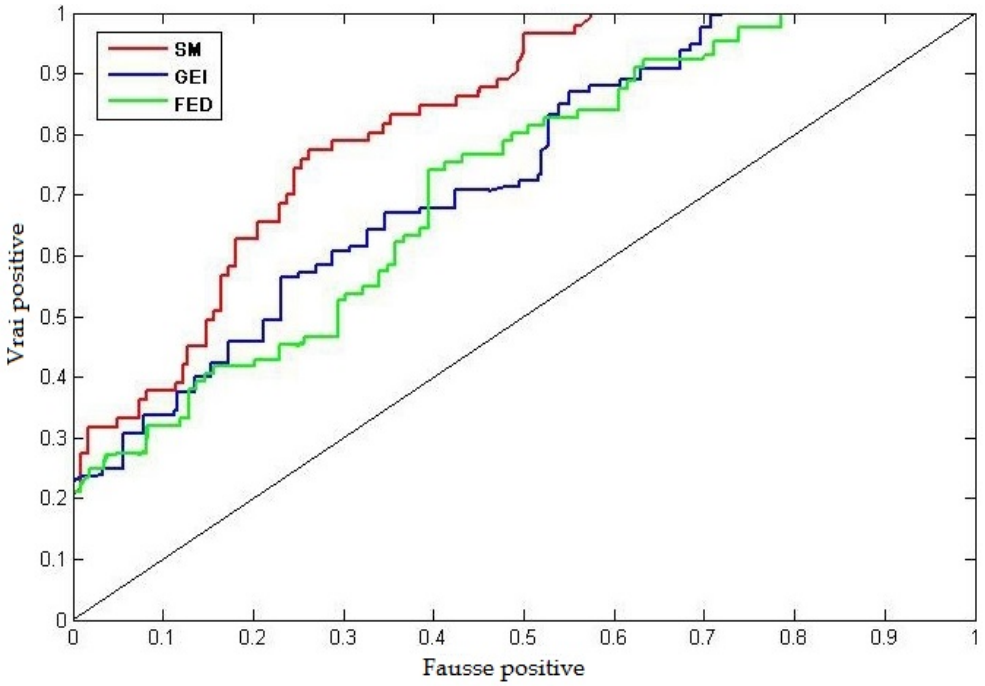


FIGURE 3.25 – Superposition des courbes ROC du descripteur proposé (SM) et de ceux existants (GEI et FED).

balancement des bras et la courbure du dos et de la tête, qui sont représentés par la projection transversale de la silhouette.

— Cette étude comparative montre que le descripteur proposé dépasse le descripteur d'état de l'art GEI par son taux de reconnaissance avec un bon rappel et une précision équilibrée entre les deux classes jeunes et âgés. Le GEI a été largement utilisé dans la littérature pour l'analyse de la démarche et l'estimation de l'âge. Nous notons cependant que l'étude de la variation de la silhouette, tout en se concentrant sur des régions spécifiques (bras, tête, pieds) ainsi que la fréquence du mouvement afin d'extraire les caractéristiques du vieillissement de la démarche, est plus pertinente que d'étudier l'ensemble du sujet comme c'est le cas du GEI.
— Le descripteur FED utilise des points discrets caractérisant le contour, alors que les autres descripteurs utilisent toute l'information de la silhouette et/ou du contour. L'information manquante due à la discrétisation du contour semble avoir causé une baisse du niveau des performances données par le descripteur FED, comparées à celles données par les autres descripteurs (GEI et SM).
— Les précisions et les rappels des deux classes jeunes et âgés sont presque égales, et nous pouvons constater que les descripteurs GEI et FED offrent une classification fiable, mais que notre descripteur SM reste le plus performant.
— En outre, les AUC enregistrés pour SM, GEI et FED sont respectivement de 0,83, 0,73 et 0,7 (Figure 3.25). Cela tend à monter la capacité de notre descripteur à mieux reconnaître les personnes âgés que le GEI et le FED.

3.6 Conclusion

En se plaçant dans des conditions réelles reflétant notre contexte de travail, à savoir l'identification des personnes lors de la traversée d'une rue où les personnes se situent à une distance importante par rapport aux caméras, l'extraction des aspects du vieillissement à partir du visage n'est pas toujours possible. Par conséquent, nous avons opté dans ce travail pour une approche de reconnaissance basée sur la démarche.

Nous avons proposé un nouveau descripteur basé sur la distribution spatio-temporelle de la silhouette humaine. En effet, par cette étude, nous avons illustré les indicateurs du vieillissement qui caractérisent la démarche humaine et son pouvoir de discrimination entre les sujets jeunes les sujets âgés.

Etant donné que le comportement du piéton lors de la traversée d'une rue est fortement influencé et corrélé avec l'état du trafic routier ainsi que par le comportement du conducteur, le chapitre suivant portera sur le processus proposé pour la collecte des paramètres du trafic basée sur du traitement de séquences vidéo.

Chapitre

4 Collecte automatique de données 'microscopiques' du trafic routier

4.1 Introduction

Le comportement des piétons lors de traversée d'une rue est fortement corrélé à la fréquence et l'intensité du trafic ainsi qu'avec les aménagements urbains implantés dans la zone de traversée. L'analyse comportementale du piéton lors de la traversée d'une rue n'est fiable que si elle prend en considération l'état du trafic routier. Ainsi, la phase de collecte des variables du trafic est requise pour la désignation d'un tel système. L'automatisation de cette phase se base sur la détection et le suivi des véhicules sur la route. Cette thématique est déjà un domaine de recherche en plein essor, surtout avec les investissements énormes réalisés par l'industrie automobile dans le transport intelligent d'une part et les progrès notables en termes de matériels d'acquisition et de capacité de calcul d'autre part. La vision par ordinateur, leader des Systèmes de Transport Intelligents (ITS), propose déjà des approches avancées, surtout en termes de détection de véhicules [158, 159].

Les avantages des capteurs vidéo par rapport aux autres détecteurs de véhicules sont reconnus depuis quelques années. Les informations riches et détaillées sur les usagers de la route et la scène du trafic que les capteurs vidéo fournissent, dépassent les capteurs traditionnels tels que les détecteurs radar et ultrasonore [160]. Le développement d'un système fonctionnel de surveillance vidéo basé sur le traitement d'images est cependant resté une cible insaisissable depuis les années 1970 [160]. En général, la détection et le suivi des usagers de la route comportent le même problème théorique, quel que soit le type d'utilisateur de la route concerné. Cependant, la détection des véhicules pose moins de difficultés que la détection des piétons. En effet, les mouvements des véhicules sont beaucoup plus réguliers et stéréotypés que ceux des piétons. Les véhicules

sont localement rigides avec une texture régulière et souvent monotone. Malgré ces différences, un schéma de catégorisation, semblable à celui des piétons, peut être adopté pour organiser les différentes étapes de désignation d'un outil automatique de collecte des paramètres du trafic.

Dans ce chapitre, nous présenterons le processus proposé pour la collecte des variables du trafic. Ce processus se base sur la détection et le suivi des véhicules en mouvement sur la route. Ensuite, nous présenterons un rapide état de l'art des principales techniques de détection et de suivi des véhicules dans des séquences d'images. Nous n'exposerons pas toutes les méthodes, mais nous nous concentrerons sur les méthodes les plus couramment utilisées et reconnues comme étant les plus efficaces. Nous clôturerons cette première partie par l'explication et l'évaluation de la méthode proposée. La seconde partie de ce chapitre concerne la phase d'analyse des trajectoires des véhicules et l'extraction des données microscopiques pour caractériser, ultérieurement, le comportement des piétons en fonction des variations du trafic routier.

4.2 Processus proposé : description générale

Dans ce processus nous préconisons davantage l'utilisation de la vision par ordinateur pour effectuer des analyses automatisées, efficientes et objectives des conflits du trafic et de collecte de paramètres caractérisant le comportement des piétons et leur interaction avec le trafic routier. Étant donné le domaine d'application qui est l'ADAS (Advanced Driver Assistance Systems) et la sécurité routière, les véhicules seront recherchés dans un environnement extérieur très variable et congestionné. Les inconvénients des méthodes de vision utilisées, comme le mouvement ou l'apparence, sont la sensibilité aux variations d'arrière-plan et l'exigence d'un temps de calcul élevé [158]. Afin de surmonter ces limitations, nous proposons ici de fusionner des paramètres de l'apparence et du mouvement afin d'améliorer la précision de détection, le temps de calcul, le taux de reconnaissance et l'efficacité dans l'application dynamique de la détection des véhicules sur la route.

Le processus proposé se compose principalement de quatre étapes inter-connectées comme exposé à la figure 4.1.

1. Extraction des paramètres du mouvement ainsi que de l'apparence pour caractériser et suivre un véhicule en mouvement sur la route.
2. Détermination des trajectoires et des positions des véhicules en cordonnées réelles. Pour cet objectif, une étape de transformation des cordonnées image en cordonnées réelles est réalisée.
3. Analyse des trajectoires et collecte des paramètres du trafic tels que la vitesse des véhicules et leur direction.
4. Validation des données par comparaison aux valeurs de réalité du terrain.

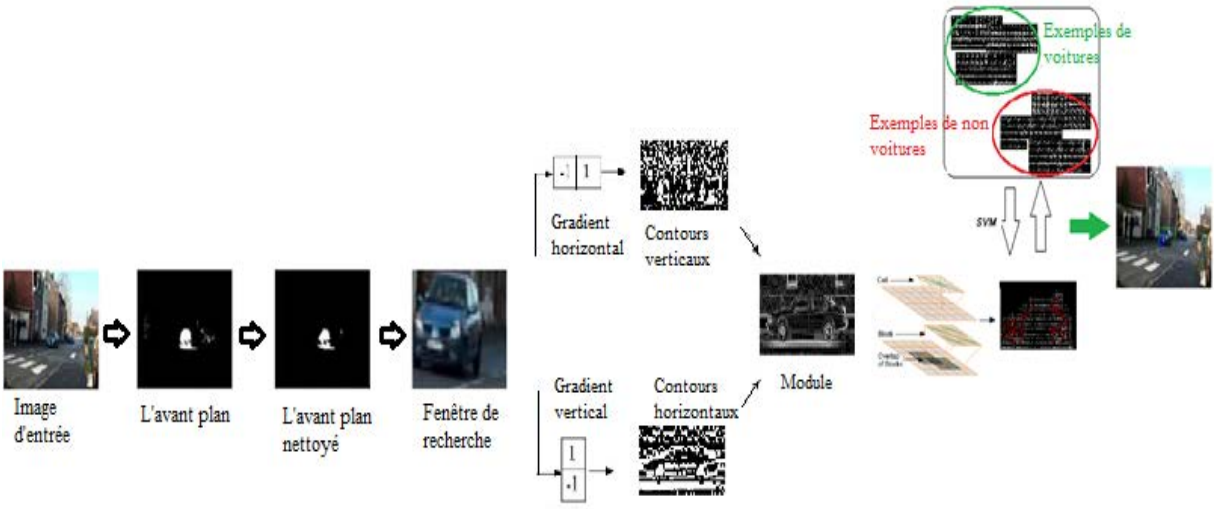


FIGURE 4.1 – Présentation générale du processus proposé pour la détection des véhicules.

4.3 Détection et suivi des véhicules

4.3.1 Approches existantes pour la détection et le suivi des véhicules

4.3.1.1 A partir de radars et de capteurs

Plusieurs approches introduisent la possibilité de détection des véhicules sur la route par des capteurs actifs. Les capteurs actifs peuvent utiliser des signaux millimétriques (capteur Lidar pour 'light detection and ranging' [161, 162]) ou électromagnétiques (capteur radar [158, 163])ou encore acoustiques (capteur ultrasonique [160]). Les capteurs actifs utilisent des signaux physiques pour détecter les véhicules. Par conséquent, ils n'ont pas besoin de lourde phase de calcul. Mais les mesures générées par ces capteurs sont assez bruitées et nécessitent un filtrage et un nettoyage étendus [158] pour arriver aux performances recherchées. Bien que les capteurs actifs produisent des résultats prometteurs dans les prototypes de véhicules, ils sont confrontés à du parasitage tels que l'interférence des capteurs lorsque plusieurs voitures utilisent une technologie similaire [164] . De plus, les capteurs radar [158, 163] et lidar [161, 162] sont coûteux et ne peuvent pas être utilisés dans tous les espaces publics.

4.3.1.2 A partir des caméras

La détection des véhicules à partir des techniques de vision par ordinateur se devise en deux étapes : la génération des hypothèses (l'extraction des caractéristiques) et la vérification des hypothèses (la classification) (Figure 4.2). Pour la phase de génération des hypothèses on distingue deux catégories principales telles que les approches basées sur l'apparence et celles basées sur le mouvement. Cette taxonomie est mise en évidence par plusieurs études sur la détection des véhicules [164, 159].

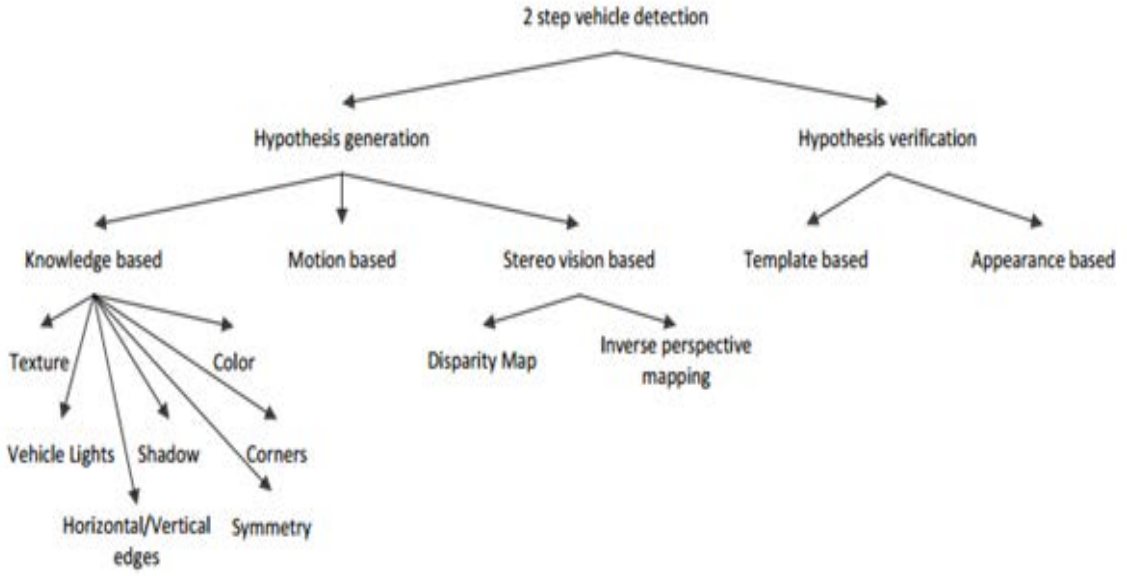


FIGURE 4.2 – Les méthodes de détection des véhicules basées sur la vision par ordinateur [164].

4.3.1.2.1 Approches basées sur le mouvement

Les approches basées sur le mouvement nécessitent un traitement sur toute la séquence d'images pour reconnaître les véhicules. Pour détecter le mouvement, les systèmes de détection des véhicules appliquent des méthodes comme 'la soustraction d'arrière-plan', 'la différence des images consécutives' de la vidéo, 'le flux optique'... La soustraction d'arrière-plan est la méthode la plus classique, adoptée par exemple dans [97]. L'approche proposée dans [162] utilise également une méthode de soustraction de fond, basée sur la médiane, pour détecter les véhicules. Par ailleurs, une analyse paramétrique, comme la taille et la vitesse des objets en mouvement, est effectuée pour discriminer la voiture des autres objets. L'approche proposée dans [165] a appliqué la méthode des différences des images de la vidéo, combinée avec le processus d'enregistrement d'image, pour détecter les véhicules en mouvement. Dans [166], la matrice d'homographie entre les images successives, a été utilisée. Pour une estimation d'homographie fiable et précise le cadre probabiliste basé sur le filtrage de Kalman est proposé, ce qui permet de détecter les véhicules en mouvement dans l'image. Cette méthode semble cependant susceptible de renvoyer de nombreux faux positifs, et l'analyse de la performance quantitative n'a d'ailleurs pas été incluse dans l'étude.

Le mouvement peut aussi être modélisé par le flux optique comme proposé dans [167]. Celui-ci a été utilisé conjointement avec des techniques basées sur l'apparence. L'estimation de l'égo-mouvement utilisant le flux optique et la détection intégrée des véhicules a été mise en œuvre dans [168]. Aussi l'estimation de l'Ego-mouvement, utilisant une caméra omnidirectionnelle pour la détection des véhicules, a été mise en œuvre dans [167].

La plupart de ces méthodes basées sur le mouvement, ont des difficultés à détecter des véhicules ayant un mouvement lent ou à l'arrêt lorsque le trafic est encombré, car

il est difficile de rafraîchir les images d'arrière-plan.

4.3.1.2.2 Approches basées sur l'apparence

Etant donné qu'un véhicule est un objet rigide avec un modèle morphologique constant dans le temps et comporte une certaine symétrie, des caractéristiques comme le contour, l'ombre et la symétrie sont utilisées pour détecter les véhicules dans plusieurs travaux. Par exemple, après avoir évalué les opérateurs de détection des contours sur l'image, la symétrie a été calculée pour reconnaître les côtés verticaux de la face arrière d'un véhicule dans l'étude [169]. Pour détecter les véhicules durant la nuit, la symétrie est aussi utilisée avec la détection de la circularité des phares et l'énergie de bordure dans [170]. Dans cette dernière étude, les caractéristiques de l'ombre et de la symétrie sont intégrées par des capteurs $3D$ et combinées avec deux modèles de système coopérant avec un filtre de modèle multiple : 1) un modèle de système utilisé pour les vitesses constantes, 2) un autre modèle de système utilisé pour les accélérations constantes. Des informations sur la forme et la symétrie du véhicule et de son ombre ont également été utilisées pour définir le modèle géométrique du véhicule. Les informations de contour permettent de mettre en évidence les côtés du véhicule, ainsi que son ombre. En dessous du contour vertical, la symétrie et le feu arrière, sont fusionnés pour la détection du véhicule. Les procédés incluant l'échantillonnage initial, la propagation, l'observation et la fusion des paramètres ainsi que l'évaluation peuvent être complétés en utilisant le filtre de particules ; ce dernier génère avec précision la distribution du véhicule [171]. La symétrie et les bords ont également été utilisés dans [172], à partir d'images monochromes capturées par une seule caméra montée à l'intérieur d'un véhicule et principalement centrée sur les conditions de lumière du jour. Il comprend la détection de la zone de la route pour éviter les fausses détections de véhicules causées par la diffraction des objets de fond. L'utilisation de 'l'ombre' comme un modèle d'apparence des véhicules repérés à l'intérieur de la zone de la route est détecté et suivi en utilisant un filtre de Kalman dans [173].

Cependant d'autres caractéristiques comme l'Histogramme de Gradient Orienté (HOG cf section 4.3.2.2 pour plus de détails) et Haar sont extrêmement bien représentés dans la littérature de détection des objets au sens large et des véhicules en particulier [164, 159, 158]. Comme instance l'utilisation du descripteur Haar dans le contexte de détection des véhicules est introduite dans [174]. Les caractéristiques de Haar ont été extraites pour détecter les faces avant des véhicules suivants un véhicule en déplacement, qui ont été capturés par une caméra tournée vers l'arrière. Les véhicules précédant sont détectés en utilisant une caméra orientée vers l'avant [175, 176, 177, 178] qui permet de capturer les faces arrière des véhicules. Les auteurs ont ensuite réussi à modéliser et reconnaitre les véhicules en se basant sur les caractéristiques de type Haar. Des profils latéraux de véhicules ont également été détectés en utilisant des caractéristiques de type Haar [179], en détectant les roues avant et arrière. Le descripteur Haar a été utilisé pour suivre des véhicules dans le plan image [180]. De plus, dans [181], les caractéristiques de Haar ont été utilisées pour détecter les différentes parties des véhicules. Toutes ces applications montrent que Haar est un descripteur puissant.

Un autre descripteur, le HOG est progressivement devenu un classique pour la détection des objets, avec une phase d'ajustement des paramètres pour chaque classe

d'objets, tel qu'un 3×3 pixels, qui est la taille de la cellule pour la détection de personne, et 8×8 pour la détection des véhicules [89]. D'autres études utilisent également le descripteur HOG pour la détection de véhicules [182, 183]. Dans [183], la symétrie des caractéristiques de HOG extraites à partir d'images données, ainsi que les caractéristiques HOG elles-mêmes, ont été utilisées pour la détection des véhicules. Au-delà de la détection des véhicules, les caractéristiques HOG ont aussi été utilisées pour déterminer la position des véhicules [184]. Toutes ces applications montrent que le HOG est également un descripteur puissant.

La fusion des descripteurs d'apparence a motivé plusieurs travaux récents dans la détection des véhicules. En effet, la combinaison de HOG et Haar est utilisée dans [168] pour la reconnaissance de véhicules. Par ailleurs, le HOG est également fusionné avec la texture, représentée par le filtre Gabor, dans [166, 185].

Ainsi, HOG et Haar sont des détecteurs d'objets très utilisés. Cependant, l'inconvénient majeur des méthodes d'apparence en générale et de ces deux descripteurs (HOG et Haar) en particulier, est le temps de calcul élevé [164, 158, 159]. Ils sont difficilement utilisables dans un contexte temps réel ou des applications dynamiques, telles que la détection des véhicules sur la route dans un objectif de sécurité routière, où le temps de calcul est un paramètre très important. La rapidité de la réponse conditionne son utilisation dans un tel contexte. Pour cela, une approche proposée dans [HOG + adaboost] [ref] exploite le principe du 'Boosting' pour diminuer la dimensionnalité de HOG et gagner en temps de calcul. Les auteurs génèrent un vecteur binaire (des 0 et 1) à partir de chaque cellule de HOG par la multiplication de l'histogramme $'h_i'$ de chaque 'bin' par son poids généré par l'algorithme adaboost. Ensuite, ces vecteurs (représentant chaque cellule) sont regroupés par blocs en suivant le même principe de construction que le descripteur HOG. Finalement, le SVM est utilisé comme classifieur pour discriminer le véhicule.

Ces dernières années, la disponibilité de processeur graphique (GPU) a permis la mise en œuvre et le parallélisme en temps réel des algorithmes de vision, coûteux sur le plan informatique, pour la détection et le suivi des véhicules [159]. Dans [186], un GPU précoce a été utilisé pour la détection des véhicules monoculaires. Dans [187], le GPU a été utilisé pour implémenter la détection de véhicules en temps réel à l'aide du descripteur HOG. La combinaison GPU et architectures FPGA fournit un détecteur raisonnablement rapide, mais il reste une solution très coûteuse qui n'est pas à la portée de toute les applications.

4.3.2 Approche proposée : Optimisation de l'Histogramme de Gradient Orienté

Comme le montre l'état de l'art de la partie précédente, l'HOG est un descripteur puissant, utilisable pour la détection des objets au sens large et celle des véhicules en particulier [158, 159, 164]. Son insensibilité à la variation de pause ou à la luminosité et sa grande capacité à modéliser la forme et l'apparence de l'objet représentent ses points forts. Cependant, l'inconvénient majeur de l'HOG est le temps de calcul et la lourdeur [158, 159, 164]. Ainsi, dans ce travail, nous allons contribuer à optimiser son utilisation classique et à gagner en termes de temps de calcul. Pour cela, notre approche consiste

à fusionner les paramètres de mouvement ainsi que ceux de l'apparence.

4.3.2.1 Les paramètres du mouvement

Sur la route et dans un contexte de traversée d'une rue, le véhicule en mouvement représente le danger majeur qui menace la sécurité des piétons. En fait, le mouvement des véhicules est le paramètre clef pour déterminer et gérer les gaps sécuritaires d'une part et prédire le moment pour traverser en toute sécurité d'autre part. Par conséquent, on suppose qu'un véhicule garé ne représente plus un danger direct. Par la suite, on ne s'intéressera qu'à la détection des voitures en mouvement sur la route. Se focaliser sur le mouvement permet de réduire la zone d'intérêt.

Le mouvement est capturé en partant d'un algorithme d'extraction d'arrière-plan. L'environnement extérieur change fréquemment à cause : des variations de conditions d'éclairage (différents moments de la journée), de la présence simultanée de plusieurs objets en mouvement (arbre, piétons, véhicules, panneaux publicitaires...), de la présence d'ombres... Tous ces éléments rendent la modélisation de l'image d'arrière-plan très critique. Dans cette optique, un modèle d'arrière-plan adaptatif est utilisé pour générer l'image d'arrière-plan dans [188]. Tout d'abord, pour apprendre le modèle d'arrière-plan initial, on utilise les premières "n" images (fixées empiriquement aux 15 premiers images [188]) de chaque séquence vidéo. Ce modèle, précédemment calculé, sera actualisé itérativement. Un pixel est considéré comme une partie de l'arrière-plan s'il est à une distance d (la valeur $d = 20$ est adoptée en se basant sur des observations expérimentales [188]) de l'une de ses couleurs de fond dans l'espace RVB. Pour l'adapter aux changements fréquents d'éclairage et de conditions météorologiques dans les séquences vidéo, les pixels de premier plan extraits sont ensuite soumis à un test de valeur de différence de couleur par rapport à la luminosité, D qui est définie comme présentée dans l'équation 4.1.

$$D = 18 \times D_{gb} + |log(\frac{V_{BG}}{V_{curr}})| \tag{4.1}$$

Où D_{gb} est la distance euclidienne entre le pixel courant et le pixel de fond dans l'espace gb normalisé ; V_{BG} est la luminosité de fond ; et V_{curr} est la luminosité actuelle du pixel.
La figure 4.3 montre l'ensemble des étapes de détection du premier plan et un exemple de détection d'une voiture.

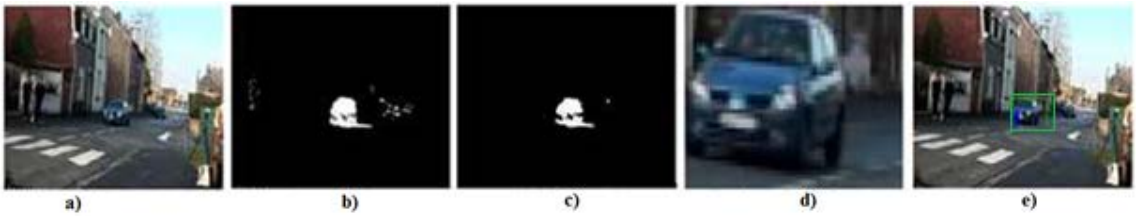


FIGURE 4.3 – Illustration d'extraction des objets en mouvement : a) Image initiale, b) Image de l'avant plan initialement obtenue, c) Image de l'avant plan post-traitée, d) Résultat d'extraction d'objets en mouvement et e) Détection du véhicule.

Pour un pixel de premier plan particulier, si $D < 0,5$ [188], ce pixel est marqué comme un pixel de changement d'éclairage (ombre/ sur-lignage) et il est rejeté. Le premier plan résultant est ensuite post-traité pour éliminer le bruit (Figures 4.3 .b et c).

4.3.2.2 Les paramètres d'apparence

a) Principe

L'idée importante derrière le descripteur HOG est que l'apparence et la forme locale d'un objet dans une image peuvent être décrites par la 'distribution' de l'intensité du gradient suivant les directions du contour. Ceci peut être fait en divisant l'image en régions adjacentes de petites tailles, appelées cellules, et en calculant pour chaque cellule l'histogramme des directions du gradient ou les orientations des contours pour les pixels à l'intérieur de cette cellule. La combinaison des histogrammes forme alors le descripteur HOG. Pour de meilleurs résultats, les histogrammes locaux sont normalisés en contraste, en calculant une mesure de l'intensité sur des zones plus larges que les cellules, appelées des blocs, et en utilisant cette valeur pour normaliser toutes les cellules du bloc. Cette normalisation permet une meilleure résistance aux changements d'illuminations et aux ombres.

b) Les étapes de réalisation

1. **Normalisation gamma des couleurs**
   Une correction gamma est appliquée sur l'image originele pour que la distribution des couleurs soit normalisée et dépasser la variation de luminosité et de contraste.
2. **Calcul de gradient**
   Pour le calcul de gradient on applique un 'filtre dérivatif $1-D$ centré', dans les directions horizontale et verticale comme illustré à la figure 4.4.

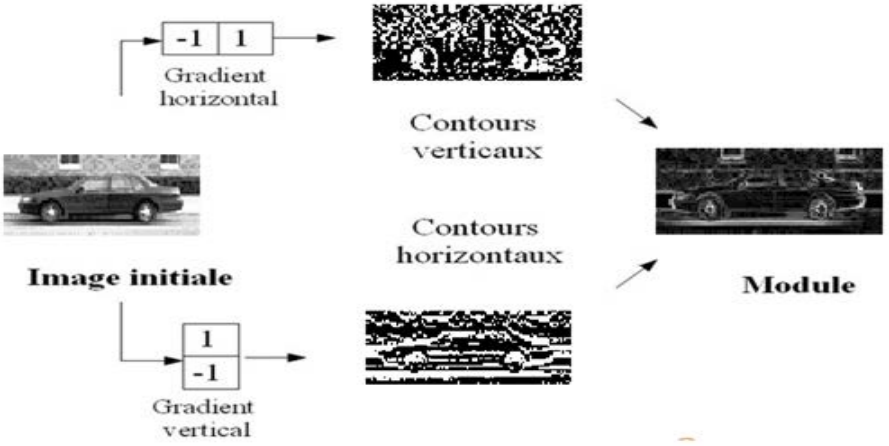


FIGURE 4.4 – Calcul des gradients horizontal et vertical d'un véhicule.

3. **Calcul d'orientation et vote par cellule**

La troisième étape est la création des histogrammes de l'orientation des gradients. Ceci est fait dans des cellules carrées de petite taille (8×8 *pixels*). Chaque pixel de la cellule vote alors pour une classe de l'histogramme, en fonction de l'orientation du gradient à ce point. Le vote du pixel est pondéré par l'intensité du gradient en ce point. Les histogrammes sont uniformes de 0 à 180^0 (cas non signé) ou de 0 à 360^0 (cas signé).

Dans le cas de la détection de véhicules, on utilise le gradient signé ($[0, 2 \times \Pi]$). L'orientation des gradients se calcule comme ci-après (Equation 4.2).

$$g(id) = \begin{cases} atan(\frac{F_y(id)}{F_x(id)}), & \text{if } F_x \geq 0 \text{ et } F_y \geq 0 \\ atan(\frac{F_y(id)}{F_x(id)}) + \sqcap, & \text{if } F_x < 0 \\ atan(\frac{F_y(id)}{F_x(id)}) + 2 \times \sqcap, & \text{if } F_x \geq 0 \text{ et } F_y < 0 \end{cases} \tag{4.2}$$

Avec F_y et F_x sont respectivement les gradients horizontal et vertical.

4. **Construction des blocs et normalisation**

   Pour la phase de construction des blocs, on regroupe plusieurs cellules dans un bloc. Les blocs sont chevauchants (se recouvrent sur un seul pixel), donc une même cellule participe plusieurs fois au descripteur final, comme membre de blocs différents.

   Les expérimentations faites par Dalal et Triggs, pour la détection de véhicules [89], ont montré que la meilleure performance est obtenue pour des blocs rectangulaires contenant 2×2 cellules de 8×8 pixels chacune.

   Une étape de normalisation des descripteurs, afin d'éviter les erreurs dues aux variations d'illumination est ensuite effectuée. Plusieurs types de normalisation sont proposés. La meilleure normalisation adaptée au contexte de détection des véhicules est la $'L1 - racine'$ [89] (Equation 4.3).

$$L1 - racine : f = \sqrt{\frac{\vartheta}{(\|\vartheta\|_1 + \varepsilon)}} \tag{4.3}$$

   Avec ϑ, le vecteur non normalisé contenant tous les histogrammes d'un seul bloc, $\|\vartheta\|_1$ est sa k-norme et ε est une constante de faible valeur.

4.3.3 Evaluation des performances

Par cette phase nous souhaitons analyser et illustrer les performances de l'approche proposée à deux niveaux :

1. Une évaluation des performances des paramètres d'apparence (HOG) en termes de représentativité des caractéristiques, de spécificité des véhicules, et en termes de taux de détection correcte(cas de mono et multi-détection).

2. Une étude expérimentale pour évaluer les performances de notre approche basée sur le HOG optimisé par la fusion de paramètres de mouvement et d'apparence dans un cas de détection dynamique.

4.3.3.1 Cas de détection statique

Dans le cas de détections statiques, nous cherchons à illustrer les performances des paramètres de l'apparence. En premier lieu, on met en relief la bonne représentativité de l'HOG et de ses caractéristiques les plus représentatives d'un véhicule ainsi que sa sensibilité à l'espace de couleur utilisé. En deuxième lieu, on évalue les potentiels de l'HOG en termes de détection d'un ou de plusieurs véhicules. Par conséquent, deux bases de données sont utilisées à ce niveau : une en niveau de gris (Figure 4.5) et l'autre en vrais couleurs (RVB) (Figure 4.6).

— UIUC [189] base de données en niveau de gris. Cette base de données contient des images des voitures en vues latérales. Les images ont été collectés à UIUC. Cette base contient 1050 images d'entraînement (550 voitures et 500 images sans voiture), 170 images de test à une seule échelle, contenant 200 voitures à peu près à la même échelle que dans les images d'entraînement et 108 images de test multi-échelle, contenant 139 voitures à différentes échelles. Des échantillons de la base UIUC sont présentés à la figure 4.5.



FIGURE 4.5 – Échantillons de la base de données UIUC : a) Exemple positif et b) Exemple négatif.

— TUDarmstadt [190] (sideviews-cars) base de données en vrais couleurs. Cette base contient des images de voitures ainsi que de non voitures capturés à l'extérieur. Un extrait de cette base est présenté par la figure 4.6



FIGURE 4.6 – Échantillons de la base de données TUD : a) Exemple positif et b) Exemple négatif.

4.3.3.1.1 Analyse et visualisation de la représentation des caractéristiques d'une voiture par l'HOG

La puissance du descripteur HOG se trouve surtout dans sa capacité à reproduire correctement l'apparence de l'objet cible. En fait, dans cette partie nous visualiserons comment le HOG voit l'objet et l'effet de l'espace de couleur ainsi que la résolution de l'image sur les performances. Pour cette objectif, nous nous basons sur l'algorithme '*HOGgles*' proposé par [191] qui présente un 'HOG dictionnaire' pour coder et, visualiser les ensembles des gradients orientés par classe (bin) sous forme d'images en niveaux de gris. L'intensité des couleurs et ainsi représenté par, le nombre de pixels de contour orientés dans chaque direction.

Nous présenterons en premier lieu des exemples de l'HOG de véhicules en niveaux de gris (Figure 4.7.a) et en RVB (Figure 4.7.b). Dans un second temps nous présenterons quelques exemples de HOG pour des objets autres que l'objet cible (Figure 4.8).

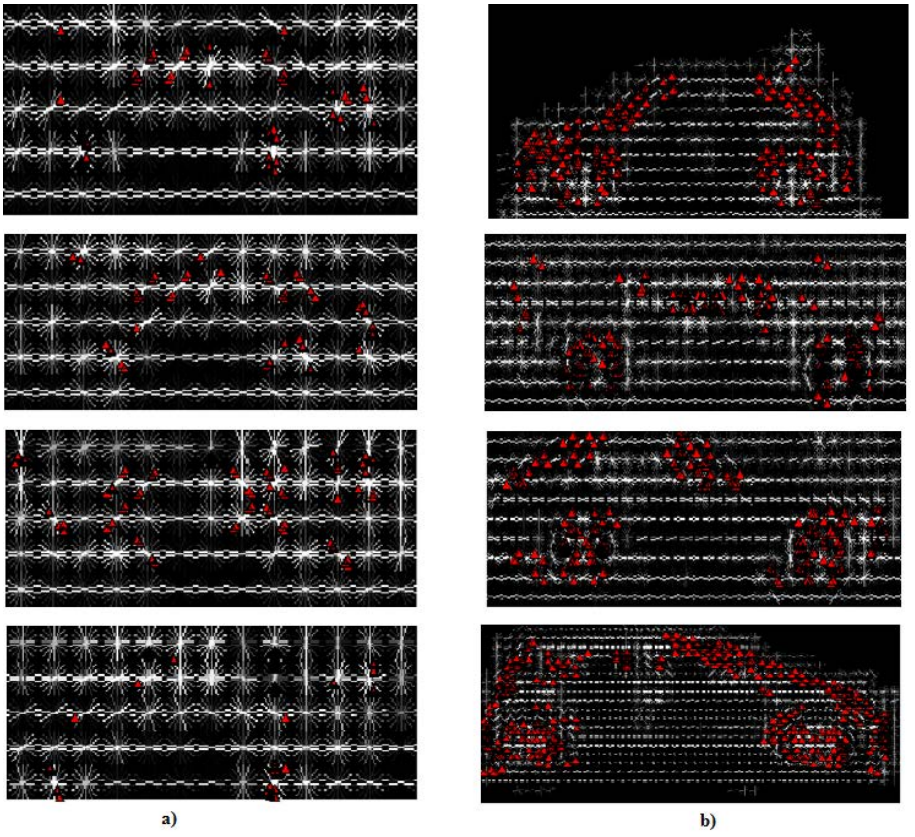


FIGURE 4.7 – Visualisation de voitures par l'HOG : a) Exemples des images aux niveaux de gris et b) Exemples des images en RVB.

On remarque que plus le nombre de pixels de contour orientés dans une direction (parmi les 18 bin) est important, plus la couleur de représentation est intense. De ce fait, les orientations de gradients les plus importantes sont représentées par la couleur blanche, alors que les faibles orientations sont de faibles intensités, voire même invisibles.

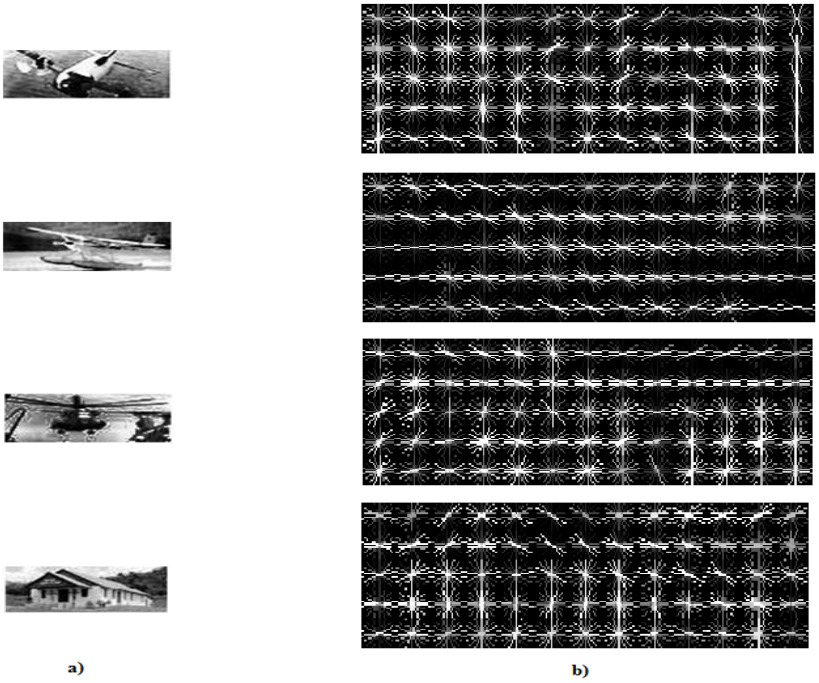


FIGURE 4.8 – Visualisation d'objets autres qu'une voiture par l'HOG : a) Images initiales et b) HOG.

En se basant sur ces représentations on constate également que les gradients les plus intenses reproduisent la forme d'un véhicule. On constate aussi que la forme du véhicule est concave et symétrique suivant l'axe du milieu vertical de la matrice, comme cela est illustré à la figure 4.9.

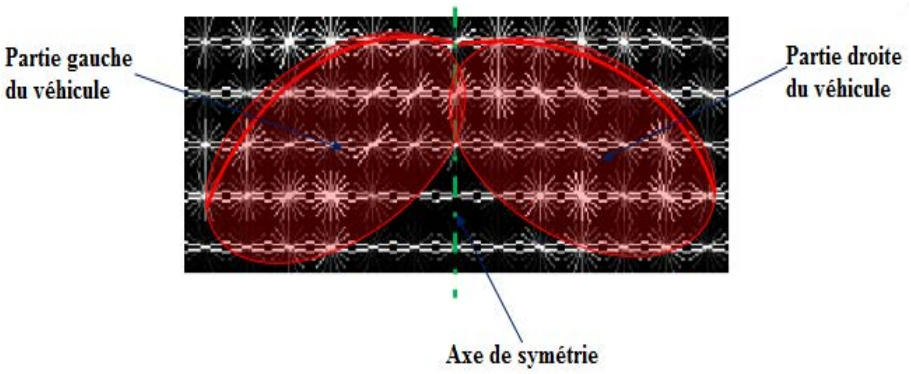


FIGURE 4.9 – Illustration de la similarité de l'HOG d'une voiture.

4.3.3.1.2 Performances en termes de reconnaissance des véhicules

Pour tester les performances du processus proposé pour la détection de véhicules, nous présenterons une première étude expérimentale, menée sur deux bases de données, avec des conditions différentes. Le premier consiste à évaluer les paramètres d'apparences retenus, en termes de mono-détection. Le deuxième test présente leurs

performances en termes de détection multiples. Ces tests sont menés sur les deux bases (UIUC et TUD) déjà présentées dans la section 4.3.3.1. Les résultats des mono et multi-détections sont présentés respectivement par le tableau 4.1, la figure 4.11 et le tableau 4.2 , la figure 4.10.

TABLE 4.1 – Taux de reconnaissance d'un véhicule unique

	UIUC		TUD	
	Apprentissage	Test	Apprentissage	test
	630 images	170 images	60 images	77 images
Taille	[40,100]	Variable	[80,200]	Variable
Taux	88,24%		100%	

TABLE 4.2 – Taux de reconnaissance multi-véhicules

	UIUC		TUD	
	Apprentissage	Test	Apprentissage	test
	630 images	30 images	60 images	5 images
Taille	[40,100]	Variable	[80,200]	Variable
Taux	100%		100%	

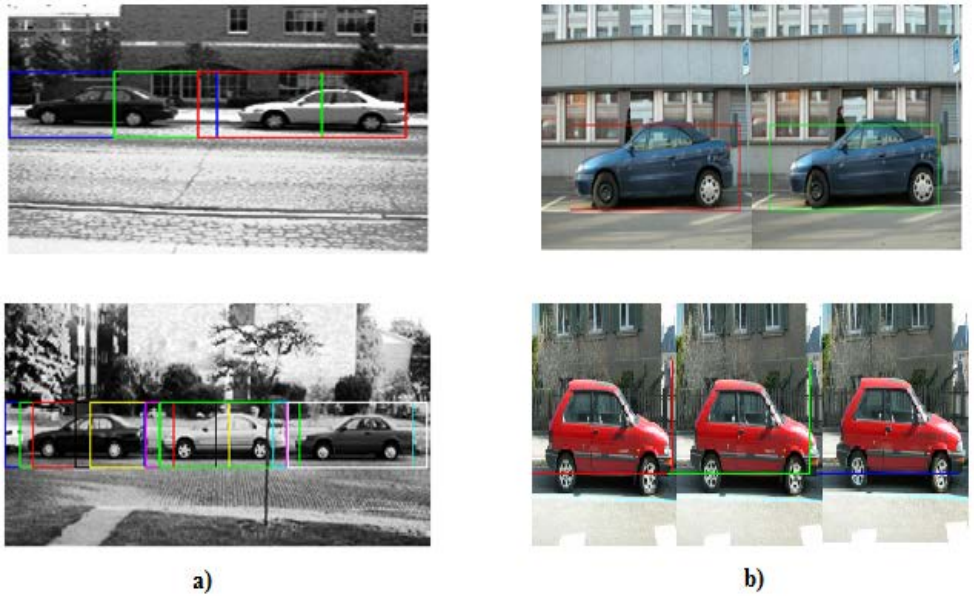


FIGURE 4.10 – Illustration de la détection de plusieurs véhicules dans une scène statique.

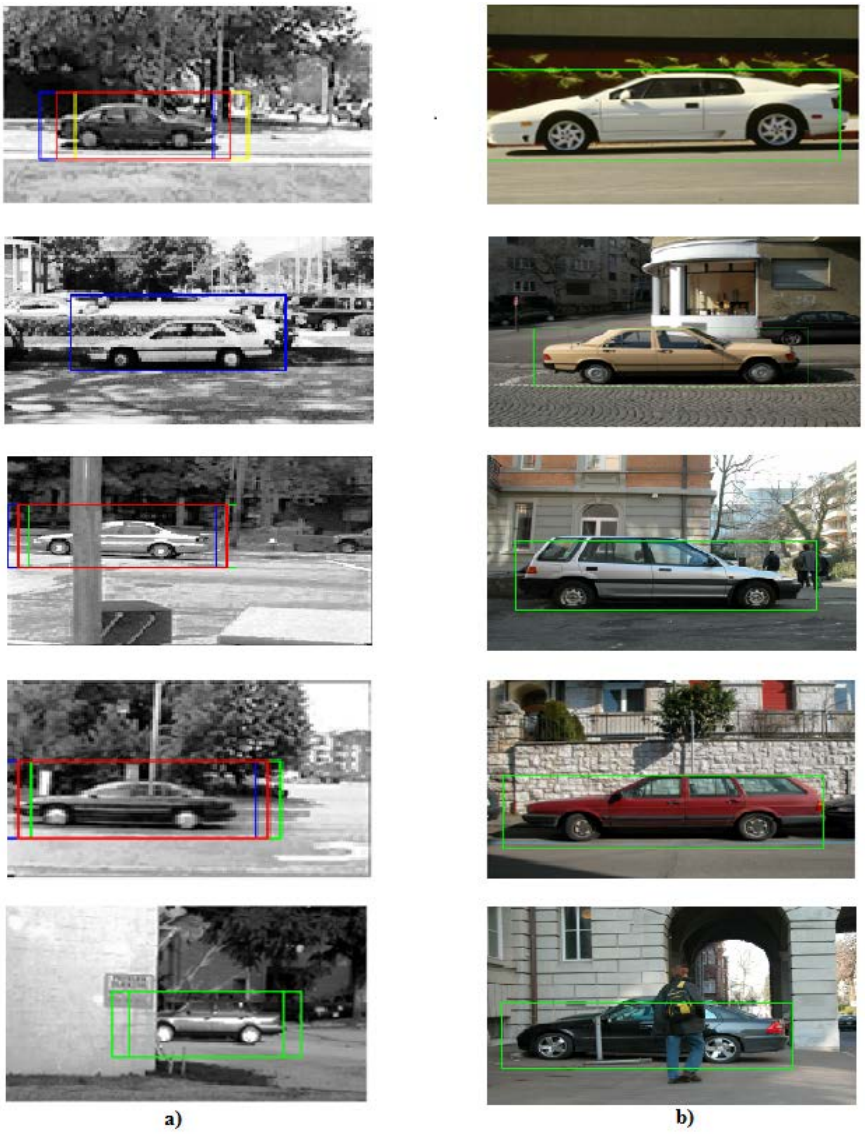


FIGURE 4.11 – Illustration de la détection d'un seul véhicule dans une scène statique.

La première constatation importante est une sur-détection. En fait, quelques fenêtres contiennent deux parties de deux véhicules différents qui sont considérées comme un seul véhicule (cas du rectangle vert de la figure 4.10). En effet, le rectangle bleu est le véhicule 1, le rouge est le véhicule 2 alors que vert est une fausse détection. Ce problème ne se retrouve qu'avec la base UIUC, qui est une base en niveau de gris.

Les fausses détections sont expliquées [191] par la précision des caractéristiques choisies : comment le descripteur voit l'objet recherché. Ces conclusions sont déjà illustrées par la visualisation de la matrice d'HOG de l'objet véhicule dans deux espaces de couleurs déférents : Niveaux de gris et RGB (Figure 4.7).

4.3.3.2 Cas de détection dynamique : performances en termes d'optimisation du temps de calcul

Dans cette section, nous étudierons les performances du processus proposé : 'HOG optimisé', pour une application dynamique. Les performances seront illustrées en termes de taux de reconnaissance et de temps de calcul.

4.3.3.2.1 Base dynamique utilisée

Nous avons enrichi la base de données CWV (CrossWalkValenciennes) [111] par des séquences capturées dans l'espace urbain dans le contexte de la traversée d'une rue (*Annexes, scénarios de test 2*). En fait, ces séquences se composent de vidéos de voitures et de piétons avec des situations importantes d'occlusion, des directions, des variations d'éclairage et de poses. La base de données utilisée incorpore également des séquences vidéo de mono et multi véhicules sur route avec différentes situations de croisement entre ces véhicules, comme illustré par la figure 4.12.



FIGURE 4.12 – Echantillons de la base dynamique.

4.3.3.2.2 Résultats

Le taux de reconnaissance est mesuré à l'aide du SVM binaire (1 = voiture, 0 sinon). Cependant, le temps de calcul est mesuré par le nombre de secondes nécessaires pour détecter toutes les voitures présentes sur une image.

Le système proposé de l'HOG optimisé, est évalué en mono et en multi-détection de voitures comme illustré par la figure 4.13. a et .b, respectivement. La variation de l'état d'éclairage est également prise en compte dans cette étude expérimentale. En fait, les figures 4.13. a et .b sont filmés à différents moments de la journée. La performance



FIGURE 4.13 – Illustrations de la détection de véhicules : cas dynamique. a) mono-détection, b) multi-détection.

du processus proposé est aussi, évaluée en fonction de la variation de direction de la voiture (trafic gauche et/ou droit).

Les résultats expérimentaux du taux de reconnaissance et du temps de calcul sont résumés dans le tableau 4.3.

En se référant au tableau 4.3, nous pouvons mettre en évidence la capacité du processus proposé en termes de détection mono et multi-voitures en situation dynamique, quelle que soit la distance par rapport à la caméra, la direction des voitures et la variation de l'état d'éclairage. Nous atteignons $84,62\%$ de taux de reconnaissance. Le tableau 4.3 résume l'étude comparative entre l'HOG calculé à partir de l'image originale et uniquement l'objet en mouvement. En fait, le temps de calcul diminue nettement de 1 heure approximativement consacrée à explorer des vidéos de 130 images en utilisant la première méthode à quelques secondes $(43,65s)$ en utilisant notre proposition. De

TABLE 4.3 – Evaluation des performances de l'approche proposée en termes de taux et de temps de détection : cas dynamique.

	HOG par blocs [90]	**Processus proposé : HOG optimisé**
Séquence 1		
Nombre d'images	130	
Espace de couleur	RVB (Vrais couleurs)	
Taille des fenêtres de recherche (pixel)	1440x1080	10x35 (taille maximale d'objets d'avant plan)
Temps de calcul (seconde)	3600	**43,65**
Taux de reconnaissance	78,46%	**84,62%**
Séquence 2		
Nombre d'images	85	
Espace de couleur	RGB (vrais couleurs)	
Taille des fenêtres de recherche (pixel)	1440x1080	31x94 (taille maximale d'objets d'avant plan)
Temps de calcul (seconde)	2592	**31,32**
Taux de reconnaissance	68,75%	**83,75%**

même, le taux de reconnaissance est important dans les deux cas avec une amélioration conséquente en utilisant le modèle proposé (+6, 16% et +15% respectivement pour les séquences 1 et 2).

4.3.3.2.3 Étude comparative

Une étude comparative avec les travaux de la littéraire est également proposée pour mettre en évidence la robustesse de notre processus, surtout en termes de taux de reconnaissance et de temps de calcul.
Nous présenterons ici une étude comparative avec 2 autres travaux [192, 187] ayant comme objectif principal la réduction du temps de calcul pour des applications dynamiques de détection de voitures telle que présentée dans le tableau 4.4.

Le premier travail présenté par [192] introduit un détecteur de voitures en temps réel basé sur le détecteur WaldBoost et le 'tracker' TLD et inclut en totalité 7 processus (WaldBoost *, Warping + RF (Randomized Forest) Learning, LrD - Détect), les échantillons négatifs LrD, FoT (Flock of Trackers), la vérification WaldBoost, la

TABLE 4.4 – Etude comparative des performances de notre approche en termes de taux et de temps de détection.

	Approche proposée	HOG+GPU [187]	[192]
Résolution des images	1440x1080	320x240	1024x768
Architecture	CPU	GPU-CPU	CPU
Temps de calcul (s/image)	0,3	0,1	0,7
Taux de reconnaissance	84,62%	85%	80%

vérification LrD. Ce système est complexe et nécessite une phase d'apprentissage exhaustive pour chaque séquence. Cinq images de pré-acquisition sont nécessaires, pour que l'algorithme se stabilise avant de commencer la mesure des performances réelles [192]. De plus, la détection d'un objet unique par ce système nécessite $0,7s$ pour une image de résolution 1024×768 en utilisant le CPU.

Le second travail [187] exploite le GPU pour paralléliser les étapes de traitement de l'HOG afin de réduire le temps de calcul. Une étude expérimentale sur une image de résolution de 320×240 est présentée. Ce travail atteint un temps de calcul égal à $0,1\ s/image$ mais il utilise une architecture sophistiquée comme le GPU et le pipeline. Aussi, la résolution de l'image utilisée est très petite comparée à celle adoptée. Cependant, l'approche proposée ne fait que $0,3\ s/image$ et atteint un taux de reconnaissance très satisfaisant de $84,62\%$ en utilisant uniquement un noyau de CPU i3.

Cette bonne amélioration du temps et du taux de reconnaissance est justifiée par l'optimisation du processus de calcul des caractéristiques de l'HOG qui sera directement focalisé sur l'objet cible (blob de premier plan). Cette méthode réduit le temps de calcul passé notamment pour l'exploration de toute la scène et donne un détecteur de voiture rapide et performant.

4.4 Modélisation de la zone de traversée (passage piéton) et détection des positions réelles des véhicules

Les conditions préalables à l'analyse précise des positions des usagers de la route ainsi que des véhicules sont la récupération des coordonnées réelles des véhicules sur la route telles qu'elles apparaissent dans les séquences vidéo. Ce besoin pratique soulève le problème de transformer les points du plan de l'image en cordonnées réelles comme cela a été fait pour le piéton (cf section 2.4).

4.4.1 Modèle de passage piéton

L'analyse comportementale des piétons lors de la traversée d'une rue à partir de traitement de vidéo nécessite la collecte des variables réelles telles que la vitesse de déplacement des véhicules en $km.h^{-1}$. Par conséquent, on cherche à calculer les positions des véhicules en coordonnées réelles et non en pixels. Ces coordonnées sont directement interprétées et ayant une signification directe dans ce contexte. En effet, la calibration de la caméra est une étape primordiale pour assurer le passage des cordonnées images en cordonnées réelles (cartésiennes).

Pour calibrer les caméras et déterminer les cordonnées des positions des véhicules dans le repère cartésien nous avons utilisé la méthode classique à base d'homographie. Cette méthode se base sur le fait d'établir une correspondance entre les points de l'image et celui d'un modèle de la zone d'intérêt construit avec les coordonnées réelles de la scène. Dans le contexte de la traversée d'une rue nous avons choisi le passage piéton comme un modèle de mappage, puisqu'il s'agit de la zone d'intérêt. Ce modèle est celui présenté dans le chapitre 2 (Figure 2.10.) et permet de visualiser une vue d'ensemble de la scène en englobant les champs de vision des 3 caméras utilisées pour filmer le piéton et le trafic gauche et droit.

4.4.2 Homographie

Le principe de transformation des coordonnées image en coordonnées cartésiennes est présenté dans le chapitre 2. C'est le même principe qui est utilisé dans la conception de l'outil de collecte des paramètres du trafic dans ce chapitre. Aussi, seul un rappel du principe général de mappage de plan image vers le plan réel, à la base de l'homographie, est présenté (Equation 4.4). Le lecteur est invité à se référer au chapitre 2 pour plus de détails.

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix} \times \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} \tag{4.4}$$

Avec 'H' la matrice d'homographie et (X_i, Y_i) et (x_i, y_i) sont respectivement les coordonnées sur le plan image et leurs correspondants sur le plan réel.

Une illustration des performances de l'homographie dans le contexte du tracé de trajectoires de véhicules est présentée à la figure 4.14, pour le trafic gauche et droit.

4.5 Données collectées pour la caractérisation du trafic

Pour étudier la collision et le risque d'accident entre les véhicules et les piétons, la littérature d'accidentologie utilise généralement les variables suivantes :

— TTC (Temps de collision : Le temps qu'il reste jusqu'à ce qu'une collision entre deux véhicules, ou un véhicule et un piéton, se produise si les trajectoires et la vitesse de déplacement ainsique la différence de vitesse sont maintenues [64]),

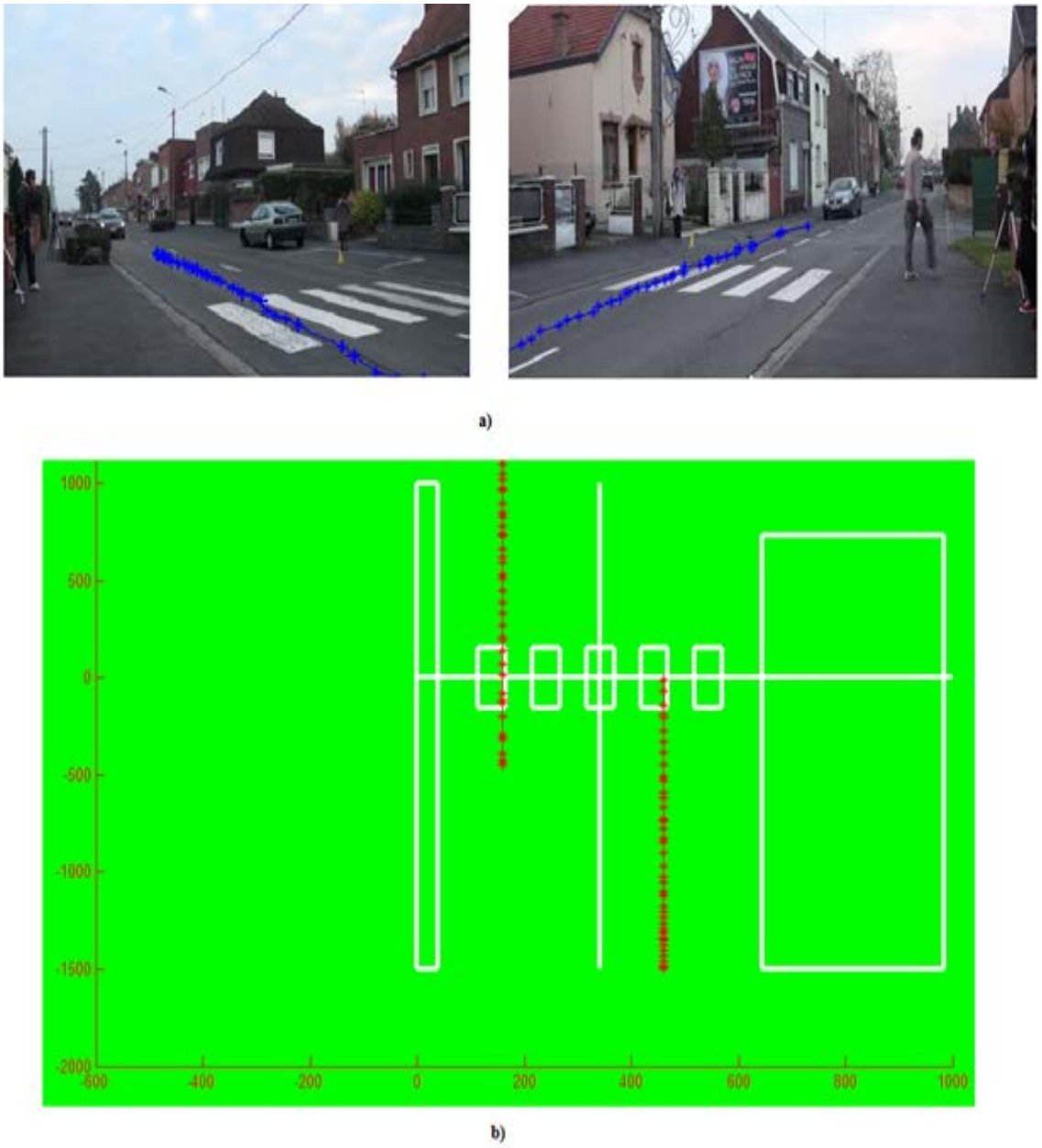


FIGURE 4.14 – Exemple de transformation de trajectoires des coordonnées images en coordonnées réelles pour le trafic gauche et droit : a) Trajectoires sur le plan image, b) Trajectoires projetées sur le modèle de passage piéton.

— *PET* (Le temps de post-empiètement : L'intervalle de temps entre la sortie du dernier piéton de la zone de conflit et l'entrée d'un nouveau piéton dans cette zone [193]),
— *DST* (Décélération du temps de sécurité : La décélération nécessaire pour atteindre une valeur de PET non négative si les mouvements des usagers de la route en conflit potentiels restent inchangés [194])
— *GT* (Temps de gap : La variation du PET calculée à chaque instant [68]).

Tous ces paramètres d'analyse des conflits se basent, implicitement ou explicitement,

sur la vitesse des véhicules, leurs directions ainsi que la position spatiale des véhicules et des piétons. Par exemple, la DST se calcul en fonction de la vitesse des véhicules, du temps restant pour que le piéton atteigne le milieu de la zone de traversée et le temps de gap. Le TTC est, quant à lui, déterminé en fonction de la différence de vitesses des véhicules et des piétons et cette différence n'est significative que si le piéton et le véhicule se rapprochent l'un de l'autre. D'où la notion importante de direction des déplacements.

En se basant sur ces études, nous avons choisi de calculer les paramètres utilisés dans l'analyse des variables les plus utilisés dans les conflits P-V. Ces paramètres sont l'estimation des positions spatio-temporelles des véhicules par rapport aux piétons (dont les paramètres et les calculs sont présentés dans le chapitre 2) ainsi que leurs trajectoires, leurs directions et leurs vitesses moyennes de déplacement.

4.5.1 Estimation des trajectoires et des positions des véhicules

Les processus de détection et de calibration des caméras sont nécessaires pour l'estimation automatique des trajectoires et des positions des véhicules. De même le principe de filtrage des trajectoires (présenté par l'algorithme 2.2) est également appliqué ici pour éviter toute position aberrante et de fausse détection.

4.5.2 Estimation des vitesses moyennes

Comme pour les piétons, la vitesse moyenne des véhicules est un paramètre très important pour la caractérisation du trafic routier. La méthode utilisée est la même que pour les calculs de vitesse des piétons (*cf partie 2.5.3*, équation 4.5).

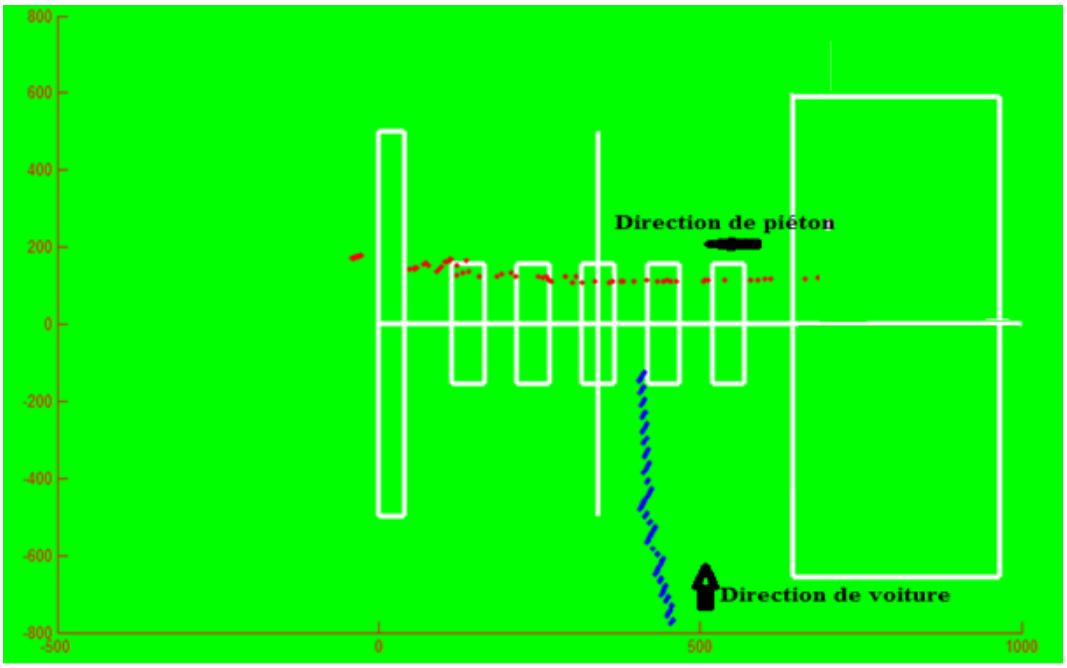
$$V = \frac{D}{T} \tag{4.5}$$

Comme pour le piéton, les deux méthodes M1 (différence entre le premier et le dernier point de la trajectoire) et M2 (somme des différences entre deux points successifs) sont utilisés.

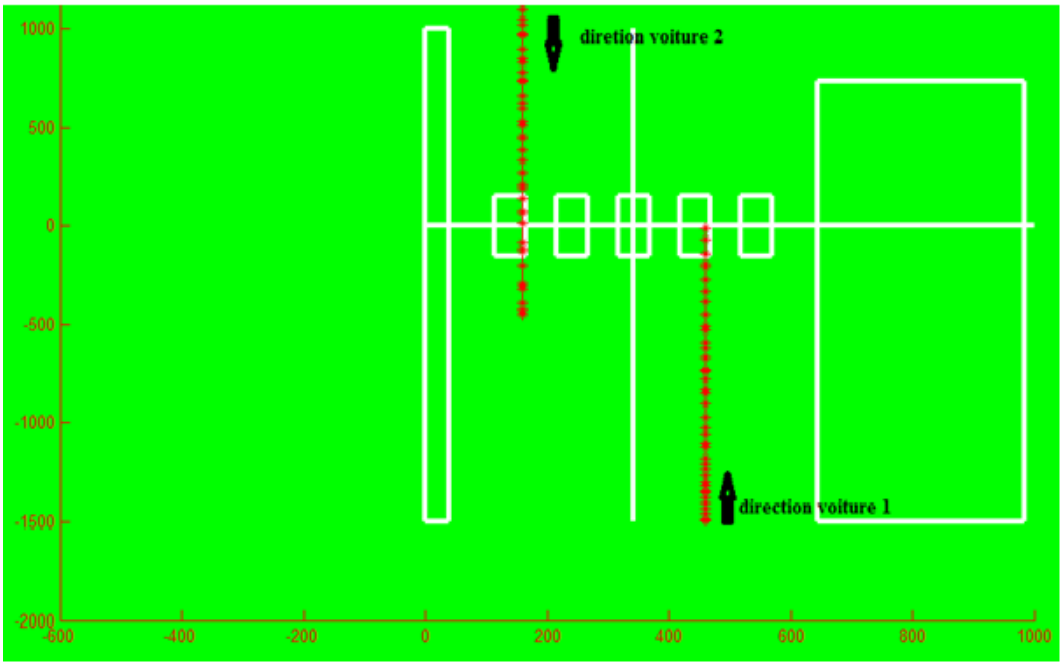
4.5.3 Estimation des directions

Par direction des véhicules on entend déterminer son sens de déplacement et le désigner par une flèche orientée sur le modèle de passage piéton, pris comme la référence de la scène de la zone de traversée. Le modèle utilisé est déjà assigné d'un repère cartésien orienté ayant l'intersection de la médiane du passage piéton comme origine (Figure 4.15). Par conséquent, la projection de la trajectoire du véhicule sur ce modèle permet d'avoir des positions annotées spatialement par des coordonnées ordonnées dans un ordre croissant ou décroissant. L'analyse de l'ordre de ces positions permet d'identifier la direction du véhicule ou du piéton. Selon le modèle de passage piéton proposé, si les positions sont dans l'ordre croissant, alors le véhicule se déplace de gauche à droite, si elles sont dans l'ordre décroissant, alors la direction du véhicule est de droite à gauche. L'analyse des trajectoires des véhicules collectés à partir des caméras 1 et 2

pour le trafic gauche et droit, est illustrée par la figure 4.15.b. La figure 4.15.a présente l'analyse des trajectoires des piétons et celles des véhicules avec la détermination de leurs directions.



a)



b)

FIGURE 4.15 – Illustration de l'identification des directions : a) Exemple de directions d'un piéton et d'un véhicule, b) Exemple de direction de deux véhicules pour le trafic gauche et droit.

4.5.4 Evaluation de la précision des estimations

4.5.4.1 Etude de cas et base d'évaluation

Nous avons utilisé la base de données CWV (CrossWalkValenciennes), présentée à la section 2.3.4.1, pour l'évaluation du processus de détection des véhicules dans les cas dynamiques. Les séquences CWV sont capturées dans un espace urbain extérieur, pour un contexte de traversée d'une rue. La base de données se compose des séquences vidéo pour le trafic gauche et droit. Pour présenter les performances de notre outil automatique de collecte de paramètres du trafic en termes de précision de la vitesse moyenne des véhicules, la base CWV est enrichie par 3 scénarios supplémentaires. Durant ces scénarios les conducteurs sont invités à rouler avec 3 vitesses différentes toute en respectant la vitesse limite autorisée dans la zone d'acquisition qui est de 50 $km.h^{-1}$.

— Scénario 1 : Deux conducteurs roulent avec une vitesse constante de 20 $km.h^{-1}$
— Scénario 2 : Deux conducteurs roulent avec une vitesse constante de 40 $km.h^{-1}$.
— Scénario 3 : Deux conducteurs roulent avec une vitesse constante de 50 $km.h^{-1}$.

Deux essais sont enregistrés par scénario.

4.5.4.2 Résultats expérimentaux

4.5.4.2.1 Evaluation de l'estimation des trajectoires et des positions

La trajectoire automatiquement estimée a été comparée à celle de la réalité du terrain mesurée manuellement par le suivi des véhicules, dont nous connaissions les caractéristiques de déplacement, pour illustrer la précision de l'outil. La comparaison a été faite en termes d'erreur moyenne ($RMSE$) et de coefficient de corrélation (R^2) qui ont été calculés comme expliqué aux Équations 2.11 et 2.12.

Le tableau 4.5 résume les résultats des calculs de la $RMSE$ et du R^2 entre les trajectoires estimées et celles réelles en utilisant les deux méthodes 'M1' et 'M2'. La figure 4.16 illustre la corrélation entre ces trajectoires.

TABLE 4.5 – Erreur moyenne et coefficient de corrélation entre les positions estimées des véhicules et celles réelles.

	$RMSE$	R^2
Trafic droit	4,35	0,99
Trafic gauche	16,62	0,97

Les résultats présentés dans le tableau 4.5 montrent que les positions des voitures détectées automatiquement sont proches de la réalité de terrain pour les cas de mono et multi-détections. En effet, le $RMSE$ ne dépasse pas 5 pixels pour la première séquence. De plus, les valeurs de R^2 illustrent que les positions estimées sont très corrélées avec celles réelles. Une erreur moyenne de l'ordre de $4,35$ pixel entre les positions estimées et celles de réalité de terrain est raisonnablement réduite dans un contexte de suivi de véhicule en mouvement sur la route avec une vitesse moyenne de 50 $km.h^{-1}$.

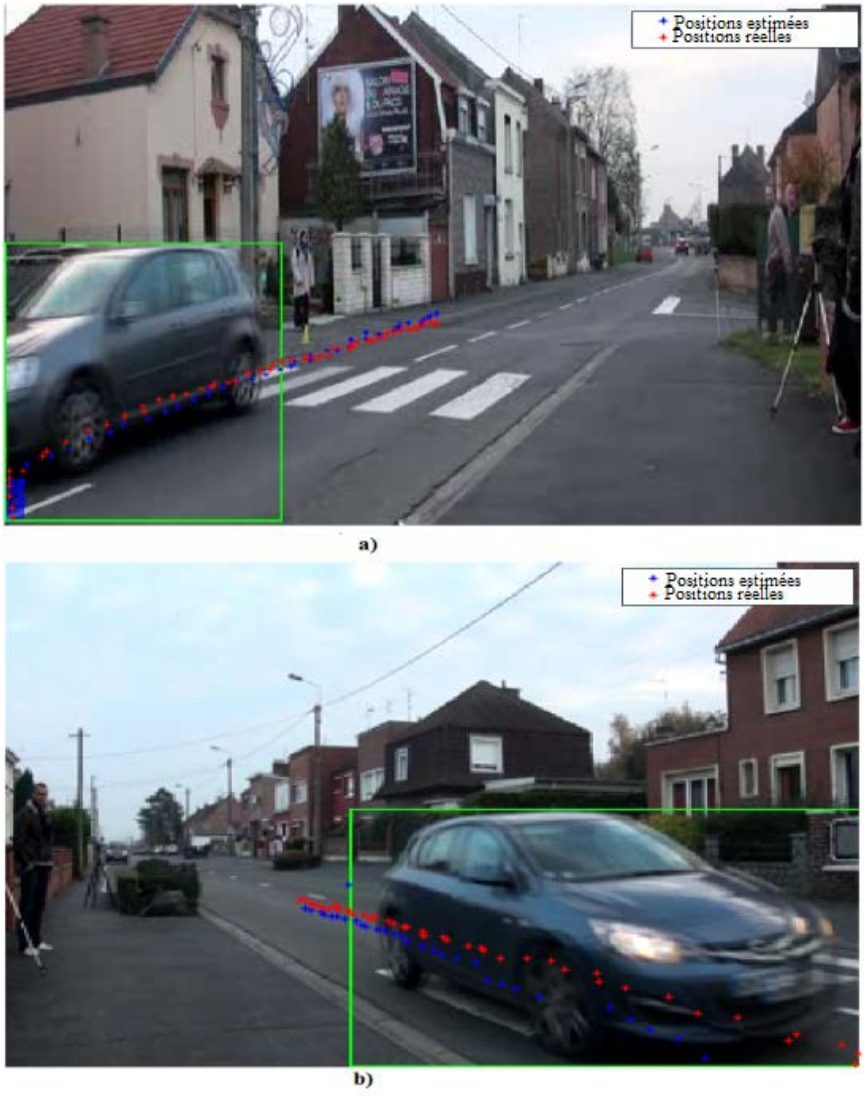


FIGURE 4.16 – Superposition des trajectoires estimées et de celles de réalité du terrain.

4.5.4.2.2 Evaluation de l'estimation des vitesses moyennes

L'évaluation de la précision de l'outil automatique de collecte des paramètres du trafic en termes d'estimation de la vitesse moyenne est présentée à travers la comparaison des vitesses estimées à celles indiquées par le compteur du véhicule et vérifié sur le terrain, suivant les 3 scénarios présentées pour l'étude de cas (*section 4.3.3.2.1*). Les estimations de vitesses sont réalisées pour le trafic gauche et droit. Les tableaux 4.6 et 4.7 résument ces estimations respectivement par les méthodes M1 et M2. Alors que le tableau 4.8 présente les valeurs de l'erreur moyenne entre les vitesses automatiques et manuelles.

En se basant sur ces résultats expérimentaux, on peut dire qu'il existe un bon accord entre les vitesses des véhicules mesurées manuellement et automatiquement ($RMSE = 1,3\ km.h^{-1}$ et $2,85\ km.h^{-1}$, respectivement pour le trafic gauche et droit). Cependant, les erreurs résiduelles peuvent être attribuées à l'inexactitude du calcul,

TABLE 4.6 – Evaluation de l'estimation des vitesses moyennes avec la méthode M1.

		M1					
		Trafic gauche			Trafic droit		
Estimation		**Scénario 1** 20 $km.h^{-1}$	**Scénario 2** 40 $km.h^{-1}$	**Scénario 3** 50 $km.h^{-1}$	**Scénario 1** 20 $km.h^{-1}$	**Scénario 2** 40 $km.h^{-1}$	**Scénario 3** 50 $km.h^{-1}$
Séquence 1	**D : Km**	0,63	1,09	1.18	0.74	1,27	1,38
	T :h	0,026	0,026	0.024	0.044	0,03	0,026
	$V : Km.h^{-1}$	24,23	41,92	49,16	16,95	42,33	53,07
Séquence 2	**D : Km**	0,93	1,05	0,98	0,80	0,428	1,15
	T :h	0,046	0,026	0,02	0,05	0,01	0,022
	$V : Km.h^{-1}$	20,21	40,38	49	16,40	42,80	52,27

TABLE 4.7 – Evaluation de l'estimation des vitesses moyennes avec la méthode M2.

		M2					
		Trafic gauche			Trafic droit		
Estimation		**Scénario 1** 20 $km.h^{-1}$	**Scénario 2** 40 $km.h^{-1}$	**Scénario 3** 50 $km.h^{-1}$	**Scénario 1** 20 $km.h^{-1}$	**Scénario 2** 40 $km.h^{-1}$	**Scénario 3** 50 $km.h^{-1}$
Séquence 1	**D : Km**	0,67	1,12	1,3	0,91	1,43	1,92
	T :h	0,026	0,026	0.024	0.044	0,03	0,035
	$V : Km.h^{-1}$	25,89	43,14	54,16	20,85	47,86	54,85
Séquence 2	**D : Km**	1,092	1,29	1,05	0,93	0,46	1,26
	T :h	0,046	0,026	0,02	0,05	0,01	0,022
	$V : Km.h^{-1}$	23,74	49,77	52,50	18,60	46,19	57,32

TABLE 4.8 – Erreur moyenne pour l'estimation des vitesses des véhicules pour le trafic gauche et droit

	Trafic gauche	Trafic droit
M1	1,30	2,85
M2	4,86	4,74

l'imprécision dans le calibrage de la caméra et les irrégularités dans les trajectoires des véhicules dues au bruit de la détection et de l'extraction des caractéristiques. On constate également que la méthode M2 semble moins performante que M1.

4.6 Conclusion

L'évolution fonctionnelle des rues et de la circulation, mais aussi le vieillissement de la population implique l'évolution des normes industrielles et des pratiques professionnelles afin de répondre aux nouveaux besoins des piétons en matière de sécurité et de mobilité. Mettre l'accent sur l'analyse du comportement des piétons implique la prise en considération des paramètres du trafic. La mise en place d'un outil automatique de collecte des données du trafic fiable à l'aide des techniques de vision par ordinateur

est un défi scientifique et technologique mais aussi un besoin sociétale, qui a motivé la définition de nos objectifs dans ce chapitre. Ainsi, le système développé dans le deuxième chapitre pour la détection et le suivi des piétons a été modifié en profondeur pour s'adapter aux particularités du mouvement des véhicules et pour discriminer les piétons et les véhicules. En effet, les paramètres de calcul du descripteur de l'HOG ont été adaptés pour le cas des véhicules par le changement de la taille des cellules et des blocs ainsi que par le nombre de classes d'orientation de gradients à prendre en considération. De même, l'optimisation de l'HOG par l'intégration de paramètres du mouvement améliore considérablement les performances en termes de taux de reconnaissance et de temps de calcul. Cet outil a été exploité pour la collecte des variables pour caractériser le trafic telles que la vitesse des véhicules, leurs directions ainsi que leurs positions spatio-temporelles. L'outil proposé a ensuite été testé sur notre base de vidéo CWV collectée à l'université de Valenciennes, en y sélectionnant plusieurs scénarios variables avec un et/ou plusieurs véhicule(s) en mouvement sur la route pour le trafic gauche et droit. Les variables de trafic seront ensuite, interprétées et corrélées avec les autres variables telles que celles des piétons pour générer des indicateurs de conflits P-V servant à inférer le comportement des piétons.

Chapitre

5 Approche probabiliste pour l'estimation des risques pris par les piétons

5.1 Introduction

L'objectif principal de notre approche automatique d'analyse comportementale des piétons lors de la traversée d'une rue est de déterminer si le piéton est en situation de danger ou sécuritaire par la corrélation (mise en relation) i) des données des piétons, ii) des véhicules et, iii) de l'environnement préalablement collectées, par l'analyse des vidéos. Plus précisément, il s'agit d'identifier les scénarios à risque et déterminer les paramètres responsables de ce danger.

Dans ce chapitre nous proposons alors, l'architecture d'un système de vision multi-couches pour la reconnaissance de scénarios à partir de séquence d'images. Le système de reconnaissance exploite une couche de reconnaissance de bas niveau et intègre des capacités de raisonnement décisionnel probabiliste. Nous proposons aussi une nouvelle stratégie de gestion de l'incertitude des capteurs de vision de la couche basse. Cette stratégie consiste à déterminer une degré de confiance aux données générées par ces capteurs à base de score de détection ainsi qu'une consistance temporelle aux observations de la même donnée. La robustesse de l'architecture proposée est illustrée par une étude expérimentale détaillée dont l'objectif est d'analyser le comportement des piétons lors de traversée d'une rue en fonction de l'aménagement urbain et du trafic routier.

5.2 Processus proposé : description générale

Dans ce chapitre nous modélisons le problème de conflit du trafic en exploitant une probable la relation de 'cause-effet' entre les observations et le comportement des piétons. cette modélisation de problème de conflit du trafic se base alors sur le concept de l'extraction automatisée d'informations pertinentes sur les objets présents dans des images et l'interprétation sémantique des séquences vidéo. Cette thématique de recherche présente un intérêt grandissant surtout avec le progrè remarquable des techniques de traitement bas niveau (*cf chapitres 2,3 et 4*) et haut niveau (IA).

Pour le traitement haut niveau de séquences vidéo, il s'agit d'interpréter les images en tirant profits de différents éléments de la scène et du contexte. Les systèmes d'interprétation pour la vidéo-surveillance sont construits à partir d'une juxtaposition de deux modules tels que la vision par ordinateur et l'IA. Comment concilier alors, le bas niveau qui est plutôt mathématique, avec le haut niveau, plutôt du domaine de l'intelligence artificielle ? Ce problème était autrefois largement vu comme un problème de contrôle uniquement (Boucher, 1999), et les efforts ciblaient l'amélioration de ce contrôle dans les systèmes de vision.

Ainsi, la relation 'cause-effet' entre les observations et le comportement des piétons permet l'utilisation d'un Réseau Bayésien (RB) pour mener ces analyses. Étant donné, aussi, que le RB présente la possibilité de rassembler et de fusionner des connaissances de diverses natures dans un même modèle avec une structure explicite et intuitive, il s'avère un choix adéquat. De même, comme le comportement du piéton est aléatoire, adaptatif et évolutif aux cours du temps, une décision bayésienne le caractérisera malgré tout. Nous proposons alors, une architecture d'un système de vision multi-couches pour la reconnaissance du comportement du piéton à partir de séquence d'images. Le système de reconnaissance exploite une couche de traitement de bas niveau pour la collecte de données et intègre des capacités de raisonnement cause-effet. La couche haut niveau est chargée d'interpréter les résultats issus des couches basses dans un contexte donné tel que la traversée d'une rue. Le système proposé met en œuvre une stratégie de conduite de décision malgré des données manquantes et bruités et intègre une technique de gestion des incertitudes des capteurs de vision pour une meilleure consolidation du bas niveau avec le haut niveau. L'architucture globale de ce système est présentée à la figure 5.1.

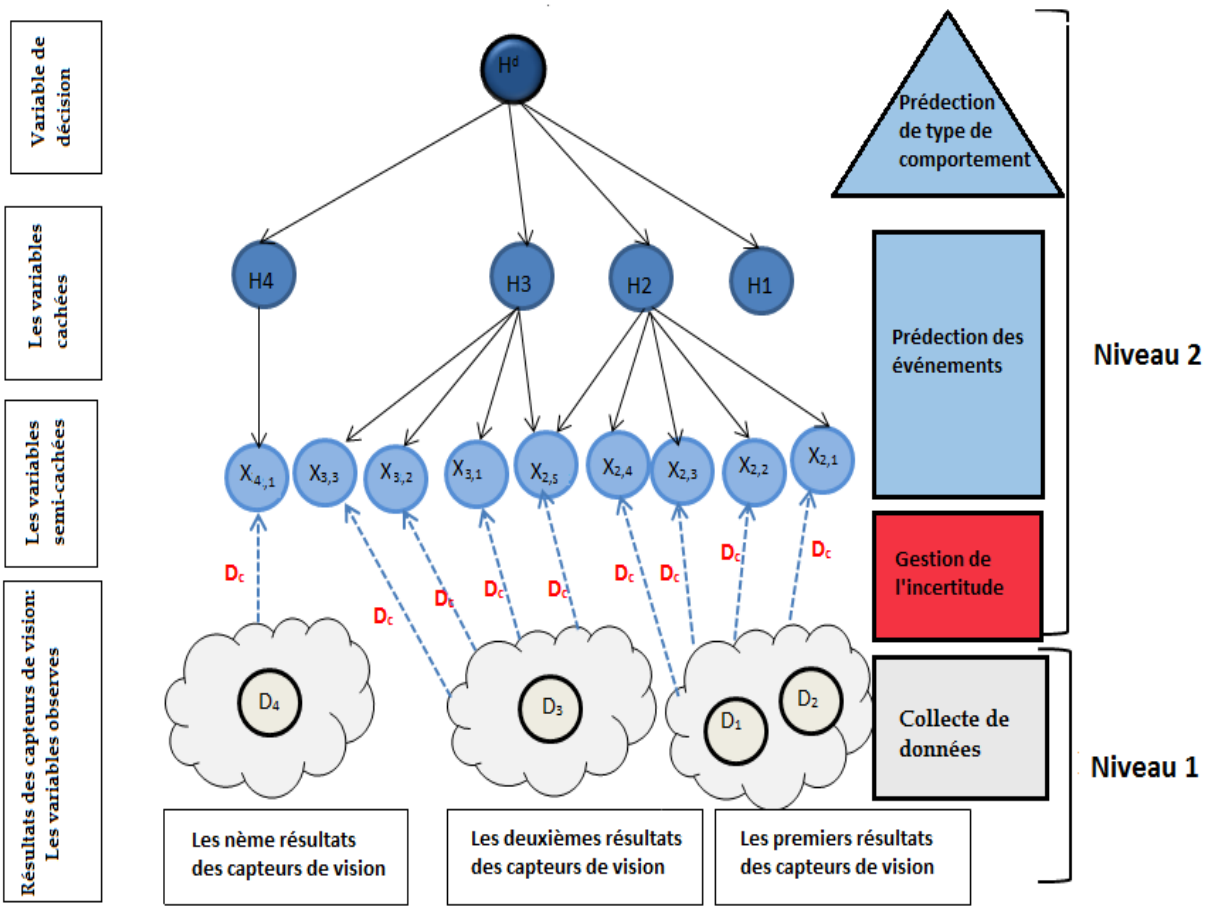


FIGURE 5.1 – Modèle générale de framework proposée pour la reconnaissance du comportement des piétons.

5.3 Approches existantes d'analyse comportementale et reconnaissance de scénarios

5.3.1 Interprétation haut niveau des séquences d'images

Depuis quelques années, les chercheurs en vision ont constaté que malgré la maturité du domaine, les systèmes développés manquent souvent de robustesse, d'adaptabilité et sont souvent très dépendants du domaine d'application. La "vision cognitive" est alors née de ce constat [195]. L'objectif affiché de cette discipline émergente est de rendre les systèmes de vision plus robustes et plus adaptables en les dotant de facultés cognitives que l'on peut résumer en quatre mots : savoir, comprendre, raisonner et apprendre. Le domaine de la vision cognitive utilise comme base la vision par ordinateur et rassemble également d'autres domaines tels que la reconnaissance de forme, l'intelligence artificielle, la robotique et les sciences cognitives.

L'interprétation haut niveau (sémantique) de l'image consiste alors, à établir un lien entre l'image, la vision et l'IA. En fait, le traitement d'images et la vision tirent partie des connaissances et des techniques d'IA pour gérer l'adaptation à un environnement

changeant, les informations incertaines, les systèmes hétérogènes de connaissances et les différents niveaux de prise de décision (Figure 5.2).

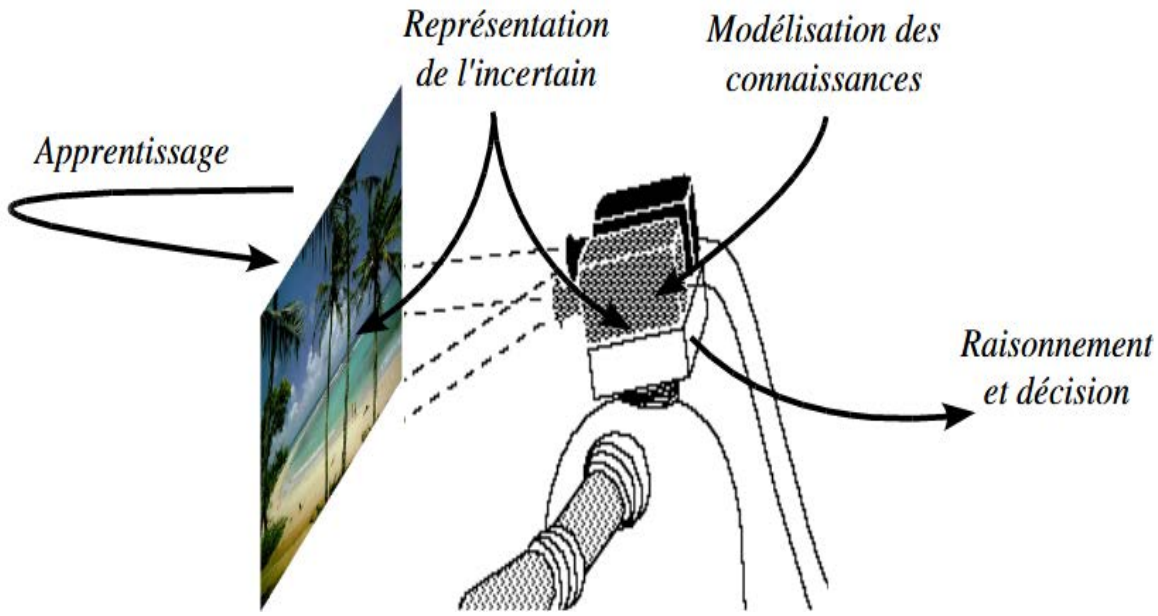


FIGURE 5.2 – Processus de traitement haut niveau de l'image [195].

L'une des applications les plus importantes des interprétations haut niveau de séquences d'images est la surveillance automatique du trafic et la sécurité routière qui a suscité un intérêt croissant ces dernières années. Dans ce domaine, l'apprentissage du "comportement" du trafic semble être la tâche la plus complexe, en particulier dans des environnements très dynamiques [196]. En outre, les facteurs de risque associés à la traversée de la route reprennent l'idée que la traversée est une phase critique qui peut causer un danger ou un accident impliquant des blessures graves ou encore la mort de l'individu.

Un comportement est essentiellement représenté par la combinaison d'activités atomiques, qui sont modélisées à partir des caractéristiques d'objet, telles que la position spatiale et temporelle, la direction ou la vitesse de déplacement. Afin d'apprendre le comportement, la corrélation entre les caractéristiques potentielles dans les dimensions spatiale et temporelle est classiquement utilisée dans le cadre de modèles graphiques probabilistes [197]. Comme les piétons ont un comportement aléatoire et variable, les piétons peuvent rapidement adapter leur comportement face à une situation de danger et peuvent également être soudainement confrontés à une situation de danger en raison d'une faible attention, du comportement des conducteurs, des manques dans les aménagements.... De plus, le comportement des piétons est souvent complexe puisqu'il est lié à des facteurs externes.

5.3.2 Principaux travaux existants

Le comportement humain, généralement défini comme un scénario [158], se compose d'un ensemble d'événements élémentaires liés à des contraintes causales et/ou temporelles. La difficulté du comportement humain réside dans sa complexité, sa variabilité spatiale et temporelle ainsi que l'incertitude qui existe sur toute tâche d'interprétation. Les approches de vision par ordinateur sont généralement axées sur l'approche numérique en utilisant les probabilités [196, 198, 199, 200, 201, 202] ou les réseaux de neurones [203] afin de faire face à l'incertitude des capteurs de vision. D'autre part, la communauté de l'IA a propose des approches symboliques plus souples permettant une capacité de reconnaissance de haut niveau [204, 205, 206].

Dans l'optique d'un système de surveillance automatique, la détection et le suivi des objets doivent être suivis d'une phase d'analyse et d'une reconnaissance du comportement des objets. L'approche développée dans [207] présente l'analyse du comportement comme étant une classification des motifs de mouvement produits par le module de suivi des objets. D'autres méthodes se sont révélées efficaces pour générer des modèles de comportement à partir des modèles du maximum d'entropie et de Markov [208]. Dans [209], la reconnaissance des comportements et des activités se fait à l'aide d'un modèle déclaratif pour représenter des scénarios et une approche logique pour reconnaître des modèles basés sur des scénarios prédéfinis.

Dans la surveillance routière automatique, les activités et les comportements des véhicules sont détectés et reconnus à des fins de surveillance et d'alerte. Une méthode simple de détection robuste d'objets en mouvement a été proposée par [210] pour la reconnaissance d'anomalies dans les scénarios intérieurs et extérieurs. L'algorithme de détection de changement bayésien (BCD) est utilisé pour éliminer le bruit, les ombres, les variations d'illumination et les mouvements répétés en arrière-plan.

Une fois que la séquences est segmentée, que les objets cibles sont reconnus et classés entre véhicules et piétons ainsi que le suivi est assuré ; la prochaine étape de la construction d'un système complet de surveillance visuelle, est l'inférence du comportement et la modélisation que ce soit de l'interaction véhicule/véhicule et/ou P-V.

Quelques exemples de travaux effectués pour détecter le comportement des véhicules à partir de vidéos sont [197, 206] et [211, 212, 213]. Dans [206] l'auteur a réalisé un système intégré basé sur des connaissances capables de traduire l'information visuelle en descriptions de langage naturel. L'accent a été mis sur l'analyse de scène de haut niveau, c'est-à-dire des représentations géométriques, comme cela pourrait être fourni par un système de vision, dans des descriptions linguistiques de mouvements d'objets. Dans une vision combinée et un système de langage naturel visant à des descriptions simultanées de langages naturels de l'imagerie dynamique, la reconnaissance des événements de mouvement doit se faire de manière incrémentielle afin de pouvoir être décrite même pendant leur progression. L'étude [211] présente un système complet pour la détection d'événements et la reconnaissance du comportement dans des vidéos issues d'une seule caméra mobile aéroportée. Cette reconnaissance de l'événement implique des humains et des véhicules et utilise un flux optique pour segmenter l'objet mobile de l'arrière-plan. Les méthodes de segmentation par flux optique sont très intensives et difficiles à réaliser en temps réel sans support matériel supplémentaire. En outre, il n'existe pas

de méthode robuste pour classer les cibles en différentes catégories.

Pour produire une analyse du comportement des objets en mouvement, un modèle générique a été proposé dans [214]. Ce modèle est constitué par deux blocs : le premier est la détection et le suivi de la région mobile ; Le second est l'intégration de la trajectoire et du contexte géospatial. Une combinaison du système de vision par ordinateur bayesian (BCVS) et des modèles couplés de Markov cachés (CHMM) [215] a été suggérée pour la modélisation des interactions entre les piétons dans les environnements extérieurs. Les HMM ont également été appliquées [206] pour organiser les activités observées à des endroits intérieurs (aux sien des bureaux) ou extérieures (suivi du trafic) en fonction de la réduction de l'entropie des distributions de composants. Cette approche a été adaptée pour inférer l'état caché à partir de vidéos ambiguës en se basant sur les orientations et les poses du corps humain. Dans [196] les auteurs ont recommandé un modèle de Markov caché dynamique (DML-HMM) comprenant le (EM) et le critère bayésien pour la classification de l'information (BIC). En outre, des réseaux probabilistes dynamiques (DPN) ont été formulés pour modéliser les corrélations temporelles et causales entre les événements discrets pour une interprétation holistique du comportement au niveau de la scène.

Pour unifier la reconnaissance des actions simples et complexes, dans [200] le HMM à action simple est codé à partir d'une grammaire stochastique qui modélise les actions complexes. Cette approche a considérablement amélioré la précision de la classification en développant une couche de reconnaissance des activités complexes en fonction de la reconnaissance des actions simples. Une autre approche utilisant une extension de la grammaire stochastique sans contexte (SCFG) a également été proposée pour modéliser les relations temporelles complexes entre les activités atomiques dans [216]. Sa principale innovation était un algorithme d'analyse multithreads adopté pour l'apprentissage de l'événement pour la reconnaissance, au lieu de l'analyseur commun.

La limitation commune de la plupart de ces approches basées sur l'HMM est la nécessité d'une grande quantité de données pour la phase d'apprentissage [217]. Par ailleurs, elles ne sont pas tellement adéquates pour les cas des analyses des comportements complexes qui sont par défaut représentés par plusieurs variables et/ou événements.

Un nouveau modèle thématique est proposé par [203] pour surmonter les inconvénients de la sensibilité, de la robustesse et de l'efficacité de l'analyse des comportements des objets. Comme un modèle thématique, on peut citer le modèle thématique de Markov (MCTM) qui s'appuie sur les réseaux bayésiens dynamiques (RBD) existants et les modèles bayésiens thématiques. Ce modèle a été conçu pour une analyse non supervisée des comportements dans des endroits publiques complexes et très fréquentés par les usagers. Dans une autre étude, les RBD ont été utilisés pour la reconnaissance du comportement dans un système de détection routier (Behavior Recognition in a Road Detection system : BRRD) [218] à travers des réseaux de capteurs de véhicules (VSN) pour inférer des événements routiers.

En outre, la détection de groupe à base de filtrage collaboratif offre une amélioration

des performances de détection. L'approche présentée dans [21] a proposé une solution élégante à la tâche de surveillance du stationnement. Ce travail a proposé un système de surveillance des piétons et des voitures qui modélise l'interaction entre deux agents utilisant un petit réseau de croyances. Les comportements et les interactions avec les objets ont été modélisés en utilisant un RB [216]. Trois modèles bayésiens hiérarchiques sont également proposés dans [204] : le mixture-modèle LDA, le mixture-modèle HDP et le modèle double HDP. Ces modèles servent à classer les pixels en mouvement et les séquences vidéo en des activités atomiques et des interactions. Les activités atomiques sont conçues comme des distributions de caractéristiques visuelles de bas niveau, telles que la position et la direction des pixels en mouvement, tandis que les interactions multi-agents sont modélisées comme des distributions des activités atomiques.

Théoriquement un système sûr (Figure 5.3) englobe trois ensembles de facteurs de risque et des interventions liés aux usagers de la route, aux véhicules et à l'environnement routier. Ce système permet de prendre des mesures de prévention plus efficaces [219, 220]. L'approche d'un tel système sûr, est adaptée et efficace dans plusieurs contextes. Il permettrait d'améliorer la sécurité routière "là où les progrès s'étaient avérés difficiles" [221].

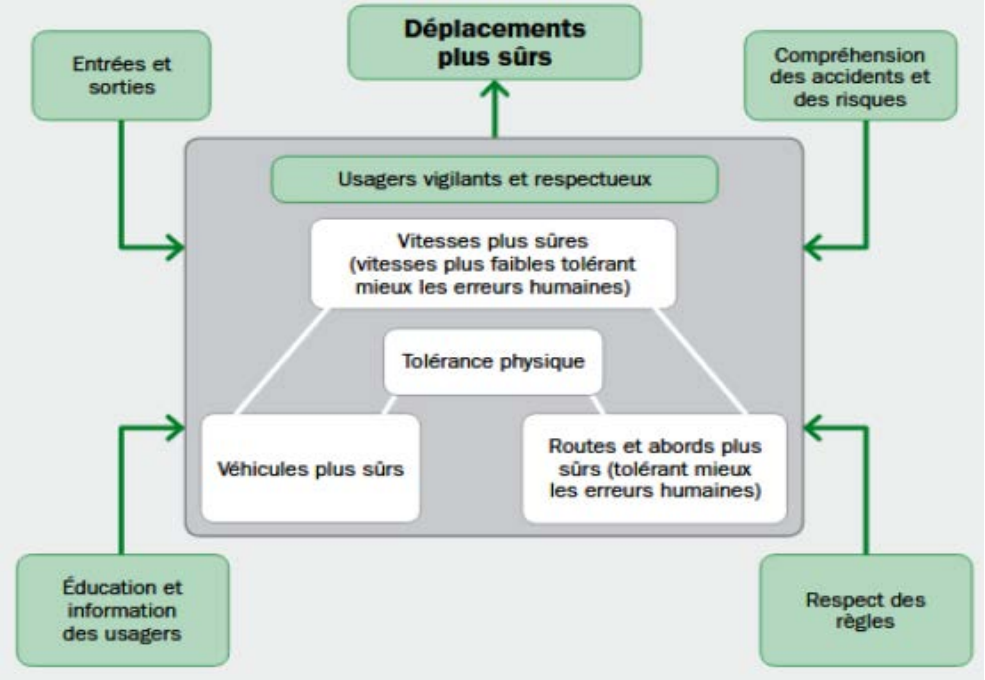


FIGURE 5.3 – Approche pour un système sûr proposé par [219].

Bien que de nombreuses tâches de surveillance visuelle difficiles aient été abordées dans les recherches ci-dessus, la tâche de considérer les activités et les interactions avec des structures temporelles complexes reste toujours un enjeu majeur [2]. En fait ni l'aspect évolutif du comportement, ni l'interaction temporelle entre les différents agents à hypothèse ne sont traités. En outre, étant donné que les observations générées par les techniques de vision sont incertaines, manquantes et parois même indisponibles à cause de problèmes d'occultation, d'éclairage, de multi-détection et suivi (split et

merge)..., l'utilisation de ces observations comme étant des variables sures n'est pas assez fiable. Ainsi, en se basant sur cette étude de l'existant on constate que le degré de confiance et de certitude associé aux capteurs de vision n'est pas propagé dans les différents niveaux de RB et pris en compte dans la décision finale générée.

5.4 Approche proposée

5.4.1 Architecture globale

En tenant compte des aspects bayésiens, du comportement humain incertain et adaptatif, nous avons choisi de mettre en œuvre une approche active basée sur un contrôle des activités du système d'interprétation en fonction du comportement du piéton et de ses décisions en relation avec l'environnement (trafic et aménagement).

Par ailleurs, nous privilégions une modélisation des imperfections des capteurs de vision, des relations causales entre les piétons et l'environnement et de l'incertitude d'estimation des conflits du trafic à partir de RB qui est considéré comme un compromis entre les approches normatives et descriptives. Cet outil propose un formalisme cohérent et intuitif qui peut représenter des informations hétérogènes, incertaines et imprécises.

Les RB, appelés aussi réseaux de croyances, appartiennent à l'ordre des modèles probabilistes graphiques. Ces structures graphiques sont utilisées pour exprimer des connaissances sur des domaines incertains. Les probabilités conditionnelles dans les graphiques sont souvent estimées en utilisant des méthodes statistiques et informatiques appropriées. Par conséquent, le RB combine les principes de la théorie des graphes, de la théorie des probabilités, de l'informatique et de la statistique. Pour les RB, on peut dire qu'ils permettent une représentation efficace et aussi un faible calcul de la distribution des probabilités conjointes sur un ensemble de variables aléatoires. L'application de RB dans différents domaines tels que l'analyse du comportement des piétons est devenue extrêmement populaire au cours de la dernière décennie. La structure graphique de ces réseaux convient parfaitement pour combiner les connaissances antérieures et les données observées. Elle traite également des données incertaines et incomplètes. Pour ces raisons, l'application de RB est également possible dans le cas de relations de cause à effet et de compréhension des différents aspects d'un problème avec la prévision d'événements futurs.

Bien que les données des capteurs de vision soient riches en détails et diversifiées, elles sont habituellement incertaines et intègrent des doutes. En effet, pour prendre en compte l'observation incertaine dans ce travail, nous proposons un nouveau coefficient de degré de confiance relatif aux performances des capteurs de vision. Le degré de confiance sera calculé en combinant le score de détection et la consistance temporelle des données observées. Ces incertitudes seront introduites dans les couches de RB en utilisant la règle de "Jeffry" pour calculer la probabilité causale [222].

L'approche proposée pour l'analyse automatique du comportement des piétons consiste à modéliser les scénarios de traversée d'une rue, puis le système doit reconnaître si le comportement actuel est à risque ou sécurisé. Le système d'interprétation proposé est alors basé sur une organisation multi-couches. Nous distinguons trois principales couches : La première couche liée aux événements ou scénarios de bas niveau,

une couche intermédiaire pour la gestion de l'incertitude des observations de vision et la propagation du degré de confiance et enfin une couche haute de reconnaissance de scénarios.

Dans la couche 1, les événements élémentaires ou observations de base seront générés à partir des données collectées par les méthodes et techniques de vison. Ces observations sont celles déjà détaillées dans les chapitres 2, 3 et 4. En effet, les paramètres piétonniers (vitesse, position, âgée), de trafic (vitesse des véhicules, direction, position, présence ou non...) et d'aménagement (passage piéton avec ou sans feu) seront tous modélisés dans cette couche pour générer des observations élémentaires qui seront combinées sous forme d'indicateurs de conflit du trafic contribuant à la prédiction du comportement du piéton.

Cependant au niveau de "la couche 2", le degré de confiance associé à ces observations élémentaires générées par les capteurs de vision sera calculé et propagé à la couche haute. Les performances des techniques de vison ainsi que la consistance temporelle de l'observation générée seront prises comme indicateurs de confiance de ces observations.

La couche 3 permet de reconnaitre la probabilité d'occurrence d'un scénario de haut niveau en exploitant les résultats de reconnaissance de la couche (1) pondérés par le degré de confiance généré par "la couche" intermédiaire (2). Cette couche peut intégrer des méthodes de combinaison très diverses, et elle peut être basée sur une simple opération logique (et, ou ..) entre des propositions des couches inférieures, ou vérifier un ordre d'occurrence d'événements suite à une phase d'apprentissage.

La classification des comportement des piétons " sécuritaire ou dangereux" *'H4'* dépend conditionnellement de 3 hypothèses (Figure 5.4) :

— *H1* : Y a-t-il un gap sécuritaire suffisant ?
— *H2* : Le piéton respecte-t-il les aménagements urbains installés sur la zone s'ils existent (exemple feu, rouge ou vert) ?
— *H3* : L'âge affecte-t-il le comportement du piéton ?

Les probabilités sont issues de l'application de la règle de décision de Bayes en rajoutant l'hypothèse d'indépendance conditionnelle des propriétés *H1*, *H2* et *H3* conditionnellement à l'événement élémentaire *H4* (Équation 5.1).

$$P(H4|H1,\ldots,H3) = \alpha P(H4) \prod_{i=1\ldots3} P(Hi|H4) \tag{5.1}$$

Avec α est un facteur multiplicatif tel que l'on ait

$$P(H4|H1,...,H3) + P(\neg H4|H1,...,H3) = 1$$

.

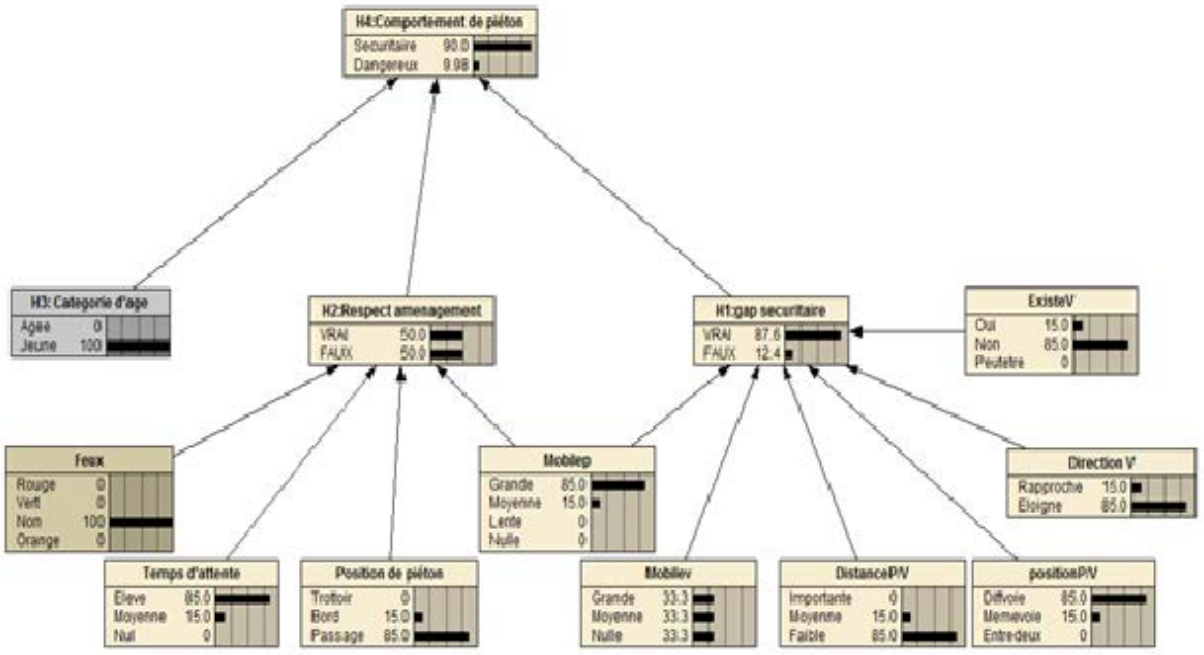


FIGURE 5.4 – Structure du réseau bayésien proposée.

L'application d'un réseau bayésien repose sur trois étapes majeures (Figure 5.5) :

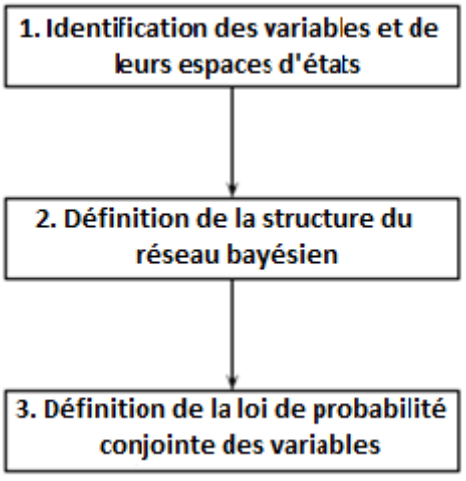


FIGURE 5.5 – Étapes de construction d'un réseau bayésien [223].

1. **Identification des variables et de leurs espaces d'états**
   La première étape de construction du réseau bayésien est l'identification de l'ensemble des variables X_i, catégorielles ou numériques, qui caractérisent le système. Une fois que les variables sont identifiées, l'étape suivante consiste à préciser " l'espace d'états " de chaque variable X_i, dans le sens de déterminer l'ensemble des valeurs possibles pour chaque variable [223].

2. **Définition de la structure du réseau bayésien**

   La deuxième étape consiste à identifier les liens entre les variables, c'est à-dire à répondre à la question : pour quels couples (i, j) la variable X_i influence-t-elle la variable X_j ? [223].

   La fixation de la structure du réseau peut se faire à partir de deux modalités :
   — L'apprentissage : La définition de la structure du RB à la base de l'apprentissage est sans aucune hypothèse sur la structure du réseau. L'expert cherche la structure qui, "une fois munie des meilleurs paramètres, rende compte le mieux possible des données observées" [223]. Dans le cas général, c'est-à-dire si l'on n'impose aucune contrainte à la structure, le problème est relativement énorme. Pour dix variables, il existe 4.10^{18} structures possibles [223].
   — Par l'expert : La structure peut être fixée manuellement par les experts du domaine en se basant sur les connaissances, sur le contexte ainsi que sur les relations causales possibles entres les variables sous-jacentes.
   Cette méthode est la plus encouragée par la communauté scientifique lors de l'utilisation du RB pour une application bien définie [223]. Par ailleurs, nous l'avons adopté lors de la fixation de la structure de l'architecture de notre approche probabiliste pour l'analyse comportementale des piétons lors de la traversée d'une rue. Nous nous sommes basés sur des références multiples d'accidentologies [77, 13, 108, 117, 116, 64, 193, 194, 68] pour identifier les indicateurs de conflit du trafic les plus pertinentes et leurs potentielles relations causales afin de construire les nœuds ainsi que les arrêtes du RB.

3. **Loi de probabilité conjointe des variables ou apprentissage des paramètres de réseau**

   La dernière étape de construction du réseau bayésien est la détermination des tables de probabilités associées aux différentes variables. D'abord, les connaissances des experts concernant les lois de probabilité des variables sont intégrées au modèle. Deux cas se présentent selon la position d'une variable X_i dans le réseau bayésien :
   — La variable X_i n'a pas de variable parente : les experts doivent préciser la loi de probabilité marginale de X_i.
   — La variable X_i possède des variables parentes : les experts doivent exprimer la dépendance de X_i en fonction des variables parentes, soit au moyen de probabilités conditionnelles, soit par une équation déterministe (que le logiciel convertira ensuite en probabilités).

5.4.2 Reconnaissance des événements et choix des variables

5.4.2.1 Agents pour la reconnaissance d'événement élémentaire

Un événement élémentaire est un fait reconnu par une observation ou par l'évolution des valeurs d'un ensemble de variables d'observation. Pour caractériser les événements, nous utilisons une variable latente (variable non mesurée). En fait, cette variable latente correspond à l'hypothèse de reconnaître l'événement particulier et les autres variables sont liées aux observations des capteurs. Les événements proposés H_i et les observations

$X_{i,j}$ pour caractériser le comportement des piétons lors de la traversée de la rue compte tenu des paramètres des véhicules et des piétons à deux reprises "t1" et "t2".

Étant donnée, qu'un comportement à risque peut être causé soit par la violation des règles de circulation par des piétons désobéissants : "Certains piétons n'ont pas attendu le feu vert pour traverser la route et/ou regardé des deux cotés de la route avant de traverser", soit par une mauvaise estimation ou manipulation du gap sécuritaire. La quantification de l'estimation par le piétons du niveau de sécurité est une tâche très difficile à faire. Le résultat est corrélé à l'âge des piétons, à sa vitesse et à sa position, ainsi qu'aux paramètres de la voiture tels que sa vitesse, sa position et sa direction. En effet, des études antérieures ont abordé un certain nombre de facteurs comportementaux par rapport à la survenue d'accidents. Ces facteurs comprennent nottament l'ignorance des feux de circulation, la traversée de la rue sur un passage piéton non équipé, l'incertitude ou le retard dans la traversée de la route par rapport à un gap sécuritaire, en courant pour traverse la route et l'inattention des piétons [215, 224, 225]. L'âge est également un facteur déterminant dans les accidents de la circulation [218].

Ces constations nous conduisent à formaliser les variables et les agents à hypothèse détaillés aux sous-sections suivantes.

5.4.2.2 Proposition des variables observées

La première étape de construction du réseau bayésien est la seule étape qui demande absolument l'intervention humaine et les connaissances de l'expert sont indispensables. Il s'agit de déterminer l'ensemble des variables X_i, catégorielles ou numériques, qui caractérisent le système. Puisque dans tout travail de modélisation, si les variables sont identifiées, il est ensuite nécessaire de préciser l'espace d'états de chaque variable X_{ij}, c'est-à-dire l'ensemble de ses valeurs possibles.

La majorité des logiciels de RBs ne traite que des modèles à variables discrètes, ayant un nombre fini de valeurs possibles, comme les fonctionnalités d'apprentissage sont réservées aux RBs à variables discrètes uniquement [223]. Il est impératif alors, de discrétiser les plages de variation des variables continues. Tenant compte de cette imposition du RB, nous nous basons sur des seuils et des références théoriques pour la discrétisation et la définition des plages des variables utilisées pour la conception de l'architecture du modèle proposé. Ces discrétisations sont détaillées comme suit.

X_{11} : "Mobilep" : Mobilité du piéton dont on calcule la vitesse à partir du processus de détection et de suivi. Pour cette variable, on a choisi 3 valeurs possibles : Grande, Moyenne, Nulle (pour l'état d'arrêt). Ces valeurs sont estimées en se basant sur la valeur de la vitesse moyenne de déplacement d'une personne normale ($1,2\ m.s^{-1}$). Cette variable déclenche les deux hypothèses H_1 et H_2.

X_{12} : "Mobilev" : Mobilité du véhicule dont on calcule la vitesse à partir du processus de détection et de suivi. Pour cette variable, on a choisi 3 valeurs possibles : Grande, Moyenne, Nulle (pour l'état d'arrêt). Cette valeur est estimée par rapport à la vitesse autorisée par la loi dans la zone de travail.

X_{13} : "Distance P/V" : Distance entre le piéton et le véhicule dont on calcule la distance euclidienne entre la position du piéton en (x, y) et la position du véhicule en

(x, y) à un instant donné. Pour cette variable, on a choisi 2 valeurs possibles : Faible, importante.

X_{14} : "Position P/V" : Position du piéton et du véhicule dans la même voie ou non en se basant sur la projection de leurs positions (x, y) sur le modèle de passage piéton.

X_{15} : "DirectionV" : Détermine la direction du véhicule en mouvement. On associe à cette variable deux valeurs : vers le passage piéton et s'éloigne du passage piéton.

X_{16} : "ExisteV" : Cette variable détermine si des véhicules sont présents. Cette information est décisive en ce qui concerne le gap sécuritaire et le type de comportement. Lorsque l'on n'a pas de véhicule, on a un gap sécuritaire et une forte probabilité que le comportement soit sécuritaire. Alors que dans les cas contraires de "Oui" et "Peut être" il faut voir les différentes valeurs des autre variables.

X_{21} : "La position du piéton" : Déterminée par la projection de la trajectoire du piéton sur le modèle de passage piéton. Pour cette variable, on a choisi 3 valeurs possibles : Sur le trottoir, au bord du passage piéton, au milieu du passage.

X_{22} : "Temps d'attente" : Temps que le piéton a passé au bord du passage piéton. Pour cette variable, on a choisi 3 valeurs possibles : Élevé (s'il est supérieur à $3s$), Moyen (si $0 <$ Temps d'attente $< 3s$), Nul.

X_{23} : "Feux" : Indique l'existence ou non de feu(x) sur la zone. Cette variable est déterminée manuellement. Pour cette variable, on a choisi 3 valeurs possibles : Rouge, Vert, Non.

X_{31} : "Catégorie d'âge" : La classe d'âge du piéton détermine si le piéton et âgé ou jeune à partir des caractéristiques extraites de la silhouette (SM). Pour cette variable, on a choisi 2 valeurs possibles : Âgé, Jeune.

Les figures 5.6, 5.7 et 5.8 illustrent les quantifications des observations des capteurs de vision et leurs discrétisations.

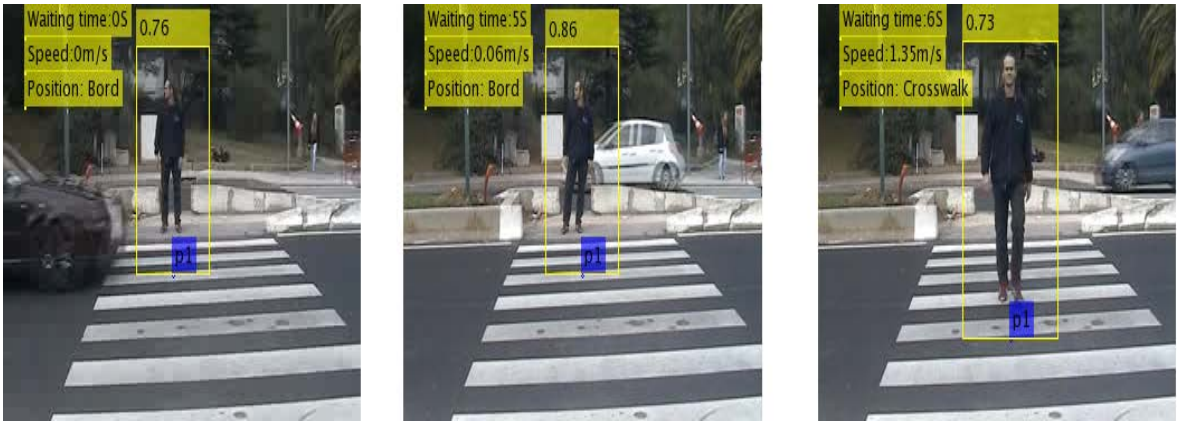
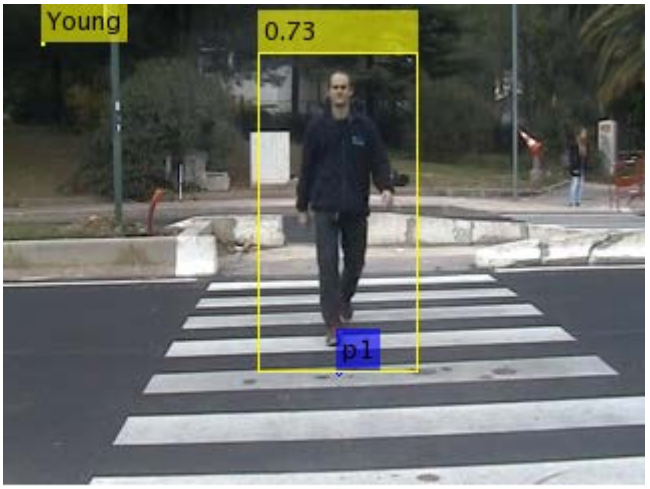


FIGURE 5.6 – Discrétisation des observations des capteurs de vision concernant les variables du piéton.



a)



b)

FIGURE 5.7 – Discrétisation des observations des capteurs de vision concernant la classe d'âge : a) Cas d'un piéton jeune, b) Cas de piétons âgées.

5.4.2.3 Proposition des agents à hypothèses

Les observations générées à partir de l'analyse des séquences vidéo seront ainsi mise en relation pour la construction des indicateurs de conflit P/V ou dite aussi des agents à hypothèses. La propositions des ces agents est la suivant.

H_1 : Il y a bien un gap sécuritaire entre le piéton et le véhicule. On lui affecte deux valeurs possibles : (vrai ou faux)

- Faux : si le piéton et le véhicule sont tous les deux en mouvement c'est–à-dire chacun d'eux a une vitesse différente de zero. Le piéton et le véhicule sont tous les deux sur la même voie et le véhicule se dirige vers le passage piéton (collision possible entre

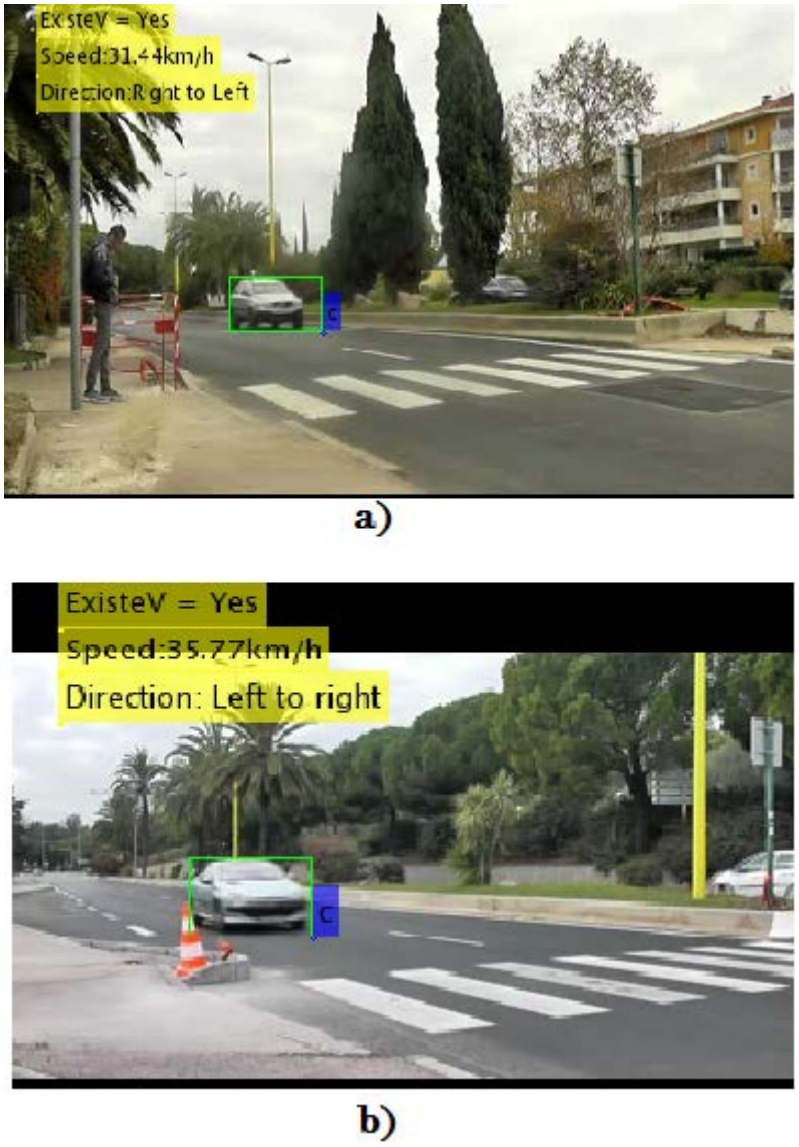


FIGURE 5.8 – Discrétisation des observations des capteurs de vision concernant les variables du trafic : a) Trafic droit, b) Trafic gauche.

le piéton et le véhicule). La distance euclidienne entre la position du piéton (x, y) et la position du véhicule (x, y) est faible.

- Vrai : Il y a un gap sécuritaire entre le piéton et le véhicule si l'on à l'une des 3 observations déjà citées (direction, positionnement sur la même voie, distance minimale).

• Vrai : si le piéton et le véhicule sont en mouvement mais sur deux voies différentes.

• Vrai : si le piéton et le véhicule sont en mouvement sur la même voie mais le véhicule s'éloigne du passage piéton.

• Vrai : si le piéton et le véhicule sont en mouvement sur la même voie mais la distance entre eux est importante.

H_2 : Le piéton respecte les aménagements

- Vrai : si le piéton est arrêté au bord du passage piéton au feu rouge. Par conséquent le piéton a une vitesse nulle pour une durée $T > 0$, sa position (x, y) est sur le trottoir et le feu est rouge.
- Vrai : si le piéton est en mouvement et sa position est soit au bord du passage piéton soit sur le passage avec le feu vert.

H_3 : L'âge affecte le comportement du piéton (le piéton est âgé ou jeune) :
- Vrai si la probabilité d'avoir un scénario dangereux est élevée. Cette probabilité est plus importante si le piéton est âgé.

H_4 : Le comportement du piéton est sécuritaire ou dangereux : Étant donnée la valeur de H_1, H_2 et H_3, déterminer si le piéton est dans une situation dangereuse ou non. Cette réponse est probabiliste, on associe un pourcentage à chaque état suivant la mise à jour du tableau de probabilité conditionnelle des variables dont dépend le comportement du piéton :

— Comportement sécuritaire : Si le piéton respecte les aménagements.
— Comportement sécuritaire : S'il y a un gap sécuritaire entre le piéton et le véhicule en mouvement.
— Comportement dangereux : correspond à la négation des hypothèses sécuritaires.

La figure 5.9 illustre les propositions des agents à hypothèse et leur discrétisation en fonction des variables observées. Alors que l'architecture globale du modèle décisionnel proposé est présentée à la figure 5.10 avec une annotation détaillée des différentes variables est hypothèses dans les différentes couches du modèle.

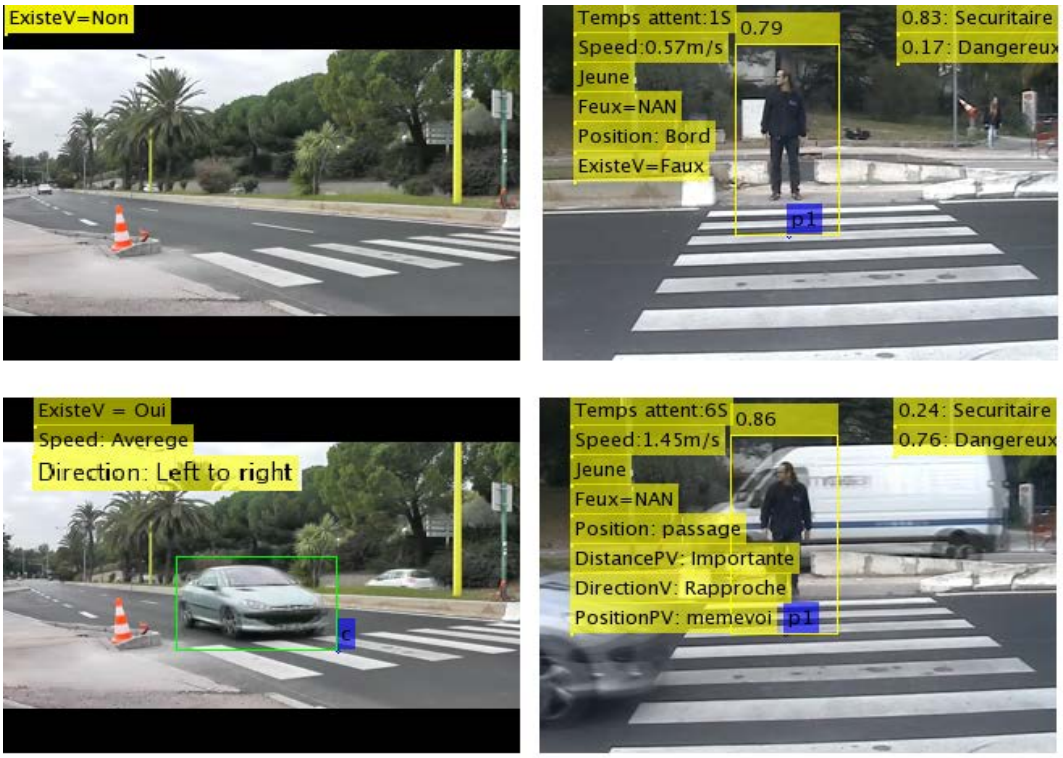


FIGURE 5.9 – Discrétisation des agents à hypothèses en fonction des variables observées.

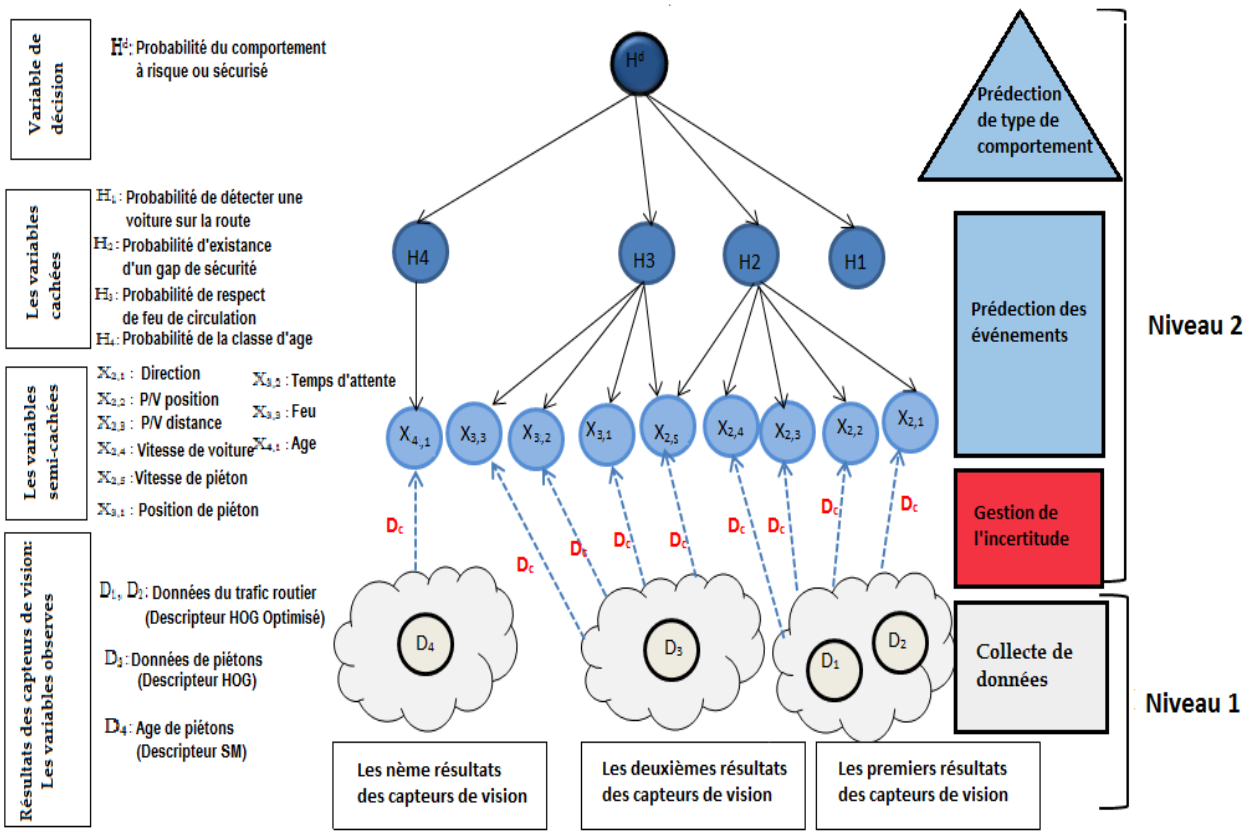


FIGURE 5.10 – Illustration de l'architecture du modèle décisionnel proposé.

5.4.3 Estimation des paramètres du réseau bayésien

L'apprentissage des paramètres du RB consiste à estimer les distributions de probabilités (ou les paramètres des lois correspondantes) à partir de données disponibles. Nous décrirons ici les méthodes les plus utilisées dans le cadre des RB, selon que les données manipulées sont complètes ou non en premier lieu et nous détaillerons ensuite notre proposition pour la gestion des imperfections de l'incertitude des capteurs de vision.

5.4.3.1 Apprentissage avec données complètes

La méthode la plus simple et fréquemment utilisée dans le cas où toutes les variables sont disponibles, est l'estimation statistique qui consiste à estimer la probabilité d'un événement par sa fréquence d'apparition dans la base de données. Cette approche, appelée maximum de vraisemblance (MV), nous donne alors (Équation 5.2) :

$$P(X_i = x_k | pa(X_i = x_j) = \theta^{MV}_{i,j,k} = \frac{N_{i,j,k}}{\sum_k N_{i,j,k}} \tag{5.2}$$

Où $N_{i,j,k}$ est le nombre d'événements dans la base de données pour lesquels la variable X_i est dans l'état x_k et ses parents sont dans la configuration x_j.

L'estimation que nous venons d'évoquer (maximum de vraisemblance) n'est valable que si les variables sont entièrement observées. Les méthodes suivantes vont donc essayer de traiter le cas où certaines données sont manquantes ou incertaines.

5.4.3.2 Gestion des imperfections de la perception et apprentissages avec données incomplètes

Les observations gérées par le système d'interprétation sont généralement imparfaites. Elles peuvent être imprécises, incertaines ou incomplètes. "L'incertitude est relative à la vérité d'une information, en d'autres termes, elle exprime un doute sur la validité de cette information. L'imprécision est liée au contenu de l'information et représente un défaut quantitatif. L'incomplétude représente l'absence d'une partie de l'information. Il est alors primordial d'intégrer dans la stratégie de conduite des moyens de gestion des imperfections que l'on résume généralement par le terme générique de la gestion des incertitudes." [226]

Les résultats des capteurs peuvent être entachés d'erreurs de détection, de position, de suivi, etc. Certains attributs sont observés systématiquement, d'autres sont observés occasionnellement et enfin d'autres ne sont jamais observés. Lors de la mise en place de la base d'apprentissage, nous intégrons cette absence d'observation totale ou partielle de certains paramètres due aux capteurs logiques. Lorsqu'ils sont incapable de prendre une décision. Cette notion de non décision est alors traduite par un signe ' ?' en tant que donnée manquante (ou incomplète) dans la base d'apprentissage.

5.4.3.3 Proposition d'une technique de gestion de l'incertitude et degré de confiance pour les observations des capteurs de vision

Malgré les nombreuses avancées dans ce domaine de recherche, le problème de contrôle entre le bas niveau et le haut niveau n'est toujours pas résolu aujourd'hui de manière générique, et les percées vers le grand public des applications de vision haut niveau ne sont pas encore à la hauteur des attentes [54]. Malgré les progrès qu'ont connues les techniques de vision par ordinateur ces derniers décennies, elles ont encore quelques lacunes quantitatives et qualitatives au niveau des données générées. Ces données peuvent être soit manquantes soit incertaines. Ces incertitudes des capteurs de vision ne sont pas encore suffisamment bien contournée et propagées dans la couche haute pour la prise de décision la plus nuancée et la plus appropriée..

La couche de RB avec son architecture actuelle ne présente aucune technique pour exploiter l'incertitude des capteurs de vision dans la prise de décision. En effet, une décision nuancée doit nécessairement prendre en compte la qualité incertaine des valeurs des variables corrélées et l'étendre aux différents niveaux du RB. Nous proposons pour cela une méthode de réfection et la diffusion d'observations incertaines quantitatives et qualitatives générées par les capteurs de vision. Cette méthode est une mesure qui calcule un degré de confiance associé aux valeurs des variables du RB. Cette métrique encapsule à la fois la précision de la détection et du suivi (performances du détecteur) ainsi que la consistance temporelle de l'observation.

La métrique proposée pour calculer le degré de confiance des observations est la multiplication du taux de performance du détecteur (score : s) par la consistance tem-

porelle (c) de cette observation en une seconde comme expliqué par l'équation 5.3.

$$D_c = s \times c \tag{5.3}$$

5.4.3.3.1 Consistance temporelle

La consistance temporelle consiste à suivre la variation d'une observation durant un intervalle de temps. En fait, une observation générée par les capteurs de vison peut avoir plusieurs valeurs selon les conditions et la précisions des méthodes et techniques utilisées. De ce fait, afin de quantifier la fiabilité des capteurs de vision nous proposons ici de calculer la fréquence d'occurrence d'une observation donnée pendant un intervalle de temps.

Sur le plan pratique on procède de la manière suivant pour le calcul de la consistance temporelle : Initialement la valeur de cette consistance est mise à $'0'$. Puis, si cette observation est présente, cette valeur sera incrémentée sinon elle sera décrémentée. Si l'observation garde la même valeur durant tout l'intervalle de temps, la consistance temporelle retenue sera égale à 1, sinon si elle est absente, elle sera mise à 0. En fait, ce coefficient d'occurrence de l'observation, a une valeur minimale égale à 0 et une valeur maximale égale à 1, c'est-à-dire une fois qu'il atteint la valeur 1 ou 0, il ne sera plus incrémentée ou décrémentée.

En effet, pour suivre le comportement aléatoire des variables des piétons et du trafic, les fréquences d'occurrence des valeurs de ces variables sont enregistrées chaque seconde pendant la traversée de la rue. Par conséquent, le coefficient de consistance temporelle, initialement égal à 0, augmentera ou diminuera de $\frac{1}{n}$ (n taux d'acquisition de la caméra : nombre d'images par seconde).

5.4.3.3.2 Score de détection

Tout processus de détection basé sur des capteurs de vision est généralement évalué à l'aide d'une métrique quantitative telle que le taux de reconnaissance correcte, le rappel, la précision... qui représentent les performances du détecteur en termes de nombre de réponses correctes. Dans le cas de l'entrepôt de données et de la prise de décision, l'évaluation quantitative des capteurs de vision ne suffit pas. Une évaluation qualitative devrait alors être ajoutée. La méthode qualitative consiste à mettre en évidence la précision de la détection correcte. La métrique adoptée représente avec quelle précision la boîte englobante inclut l'objet cible.

Le score de détection indique la métrique qualitative utilisée, pour représenter la précision de détection afin de calculer le degré de confiance associé au capteur de vision. HOG est calculé pour chaque fenêtre de recherche en balayant l'image par fenêtre comme cela est présenté dans les chapitres 2 et 4 en se référant au travail [89]. Plus d'une détection peut être générée à la fin du balayage d'image. Ces détections seront fusionnées en fonction du processus de suppression non maximum tel qu' illustré par l'algorithme 3 [89]. Le score de détection (le poids de détection ϖ_i) est calculé comme suit (Équation 5.4 [89]) :

$$\varpi_i(y) = \frac{|H_i|^{-1/2} t(w_i) exp(-D^2\left[y, y_i, H_i\right]/2)}{\sum_{i=1}^{n} |H_i^{-1/2} t(w_i) exp(-D^2\left[y, y_i, H_i\right]/2)} \tag{5.4}$$

Avec y_i, $i = 1...n$ est la liste des détections (en 3D et avec les différentes échelles) générées par le processus de détection et H_i est la matrice de covariance. H_i est une matrice diagonale. Soit $diag[H]$ qui représente les 3 éléments diagonaux de H. Nous utilisons des matrices de covariance selon l'échelle, comme présenté par l'équation 5.5.

$$diag[H_i] = \left[(exp(s_i)\sigma_x)^2, (exp(s_i)\sigma_y)^2, (\sigma_x)^2\right] \tag{5.5}$$

(a) Let start scale $S_s = 1$; compute end scale $S_e = min[W_i/W_n, H_i/H_n]$, where W_i and H_i are image width and height, respectively
(b) Compute the number of scale steps S_n to process

$$S_n = floor(\frac{(log(S_e/S_s))}{(log(S_r))} + 1)$$

—
For each scale $S_i = [S_s, S_r, ..., S_n]$

(a) Rescale the input image using bilinear interpolation
(b) Extract features (HOG descriptor) and densely scan the scaled image with stride Ns for object/non-object detections
(c) Push all detections with $t(w_i) > c$ to a list

—
Non-maximum suppression

(a) Represent each detection in 3-D position and scale space y_i
(b) Compute the uncertainty matrices Hi for each point
(c) Compute the mean shift vector iteratively for each point in the list until it converges to a mode
(d) The list of all of the modes gives the final fused detections
(e) For each mode compute the bounding box from the final center point and scale

Algorithme 3: Calcul du score de détection [89].

5.4.3.3.3 Injection du degré de confiance

Pour l'injection d'incertitude des observations, dans l'apprentissage des paramètres du RB nous distinguons plusieurs cas selon la réponse générée par les capteurs de vision.

1. Absence de réponse du capteur de vision : Pas de détection de l'objet cible (problème d'occultation, ré-identification, confusion...). Dans ce cas automati-

quement toutes les observations relatives à cet objet seront considérées comme données manquantes.

2. Absence de l'objet cible (pas de véhicule sur la route à l'instant 't1') dans ce cas, toutes les observations relatives au véhicule seront considérées comme manquantes et la prise de décision sera réalisé, indépendamment des valeurs des autres variables.
3. Observation ayant un degré de confiance négatif ($c < 0$). Cette observation sera considérée comme manquante.

   Toute observation quantitativement manquante sera introduit dans le RB avec des degrés de confiances égaux à toutes les valeurs possibles de cette observation. Le fait d'utiliser des degrés de confiances égaux traduit bien la notion de non décision du processus de vision. Il s'agit de la même probabilité pour l'observation d'avoir les valeurs $v1$, $v2$, ... Par conséquent, cette observation aura une influence "neutre" dans la prise de décision ce qui est mieux que de la considérer comme non existante.
4. Observation ayant un degré de confiance positif pour une valeur bien déterminée. Dans ce cas, l'observation sera introduite dans le RB que ce soit pour la phase d'apprentissage ou de test avec le degré de confiance associé à cette valeur D_{cy1} et les autres $'n'$ valeurs possibles auront par conséquent, un degré de confiance $'D'_{cyi}$ complémentaire à celui généré par le système, soit $\sum_i sD_{cyi} = 1$, alors

$$D_{cyi} = \frac{1 - D_{cy1}}{n}; i = 2...n$$

   Prenons le cas d'une observation ayant deux valeurs possibles (existe : oui ou non). Si le système génère que l'observation existe avec une confiance = 0.8, alors l'observation sera injectée dans le RB avec 0.8 comme degré de confiance de la valeur 'oui' et 0.2 comme degré de confiance de la valeur 'non'.

   => Le fait d'associer un degré de confiances d'incertitude aux valeurs des observations permet de prendre en considération l'incertitude des capteurs de vision et de la propager dans la prise de décision du RB. Cette méthode permet de dépasser les réponses butoires (binaires) utilisées classiquement c'est-à-dire si la valeur 'oui' =1 alors la 'non' = 0, ce qui n'est pas toujours le cas, surtout pour les données générées par les processus de vision qui ont toujours un taux d'incertitude de la réponse trouvée.

Pour l'injection de ces degrés de confiance, que se soit lors de la phase d'apprentissage des paramètres ou de test, l'utilisation d'un calcul classique de probabilité conditionnelle ne peut pas traiter les données recueillies par la couche de vision incertaine. Pour cela, pour apprendre les paramètres, nous utilisons la règle de Jeffrey [222].

En considérant un RB avec un ensemble de variables X modélisant un domaine particulier. N définit une distribution commune $P(X)$. En donnant $Q(Y)$, une observation d'une distribution de probabilité sur les variables $Y \subseteq X$, la règle de Jeffrey [222] dit que la distribution de toutes les autres variables sous cette observation devrait être mise à jour (Équation 5.6 [222]).

$$Q(X \setminus Y) = \sum_i P(X \setminus Y \mid y_i) Q(y_i) \tag{5.6}$$

Avec $Q(y_i)$ qui est la probabilité que Y ait la valeur y_i. Dans ce travail nous conciderons que $Q(y_i)$ est égale au degré de confiance associé à l'observation Y à l'instant 't'.

5.5 Évaluation des performances du modèle proposé pour l'estimation des risques

5.5.1 Protocole de test

Une étude expérimentale est menée sur plusieurs séquences vidéos de scénarios de traversée très variés (5 scénarios). Les séquences de test sélectionnées sont des piétons d'âge varié (personnes âgées, jeunes). Les caractéristiques des scénarios sont présentées dans le tableau 5.1. Ces scénarios sont réels de comportements de traversée dangereux et sécurisés.

La première séquence est principalement utilisée pour l'apprentissage des paramètres du RB. Les quatre autres séquences sont utilisées pour tester la fiabilité de la décision générée (détaillée dans la prochaine sous-section).

TABLE 5.1 – Caractéristiques des séquences de test.

	Age	Feu	Couleur de feu	Nombre de piétons	Nombre de voitures	Instant où la voiture se rapproche du piéton
Séquence 1	Âgé	Non	Abs	1	1	Au début
Séquence 2	Jeune	Non	Abs	1	2	Au début
Séquence 3	Jeune	Oui	Rouge	1	1	Au milieu
Séquence 4	Âgé	Non	Abs	2	0	-
Séquence 5	Âgé	Oui	vert	1	1	A la fin

Nous avons préalablement collecté les informations des séquences pour chaque événement élémentaire et estimons les probabilités conditionnelles du réseau de manière supervisée. Le scénario de traversée est divisé en 10 parties égales à partir du moment où le piéton se trouve juste au bord du trottoir jusqu'à la fin de la traversée. Nous avons 10 observations pour chaque événement, avec l'évaluation du comportement correspondant, qui sont utilisées pour apprendre le processus du RB. L'algorithme EM converge après la 4^{eme} itération. Le modèle proposé doit produire à la fois des décisions de comportement et des valeurs des événements. La réponse produite doit prendre en compte certaines observations manquantes et/ou leur incertitude. Les performances du modèle sont comparées aux résultats de la réalité de terrain et illustrées par les images vidéo correspondantes.

Pour l'évaluation des performances de reconnaissance du comportement générées par l'approche probabiliste proposée, nous calculons des facteurs de reconnaissance pour chaque comportement des différentes cibles. Ces facteurs sont les alarmes positives (TPA) et les alarmes négatives (FPA). TPA est la somme des probabilités de vrai comportement des piétons enregistrées pour chaque instant $Ti(i = 1...10)$ lors de la

traversée. FPA est alors le complement de cette probabilité. Ces deux métriques sont calculées tel qu'explicité aux équations 5.7 et 5.8. Ces valeurs sont comparées au type de comportement de la vérité de terrain qui est connue a priori, suite à l'analyse d'expert.

$$TPA = \frac{\sum_{i=1}^{10} P(H_i)}{10} \tag{5.7}$$

$$FPA = \frac{\sum_{i=1}^{10} (1 - P(H_i))}{10} \tag{5.8}$$

Avec $P(H_i)$ qui est la probabilité prédire correctement la valeur de l'agent à hypothèse H_i.

Dans cette partie nous présentons deux modèles de scénarios exploitant notre architecture logique de reconnaissance et d'estimation du comportement à risque. Le premier scénario (S2 et S3) est lié à la reconnaissance du comportement des piétons jeunes alors que le deuxième (S4 et S5) est pour étudier le comportement des piétons âgés.

Les scénarios testés sont ceux pour 2 personnes âgées et 2 jeunes avec et sans feu. Toutes les séquences sont fournies avec des informations représentant le calibrage de la caméra, l'information géométrique de la scène et les annotations temporelles des données et des scénarios observés telles que l'écart de sécurité, l'équipement urbain, l'âge, le type de comportement des piétons, les positions des voitures et des piétons, les vitesses des voitures et des piétons, le temps d'attente, l'indication du feu....

5.5.2 Résultats

5.5.2.1 Traversée d'une rue pour les piétons jeunes

Pour étudier le comportement des piétons jeunes lors de la traversé d'une rue et estimer leurs attitudes face à des situations de danger, nous avons utilisé deux séquences vidéos. La première filme la traversée sur un passage piétons sans feu alors que pour la seconde, le même passage piétons est filmé après avoir été équipé par un feu qui règle la circulation.

— **Séquence S2 :Un piéton jeune sur un passage piéton sans feu :** Au début, le jeune piéton est en situation de sécurité sur le trottoir et une seule voiture loin de lui est présente sur la route. Après quelques secondes, lorsque le piéton procède à la traversée de la rue, une autre voiture se rapproche de lui. Cette action met le piéton en danger. En fin de séquence, la voiture, s'éloigne et le piéton continue à traverser la rue en toute sécurité sans aucun danger, car aucune voiture n'est présente à ce moment-là sur la route.

Le tableau 5.2 montre les résultats de la reconnaissance du scénario modélisé à travers une première séquence de traversée de la rue. Il résume les prédictions des agents à hypothèse H_2 , H_3 , H_4 pris comme indicateurs du conflit du trafic et H^d la décision sur le type de comportement du piéton durant les 10 instants $T_{1...10}$ de la traversée.

Il est à noter ici que l'absence de feu est traduite par des probabilités égales à 50% pour le respect d'aménagement et le non respect. En fait, la probabilité équilibrée reflète

ici l'absence de décision, aucune réponse n'est favorisée. Le tableau représente aussi, à partir de la réalité de terrain, les erreurs de prédiction du type de comportement du piéton (FPA), et le niveau de difficulté (subjectif) de la séquence étudiée. L'algorithme proposé a correctement inféré le comportement des piétons avec un taux de réponses positives de 7.8/10.

Les probabilités des comportements à risques ainsi que sécurisés pour ces 10 instants sont illustrées par les courbes tracées à la figure 5.11.

TABLE 5.2 – Résultats des prédictions générées par l'approche probabiliste proposée pour la séquence S2.

		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
H_2(%)	Vrai	59,2	76,5	82,1	51,5	51,5	0	0	53,6	82,3	87,6
	Faux	40,8	23,5	17,9	48,5	48,5	100	100	46,4	17,7	12,4
H_3(%)	Vrai	50	50	50	50	50	50	50	50	50	50
	Faux	50	50	50	50	50	50	50	50	50	50
H_4		Jeune									
H^d(%)	Sécurisé	72,0	83,0	86,6	67,1	67,1	22,6	23,7	68,4	86,7	90,02
	Risqué	28,0	17,0	13,4	32,9	32,9	77,4	76,3	31,6	13,3	09,98
TPA		7.8/10									
FPA		2.2/10									
Difficulté		5/5									

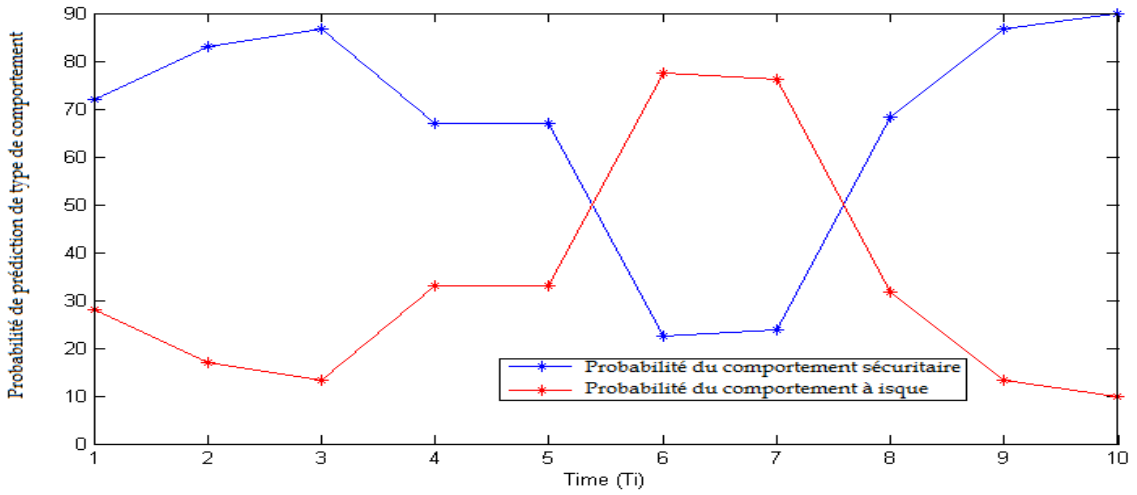


FIGURE 5.11 – Estimation du comportement lors du scénario 2 de traversée d'une rue : cas d'un sujet jeune sur un passage piéton sans feu.

La figure 5.12 représente les indicateurs de la couche basse de reconnaissance des scénarios. Les différentes observations et agents à hypothèses sont illustrés par les annotations textuelles générées par le RB ainsi que les figures qui correspondent aux 10 instants de traversée.

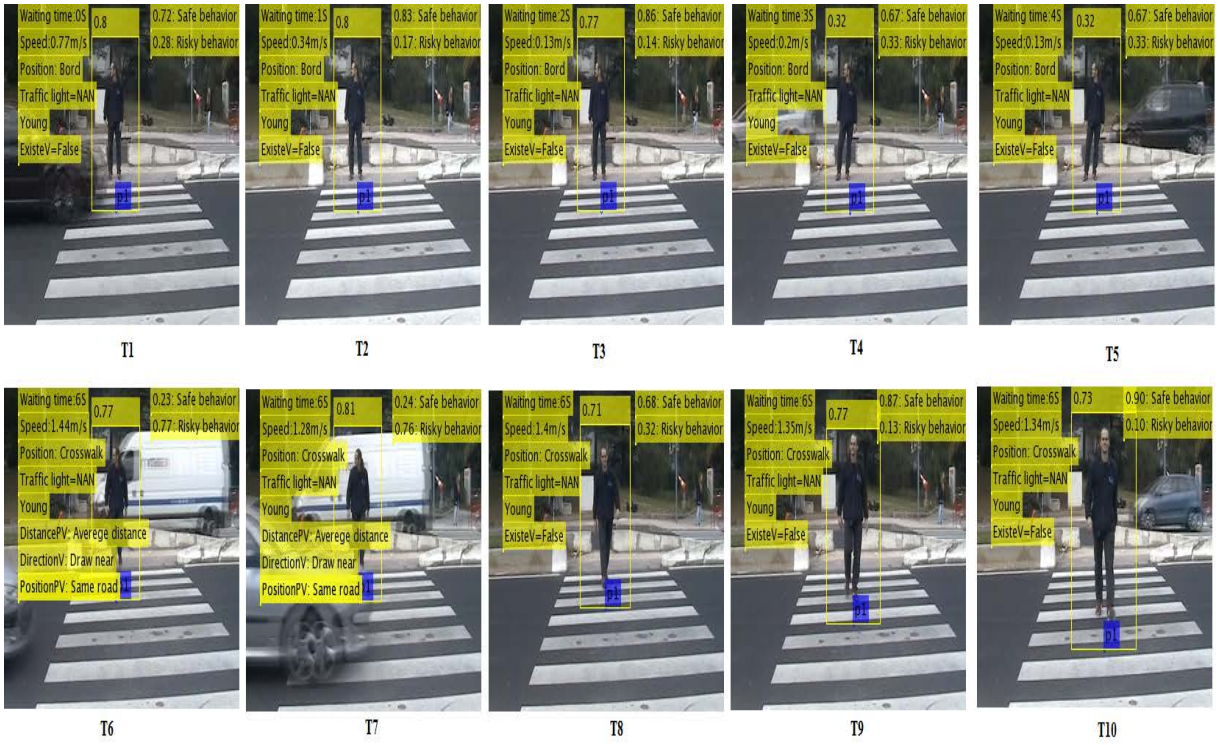


FIGURE 5.12 – Illustration des reconnaissances des événements élémentaires et estimation du type de comportement : cas d'un sujet jeune sur un passage piétons sans feu (S2).

— **Séquence S3 : Un piéton jeune sur un passage piéton avec feu :** Cette séquence illustre un comportement à risque d'un jeune qui ne respecte pas l'indication du feu. Il traverse la rue à une vitesse rapide au feu rouge. Bien qu'au début de la traversée, aucune voiture ne soit pas présente sur la route, à la fin, une voiture se rapproche de lui, ce qui rend la situation plus à risque.

Les résultats de la reconnaissance du scénario modélisés à travers cette séquence sont résumés dans le tableau 5.3 qui montre, les prédictions des agents à hypothèse H_2 , H_3 , H_4 et H^d durant les 10 instants $T_{1...10}$ des traversées. L'erreur de prédiction du type de comportement des piétons (FPA) est également présentée dans le tableau 5.3 pour cette séquence jugée subjectivement de difficulté faible. L'algorithme proposé a correctement inféré le comportement des piétons avec un taux de réponses positives de $7,13/10$.

Les résultats d'estimation du comportement à risque ainsi que sécuritaire pour ces 10 instants sont illustrés à la figure 5.13.

5.5.2.2 Traversée d'une rue pour les piétons âgés

Pour étudier le comportement de piétons âgés lors de la traversé de la rue et estimer leurs attitudes face à des situations de danger, nous avons utilisé deux séquences vidéos. La première filme la traversée sur un passage piétons sans feu alors que la seconde filme le même passage piétons après avoir été équipé par un feu qui règle la circulation.

TABLE 5.3 – Résultats des prédictions générées par notre approche probabiliste pour la séquence S3.

		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
H_2(%)	Vrai	59,2	50	50	50	56,6	23,3	23,3	5,56	50	83,3
	Faux	40,8	50	50	50	43.,4	76,7	76,7	94,4	50	16,7
H_3(%)	Vrai	99,5	83,3	83,3	83,3	50	83,3	83,3	50	83,3	83,3
	Faux	0,5	16,7	16,7	16,7	50	16,7	16,7	50	16,7	16,7
H_4		Jeune									
H^d(%)	Sécurisé	86,2	77,4	77,4	77,4	57,2	65,8	65,8	37,2	77,4	91,9
	Risqué	13,8	22,6	22,6	22,4	42,8	34,2	34,2	62,8	22,6	08,1
TPA		7,13/10									
FPA		2,87/10									
Difficulté		2/5									

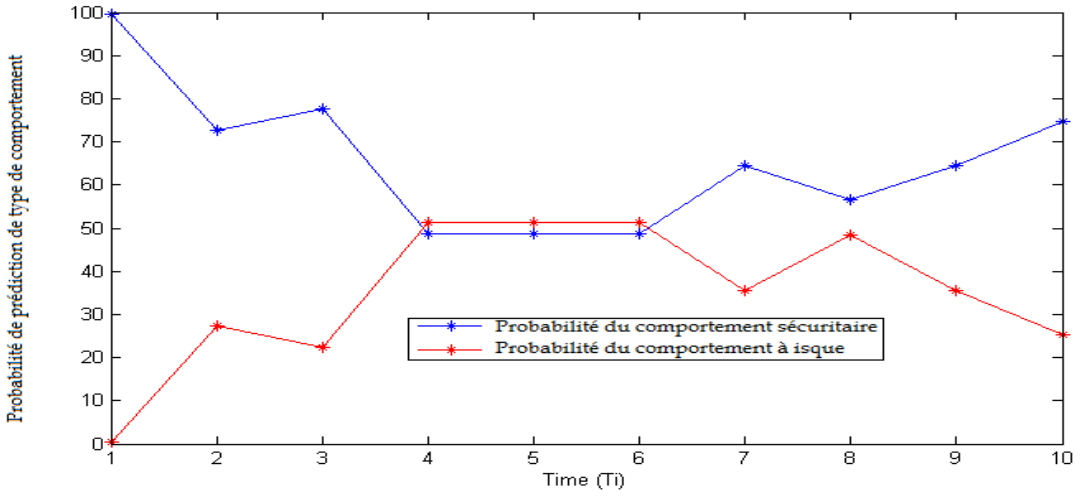


FIGURE 5.13 – Estimation du comportement lors du scénario 3 de traversée d'une rue : cas d'un sujet jeune sur un passage piéton avec feu.

— **Séquence S4 : Deux piétons âgés sur un passage piétons sans feu.** Durant cette séquence les deux piétons âgés marchent très lentement et semblent avoir peur et présentent une démarche hésitante au début de la traversée. L'avancée en âge et un tel attitude placent les piétons en une situation de danger, même si aucune voiture n'est détectée au moment de l'initiation de la traversée. Ensuite, les piétons semblent bien contrôler l'état du trafic routier et continuent la traversée en toute sécurité.

Le tableau 5.4 présente les résultats de la reconnaissance du comportement des pitons durant le scénario modélisant la traversée pour les personnes âgées sur 10 périodes temporelles. Il résume les prédictions des agents à hypothèse H_2 , H_3 , H_4 respectivement, l'existence de gap sécuritaire, le respect d'aménagement urbain (feu), la classe d'âge, pris comme indicateurs de conflit de trafic et H^d la décision. Le tableau présente également à partir de la réalité de terrain, les erreurs de prédiction du type de compor-

tement des piétons (FPA), et le niveau de difficulté (subjectif) de la séquence étudiée. L'algorithme proposé a correctement inféré le comportement des piétons avec un taux de réponses positives de $8,5/10$ pour ce scénario, estimé moyennement difficile.

Les probabilités de comportements à risques ainsi que sécurisés pour ces 10 instants sont illustrées par les courbes tracées à la figure 5.14.

TABLE 5.4 – Résultats des prédictions générées par l'approche probabiliste pour la séquence S4.

		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
$H_2(\%)$	Vrai	42,8	44,2	45,1	43,4	100	100	100	100	100	100
	Faux	57,2	55,8	54,9	56,6	0	0	0	0	0	0
$H_3(\%)$	Vrai	50	50	50	50	50	50	50	50	50	50
	Faux	50	50	50	50	50	50	50	50	50	50
H_4		Âgé									
$H^d(\%)$	Sécurisé	61,2	62,1	62,7	61,6	100	100	100	100	100	100
	Risqué	38,8	37,9	37,3	38,4	0	0	0	0	0	0
TPA		$8,5/10$									
FPA		$1,5/10$									
Difficulté		$3/5$									

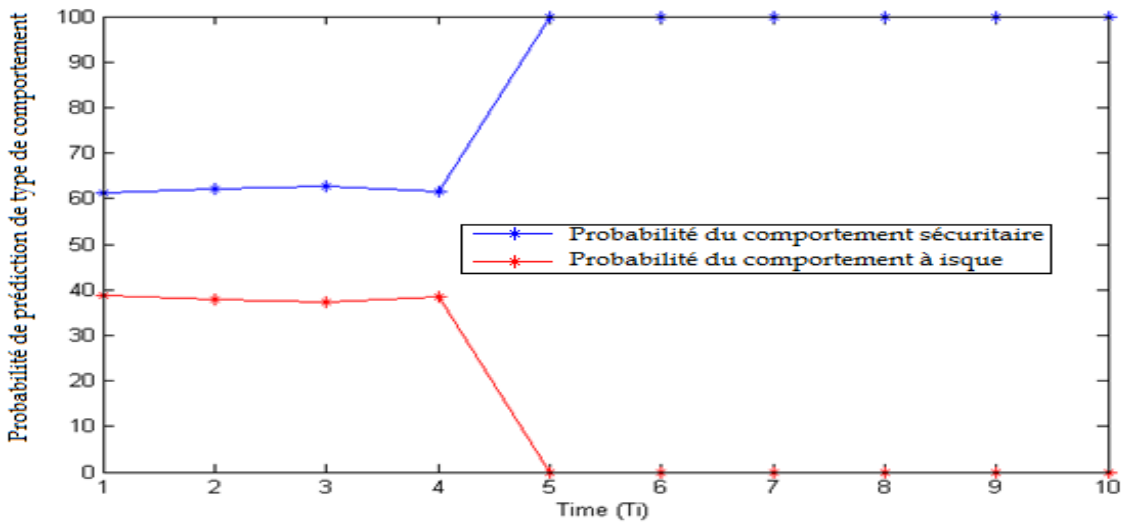


FIGURE 5.14 – Estimation du comportement lors du scénario 4 de traversée d'une rue : cas d'un sujet âgé sur un passage piéton sans feu.

— **Séquence S5 : Un piéton âgé sur un passage piéton avec feu :** Cette séquence illustre le comportement d'un piéton âgé face à un aménagement urbain tel qu'un feu sur la sécurité du piéton. En effet, au feu vert le piéton procède à la traversée de la rue en toute sécurité jusqu'à l'arrivée d'une voiture qui se rapproche de lui mais qui s'est arrêté, juste avant le passage piétons car le feu est rouge pour lui. Les piétons et les conducteurs respectent l'indication de feux et le piéton effectue une traversée sécurisée.

Le tableau 5.5 présente les résultats durant les différentes étapes de reconnaissance du scénario de traversée pour un sujet âgé sur un passage piéton équipé d'un feu. La figure 5.15 illustre l'évolution des probabilités du comportement sécuritaire et à risque sur les 10 périodes de traversée.

TABLE 5.5 – Résultats des prédictions générées par l'approche probabiliste pour la séquence S5.

		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
$H_2(\%)$	Vrai	87,0	69,8	75,3	38,7	38,7	38,7	62,1	52,9	62,1	67,8
	Faux	13	30,2	24,7	61,3	61,3	61,3	37,9	47,1	37,9	32,2
$H_3(\%)$	Vrai	57,2	25,0	26,7	26,7	26,7	26,7	16,7	16,7	16,7	40
	Faux	42,8	75,0	73,3	73,3	73,3	73,3	83,3	83,3	83,3	60
H_4		Âgé									
$H^d(\%)$	Sécurisé	99,7	72,8	77,6	48,7	48,7	48,7	64,4	56,6	64,4	74,8
	Risqué	0,33	27,2	22,4	51,3	51,3	51,3	35,6	48,4	35,6	25,2
TPA		6,5/10									
FPA		3,5/10									
Difficulté		4/5									

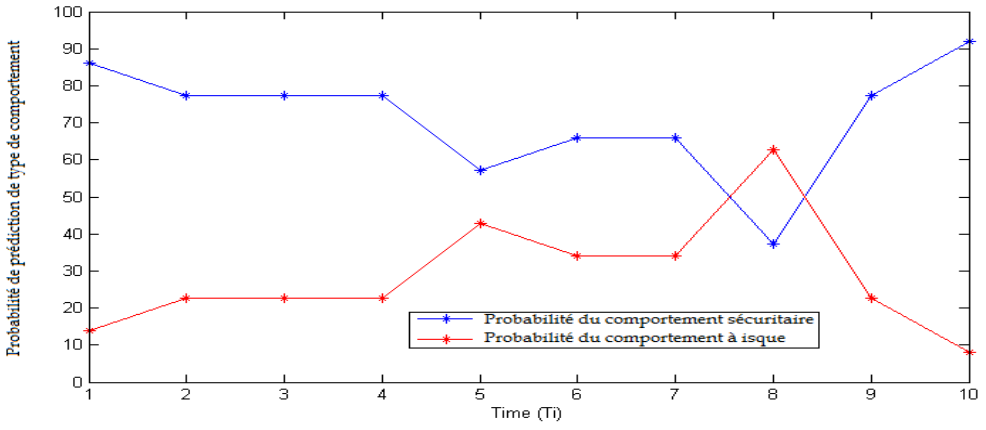


FIGURE 5.15 – Estimation du comportement lors du scénario 5 de traversée d'une rue : cas d'un sujet âgé sur un passage piéton avec feu.

5.5.2.3 Discussion

Étant donnée que, les données sont collectées par des capteurs de vision, elles ne sont pas aussi précises et elles sont partiellement bruitées. Elles peuvent être manquées et non identifiées pendant un certain temps en raison des limites des méthodes de détection et de suivi. En effet, cette incertitude associée aux données de vision doit être modélisée et injectée dans la couche de décision pour une reconnaissance plus fiable du comportement du piéton. L'introduction du degré de confiance associée aux capteurs de vision à base de règle de Jeffry permet au système d'intégrer naturellement certaines des limitations des détecteurs sans réduire la quantité de données d'apprentissage. Cette technique augmente considérablement la qualité de la décision. Nous notons que les comportements des cibles ont été correctement annotés avec des remarques textuelles durant les différentes instants 't_i' pendant la séquence vidéo. En effet, la figure 5.11 représente à $T6$ et $T7$ une interaction dangereuse entre un piéton et un véhicule. Les cibles sont proches les unes des autres et leur vitesse relative est élevée. Le système reconnaît correctement que ce comportement peut être dangereux. Cependant, à l'instant $T1$, même si la voiture est sur la route avec une vitesse moyenne, le comportement de piéton est reconnu comme un comportement sécurisé car il est encore au bord du passage. De plus, tous ces comportements ont été correctement reconnus. Ainsi, nous notons une bonne coïncidence entre la description textuelle des scénarios testés et la description enregistrée pour les observations, les événements et le comportement en utilisant le modèle d'analyse du comportement proposé.

Les systèmes d'analyse vidéo se développent pour l'analyse du comportement des piétons notament, à condition de pouvoir calculer automatiquement les indicateurs de conflit et de signaler les interactions importantes à un observateur humain pour un examen plus approfondi des interactions entre le trafic et le piéton lors de la traversé de la rue. La qualité de ces indicateurs de conflit tels que le temps de collision, Post Encroachment Time, la marge de sécurité et décélération de temps de sécurité [194], ou le temps d'attente et la vitesse de traversée [64], ont été proposés et évalués pour leur capacité à représenter la gravité des conflits du trafic [68], mais aucun des indicateurs de conflit n'est individuellement capables de représenter toutes les interactions dangereuses entre les usagers de la route [68]. De plus, tous les indicateurs de conflit estimés sont utilisés comme données 100% sûres et complètes pour analyser le comportement des piétons sans tenir compte des lacunes des capteurs de vision. Par conséquent, cela est considéré comme une limitation majeurs de ces travaux et illustre notre contribution dans cette thèse.

Nos contributions pertinentes consistent à utiliser le RB pour concevoir un outil d'analyse du comportement qui combine plusieurs indicateurs de conflit et établit avec succès une relation de causalité entre eux pour qualifier des événements comportementaux. En outre, le modèle proposé tient compte des observations incertaines des capteurs de vision et introduit un nouveau degré de confiance associé aux performances des capteurs de vision et la consistance temporelle des observations pendant une période de temps (1 seconde). Ce coefficient d'incertitude proposé, prend en considération les contraintes de type qualitative ainsi que quantitative des capteurs de vision. Les obser-

vations de la couche basse sont qualitativement caractérisées par le score de détection, qui mesure la précision de la détection de l'objet cible. Pour ajouter des capacités de type quantitatif tels que la durée ou la date d'une telle observation, nous avons utilisé la consistance temporelle de cette observation durant une seconde et les enregistré durant 10 instants différents de la traversée. Ces degrés de confiance sont ensuite, injectés au niveau décisionnel en utilisant l'apprentissage des nouveaux paramètres RB selon les règles de Jeffry.

On se référant aux résultats expérimentaux préliminaires présentés par les tableaux 5.2 à 5.5, il semble se dégager au regard de ces observations ponctuelles une tendance à prendre plus de risque par les sujets jeunes versus les sujets âgés, même si ces derniers sont plus exposés aux risques surtout sans le feu. La différence d'âge pourrait expliquer cette différence de comportement chez les sujets âgés et les sujets jeunes.

En analysant des scénarios typiques de traversée d'une rue intégrants des sujets jeunes et d'autres âgés on peut noter également que les jeunes violent plus facilement les règles de circulation que les âgés (cas de la séquence S4). Par contre, même si les âgés semblent être plus prudents, ils sont plus susceptibles de ne pas faire attention aux feux de circulation ou d'éviter de traverser la route aux endroits désignés. Une fréquence plus élevée d'accidents chez les personnes âgées pourrait, aussi être due au fait que les règles de circulation semblent plus assimilées chez les jeunes que chez les âgés. En fait, les jeunes peuvent alors facilement manipuler les marges de sécurité et les exploiter pour faire une traversée sécurisée même au feu rouge.

5.6 Conclusion

Durant ce chapitre, nous avons proposé et évalué une approche probabiliste basée sur la juxtaposition des fonctionnalités de la vision par ordinateur et celles des réseaux bayésiens. Cette approche a comme objectif la mise en relation des différents indicateurs de conflits du trafic possible entre un piéton et un véhicule en mouvement sur la route tenant compte des aménagements urbains (feu) et l'effet d'age en particulier.

Une originalité de cette partie concerne l'intégration des imperfections des modules de traitement d'images en les considérant comme des données manquantes lors du processus d'apprentissage supervisé en cas d'absence de réponse générée par les capteurs de vision (cas d'occultation totale). En outre, nous avons proposé en particulier, le coefficient de degré de confiance associé aux observations de la couche de vision (la couche basse) exploitant le lien entre la précision de détection (score de détection) des objets cible (piétons, véhicules, classe d'age) et la consistance temporelle de ces observation durant un petit laps de temps (1 seconde). Nous obtenons au final une bonne concordance entre les observations de terrain et l'analyse automatique réalisée par notre système.

Conclusion et perspectives

La sécurité routière et l'analyse comportementale sont les deux thèmes principaux de cette thèse. Elles sont également deux piliers principaux d'un système de transport fonctionnel. Traditionnellement, la performance d'un système de transport est mesurée en termes de mobilité. Cependant, la mobilité conduit souvent à la réduction de la sécurité des usagers de la route non motorisés [227]. Des études récentes ont montré que le coût des collisions routières dépasse le coût de la congestion du trafic de près de dix fois [227]. A l'inverse, il existe une demande croissante pour la construction d'un système de transport sécurisé. L'accent mis sur la sécurité donne une impulsion à l'accidentologie en général et à l'étude du comportement à risque des piétons en particulier, notamment lors de la traversée d'une rue et révèle des facteurs de risques.

L'analyse du comportement des piétons lors de la traversée d'une rue, est une étape clé pour la construction d'un système de transport durable, sécurisé et favorisant les déplacements doux (la marche). Elle est d'une importance grandissant, pourtant les méthodes d'analyse actuelles et les données disponibles liées à cette thématique souffrent encore de limites dans la collecte des données et leur analyse. L'analyse de l'existant a montré que la majorité des travaux effectués dans ce contexte sont soit manuels, soit semi-automatiques basés sur des observations in situ, des statistiques, des simulateurs ou encore de la réalité projetée et/ou des "vidéos tapping".

Se démarquant de l'existant, nous proposons une nouvelle architecture de système d'analyse du comportement des piétons lors de la traversée de la rue. L'approche proposée se base sur le traitement automatique d' images de caméras HD grand public et exploite les techniques et les avancées de la vision par ordinateur. Le système proposé est centré sur un certain nombre de propositions et d'applications de techniques de vision par ordinateur à des fins de collecte de données du trafic, des piétons et de l'analyse automatisée de niveau de sécurité de la traversée. Les capteurs vidéo ont été utilisés comme principale source de données dans ce système. Le principal avantage des techniques de vision par ordinateur est le potentiel de collecte des données microscopiques

des usagers de la route et du trafic à un degré d'automatisation et avec une précision qui ne peut être réalisable par des techniques manuelles ou semi-automatique et surtout de générer des indicateurs de conflit du trafic objectifs, sans l'intervention subjective de l'être humain. Dans ce contexte nous avons proposé en premier lieux 3 sous systèmes de collecte de données. Le premier sous-système est conçu pour la collecte des paramètres piétonniers tels que la vitesse instantanée et moyenne, les frontière des phases de traversée (rapprochement, attente, traversée), la direction, les positions spatio-temporelles, les trajectoires et le temps d'attente. Le deuxième sous-système consiste à estimer la classe d'âge du piéton. Ce système est basé sur l'analyse de la marche et de la posture de la personne, puisque nous sommes dans un contexte extérieure où le piéton est à une distance raisonnablement grande par rapport à la caméra rendant l'étude de son visage quasi-impossible. Dans ce cadre, nous avons proposé et évalué un nouveau descripteur basé sur le modèle de la silhouette. Ce descripteur a montré des performances compétitivement comparables par rapport à l'existant. Le troisième sous-système est destiné à la collecte automatique des paramètres du trafic routier gauche et droit. Le système se base sur une proposition d'un HOG optimisé pour la détection des véhicules en mouvement sur la route. Par la suite, des paramètres tels que la vitesse, la direction ainsi que la trajectoire des véhicules sont déterminés.

Cette thèse ouvre la voie d'une exploitation combinée de la vision par ordinateur et de l'intelligence artificielle par la proposition d'une approche probabiliste multi-couches pour l'analyse du comportement des piétons lors de la traversée d'une rue. Cette approche met en relation les paramètres (piétons et trafic) déjà collectés et génère des agents à hypothèses pour prédire finalement le type de comportement du piéton durant les différents instant de sa traversée. Une telle architecture se base sur l'ordonnancement hiérarchisé d'un RB et met en œuvre les relations causales entre les variables d'environnement et le comportement du piéton. L'utilisation d'un RB dans ce contexte était favorable vue la structure intuitive et explicite du RB qui modélise les relations cause-à-effet d'une part et la grande capacité à jumeler et combiner des données hétérogènes, de nature différente, voir même manquantes.

La couche basse, pour la collecte des données est fondée sur les capteurs de vision. Par ailleurs les observations et les événements élémentaires obtenus sont de nature incertaine et imprécise à cause des limitations des techniques de détection et de suivi. Pour palier à cette limitation, nous proposons ici une nouvelle métrique de calcul d'un coefficient de degré de confiance associé à ces observations. Ce coefficient est calculé en fonction de la précision qualitative de détection de l'objet cible, dit score et la consistance temporelle de cette même observation durant un petit laps de temps (1 seconde). Ce degré de confiance est ensuite intégré dans la couche haute, que ce soit lors de la phase d'apprentissage ou d'inférence à partir de la règle de 'Jeffry'.

Une telle architecture se base sur l'ordonnancement hiérarchisé de RB et met en évidence les relations causales entre les variables d'environnement et le comportement du piéton. A différente étapes du processus de traitement, les résultats automatiques ont été comparés aux résultats mesurés sur le terrain sur des essais soigneusement choisies pour leur diversité voire leurs difficultés. De même, une comparaison à des outils

de la littérature a également été réalisée. Les résultats montrent un bon niveau de performances aux différentes étapes testées.

Si certains résultats plus particulièrement centrés sur les techniques développées ont déjà donné lieu à la rédaction d'articles, publiés ou soumis à des conférences et/ou des revues internationales, le recueil de données permis par ce type d'expérimentation reste très important. En conséquence, l'exploitation du travail réalisée dans le cade de cette thèse nécessite d'être complété. Les prolongements les plus directs consistent à poursuivre la publication des résultats, notamment dans des revues de vision par ordinateur appliquée. C'est le cas notamment de la dernière partie de la thèse, la possibilité d'une analyse du comportement du piéton automatique et à grande échelle à partir d'une approche probabiliste basé sur le RB et intégrant la gestion des incertitude des capteurs de vision.

Des améliorations aux niveaux de la couche basse sont également envisageables pour augmenter les performances des observations générées par les capteurs de vision. En effet, pour la phase de détection des personnes, l'HOG à plusieurs échelles tel qu'il est présenté par [228] peut être utile puisqu'il se base sur une décomposition pyramidale permettant de générer, à partir de l'image initiale, des images à différentes résolutions conservant l'orientation de gradient. Cette technique semble être efficace aussi dans le contexte de détection des véhicules. La représentation pyramidale de l'HOG permet de dépasser le problème de variation d'échelle et d'assurer la détection des véhicules à une grande distance par rapport à la caméra, ce qui permet par ailleurs d'activer le système d'extraction des paramètres du trafic plus tôt et par conséquence d'augmenter les marges de sécurité pour l'inférence du danger potentiel menaçant le piéton et gagner en temps de calcul.

Ces considérations guident le choix des suites qu'il conviendrait de donner à ce travail. Il s'agit de poursuivre l'exploitation de l'approche probabiliste d'analyse comportementale à deux niveaux :
i) Pour traquer au plus près les causes de mise en danger des usagers de la route les plus vénérables tels que les piétons et ceux âgés en particulier. L'approche expérimentale in situ des traversées garde un potentiel considérable de mise en relations des indicateurs de conflit du trafic déjà utilisés. L'implication d'autres indicateurs et agents à hypothèses caractérisant davantage les détails d'un conflit véhicule-piéton peuvent également être un axe d'amélioration envisageable. Il en est de même concernant l'étude d'effet d'aménagements urbains tel qu'initier avec l'exemple du feu comme les carrefours et les croisements, sur le comportement des piétons car il représentent un danger supplémentaire.
ii) Pour la conception et l'évaluation d'un outil facilement accessible, permettant une approche écologique de l'analyse du comportement du piéton âgé lors de la traversée de rue. Pour cela, l'approche réalisée à partir de l'analyse des séquences vidéo issues de trois cameras du commerce nécessite d'être tester sur un recueil important de traversées de rues. Le dispositif à 3 caméras doit être implémenté dans différentes régions pour filmer en contenu durant plusieurs heures tous les piétons avec plusieurs

situations de traversée sécurisées et dangereuses. L'analyse de ces recueils à partir de l'approche multi-couches développée dans le cadre de cette thèse permettra l'étude comportementale des piétons lors de la traversée de rues et produire des conclusions et des règles comportementales permettant de dégager les éléments à risques et évaluer objectivement l'effet des aménagements.

Bibliographie

[1] R. Thomas. *Marcher en ville : Faire corps, prendre corps, donner corps aux ambiances urbaines.* éditions des archives contemporaines, 2010, p. 196.

[2] *Groupe de travail du CCATM sur les usagers vulnerables de la route : groupe d'experts sur les piétons, "Mesures de prevention pour assurer la sécurité des piétons au Canada".* Le Conseil canadien des administrateurs en transport motorisé (CCATM), 2013.

[3] F. Ferrandez. *L'étude detaillée d'accidents orientée vers la sécurité primaire.* Presses ENPC, pp. 88-89, 1995.

[4] D. Fleury. *Sécurité et urbanisme - La prise en compte de la sécurité routiére dans l'amenagement urbain.* Presses ENPC, 1998.

[5] H. de Silva. *Why we have automobile accidents.* Wiley ch. Pedestrians, pp. 220-239, 1951.

[6] *Organisation mondiale de la Santé, "Sécurité des piétons : manuel de sécurité routière pour les décideur et les intervenants".* Catalogage a la source : Bibliotheque de l'OMS, 2013.

[7] A. Julien et J. Carré. « Cheminements piétonniers et exposition au risque ». In : *Recherche Transports Sécurité* 76 (2002 2002), p. 173–189.

[8] OCDE. *La mobilite des Francais, panorama issu de l'enquete nationale transports et deplacements 2008.* Service de l'Observation et des Statistiques (SOeS) du Commissariat General au Developpement Durable (CGDD), 2010.

[9] *SUD INSEE L'essentiel, "Marseille, Nice et Toulon : 530.000 piétons chaque jour".* INSEE, 2002.

[10] J. M. Lilley, T. Arie et C. E. D. Chilvers. « Accidents involving older people : A review of the literature ». In : *Age and Ageing* 24 (4 1995), p. 346–365.

[11] C. Zegeer et M. Bushell. « Pedestrian crash trends and potential countermeasures from around the world ». In : *Accident Analysis and Prevention* 44 (1 2012), p. 3–11.

[12] *Groupe de travail intersecretariats sur les statistiques de transports (IWG), "Glossaire des statistiques de transport. Transmis par Eurostat, Organisation des Nations Unies-Commission Economique pour l'Europe - Comite des transports interieurs"*. 2001. URL : `http%20://www.unece.org/trans/main/wp6/pdfdocs/2001-8f.pdf`.

[13] C. F. WAKIM. « Etude de la prédiction de chocs véhicule-piéton ». Thèse de doctorat, Université Paris Sud-Paris XI, 2005.

[14] C. Mathers et D. Loncar. *Updated Projections of Global Mortality and Burden of Disease, 2002-2030.* Methods et Results.', World Health Organization, 2002.

[15] A. Jain, A. Gupta et R. Rastogi. « Pedestrain Crossing Behavior Analysis at Intersections ». In : *International Journal for Traffic and Transport Engineering* 4 (1 2014), p. 103–116.

[16] *Organisation mondiale de la santé, "Rapport de Situation de la Sécurité Routière dans le Monde 2013"*. Decennie d'action pour la sécurité routière 2011-2020, 2013.

[17] « L'observatoire national interministeriel de la sécurité routière, "L'accidentalité routière en 2015 Bilan sommaire" ». 2015.

[18] « L'observatoire national interministeriel de la sécurité routière, "L'accidentalité routière en 2016 Bilan sommaire" ». 2016.

[19] *Groupe de travail intersecretariat de la Commission economique pour l'Europe sur les statistiques du transport, "Glossaire des statistiques de transport"*. New York, Nations Unies, 2003.

[20] J. RFS. « Overcoming barriers to pedestrian safety ». In : *Proceedings of the Australasian College of Road Safety National Conference : A Safe Systems Approach : Expanding the Reach* (2012), p. 1–8.

[21] G. Richelet. *Dictionnaire francois.* A Geneve, Chez Jean Herman Widerhold, 1680.

[22] *Institue National de la Statistique et des etudes economiques.* URL : `www.insee.fr`.

[23] *Commissariat General au Developpement Durable, "La Revue du CGDD, La mobilite des francais, Panorama issu de l'enquete nationale Transports et deplacements 2008"*. CGDD, 2010.

[24] G. Kemoun et al. « Evaluation de la marche chez le sujet âgé ». In : *La marche humaine et sa pathologie* (1994), p. 322–328.

[25] D. Strubel. « Perte de l'autonomie a la marche ». In : *La marche humaine et sa pathologie* (1994), p. 313–322.

[26] N. Ward. « Compensatory mechanisms in the aging motor system ». In : *Ageing Research Reviews* 5 (3 2006), p. 239–254.

[27] J. Delzenne. « La traversée de rue chez le piéton sénior : Conception d'un simulateur, étude biomécanique et comportementale ». Thèse de doctorat, Université de Valenciennes et du Hainaut-Cambresis, 2013.

[28] *Association Transports et Environnement, Les aines : les oublies de la circulation, "Diagnostic sécurité aines-Quartiers des Eaux-vives Ville de Geneve"*. 2008.

[29] J. Oxley et al. « Differences in traffic judgements between young and old adult pedestrians ». In : *Accident Analysis et Prevention* 29 (6 1997), p. 839–847.

[30] S. U. Jensen. « Pedestrian safety in Denmark ». In : *Transportation Research Record : Journal of the Transportation Research Board* 1674 (1 1999), p. 61–69.

[31] R. C. Harruff, A. Avery et A. S. A. Pandya. « Analysis of circumstances and injuries in 217 pedestrian traffic fatalities ». In : *Accident Analysis et Prevention* 30 (1 1998), p. 11–20.

[32] T. Koepsell et al. « Crosswalk markings and the risk of pedestrian motor vehicle collisions in older pedestrians ». In : *The Journal of the American Medical Association* 288 (17 1998), p. 2136–2143.

[33] G. Dunbar, C. A. Holland et E. A. Maylor. *LOlder pedestrians : A critical review of the literature-Road Safety Research Report No. 37*. Department for Transport, London, 2004.

[34] I. Ajzen. « The theory of planned behavior ». In : *Organizational Behavior and Human Decision Processes* 50 (2 1991), p. 179–211.

[35] D. Evans et P. Norman. « Understanding pedestrians' road crossing decisions : An application of the theory of planned behaviour ». In : *Health Education Research* 13 (4 1998), p. 481–489.

[36] C. A. Holland et P. M. A. Rabbitt. « People's awareness of their age-related sensory and cognitive deficit and the implications for road safety ». In : *Applied Cognitive Psychology* 6 (3 2007), p. 217–231.

[37] E. Avineria, D. Shinarb et Y. O. Susiloa. « Pedestrians' behaviour in cross walks : The effects of fear of falling and age ». In : *Accident Analysis and Prevention* 44 (1 2012), p. 30–34.

[38] M. S. Cloutier, M. Tremblay et P. Morency. « Amenager les intersections en milieu urbain : Quelle place pour le piéton : Qualité et sécurité du deplacement piéton : Facteurs, enjeux et nouvelles actions ». In : *Colloque International du GERI COPIE 2011* (2011).

[39] S. J. Lee et J. Hidler. « Biomechanics of overground vs treadmill walking in healthy individuals ». In : *Journal of Applied Physiology* 104 (3 2008), p. 747–755.

[40] D. Akin et V. Sisiopiku. « Modeling Interactions Between Pedestrians and Turning Vehicles at Signalized Crosswalks Operating Under Combined Pedestrian-Vehicle Interval ». In : *Transportation Research Board Annual Meeting, Washington, DC.* (2008).

[41] B. Randles et al. « Modeling Interactions Between Pedestrians and Turning Vehicles at Signalized Crosswalks Operating Under Combined Pedestrian-Vehicle Interval ». In : *SAE ICE Detroit Michigan* (2001), p. 9.

[42] D. Silcock, R. Walker et T. Selby. « Pedestrians at risk ». In : *PTRC proceedings of European transport conference* (1998), p. 209–219.

[43] W. Daamen et S. P. Hoogendoorn. « Experimental Research of Pedestrian Walking Behaviour ». In : *Transportation Research Record* 1828 (03-3113 2003).

[44] l. Broggi, P. Cerri et H. G. J. S. Ghidoni andP. Grisleri. « A New Approach to Urban Pedestrian Detection for Automatic Braking ». In : *IEEE Transactions on Intelligent Transportation Systems* 10 (43 2009).

[45] C. Musselwhite et H. Haddad. « Strategic Promotion of Ageing Research Capacity Prolonging Safe Driver Behaviour Through Technology : attitudes of older drivers ». In : *Meeting the challenges of an ageing society* (2008).

[46] S. Schmidt et B. Férber. « strians at the kerb - recognising the action intentions of humans ». In : *Transportation Research Part F Traffic Psychology Behavior* 12 (2009), p. 300–310.

[47] R. Lobjois et al. « Les effets du vieillissement sur les comportements de traversée de rue ». In : *Congrés International de l'Association des Chercheurs en Activités Physiques et Sportives* (2009), p. 405–406.

[48] R. Lobjois et V. Cavallo. « e-related differences in street-crossing decisions : The effects of vehicle speed and time constraints on gap selection in an estimation task ». In : *Accident Analysis and Prevention* 39 (5 2007), p. 934–943.

[49] R. Lobjois et V. Cavallo. « The effects of aging on street-crossing behavior : From estimation to actual crossing ». In : *Accident Analysis and Prevention* 41 (2 2009), p. 259–267.

[50] R. Lobjois, N. Benguigui et V. Cavallo. « The effects of age and traffic density on street-crossing behavior ». In : *Accident Analysis and Prevention* 53 (1 2013), p. 166–175.

[51] A. Dommes et V. Cavallo. « The role of perceptual, cognitive, and motor abilities in street-crossing decisions of young and older pedestrians ». In : *Ophthalmic and Physiological Optics* 31 (3 2011), p. 292–301.

[52] A. Dommes et V. Cavallo. « Can simulator-based training improve street-crossing safety for elderly pedestrians ». In : *Transportation Research Part F : Traffic Psychology and Behaviour* 15 (2 2012), p. 206–218.

[53] A. Dommes et al. « Age-related differences in street-crossing safety before and after training of older pedestrians ». In : *Accident Analysis and Prevention* 44 (1 2012), p. 42–47.

[54] K. A.Ismail. « Application of computer vision techniques for automated Road safety analysis and traffic data collection ». Phd, thesis, The university of british columbia, 2010.

[55] G. Reymond et al. « Validation of Renault's dynamic simulator for Adaptative Cruise Control experiments ». In : *Driving Simulation Conference* (2000).

[56] *Honda Worldwide, "Honda develops world's first intelligent night vision system able to detect pedestrians and provide driver cautions"*. 2004. URL : `http%20://world.honda.com/news/2004/4040824_01.html`.

[57] Y. Nishimura et N. Okamoto. « A study on analysis of road image including a pedestrian crossing ». In : *IEEE Canadian Conference on Electrical and Computer Engineering* (2002), p. 934–939.

[58] S. Kohler et al. « Early detection of the Pedestrian's intention to cross the street ». In : *IEEE International Conference on Intelligent Transportation Systems (ITSC)* (2012), p. 1759–1764.

[59] C. Motamed. « Motion Detection and Tracking Using Belief Indicators for an Automatic Visual-Surveillance System ». In : *Image and Vision Computing Journal* 24 (11 2006), p. 1192–1201.

[60] L. Simon, S. Tarek et H. Z. Mohamed. « Automated Collection of Pedestrian Data Through Computer Vision Techniques ». In : *TRB Transportation Research Record : Journal of the Transportation Research Board* 2299 (1 2012), p. 121–127.

[61] M. Tarawneh. « Evaluation of Pedestrian Speed in Jordan with Investigation of Some Contributing Factors ». In : *Journal of Safety Research* 32 (2 2001), p. 229–236.

[62] A.Svensson. « A Method for Analyzing the Traffic Process in a Safety Perspective ». Phd thesis, University of Lund, 1998.

[63] C. Hyden. « The Development of a Method for Traffic Safety Evaluation : The Swedish Traffic Conflicts Technique ». Phd thesis, Department of Traffic Planning et Engineering, Lund University, Lund, Sweden, 1987.

[64] J. Hayward. « Near-miss Determination through Use of a Scale of Danger ». In : *Highway Research Record* (1968), p. 24–34.

[65] M. Minderhoud et P. Bovy. « Extended Time-to-collision Measures for Road Traffic Safety Assessment ». In : *Accident Analysis and Prevention* 33 (1 1968), p. 89–97.

[66] A. Varhelyi. « Dynamic Speed Adaptation Based on Information Technology-A Theoretical Background ». Phd thesis, Lund University, Lund, Sweden, 1996.

[67] B. Allen, B. Shin et P. Cooper. « Analysis of Traffic Conflicts and Collisions ». In : *Transportation Research Record : Journal of the Transportation Research Board* 667 (1978), p. 67–74.

[68] J. Archer. « Methods for the Assessment and Prediction of Traffic Safety at Urban Intersections and their Application in Micro-simulation Modelling ». Phd thesis, Royal Institute of Technology, 2004.

[69] H. topp. « Traffic Safety Work with Video-Processing ». Phd thesis, University Kaiserslautern, 1998.

[70] C. Wakim. « Procède et systéme embarqué de prédiction de choc entre un véhicule et un piéton ». In : *Brevet déposé a l'INPI (no : 05-06504)* (2005).

[71] C. Wakim. « Procède et systéme anti-collision adaptatif pour véhicule automobile ». In : *Brevet déposé a l'INPI (no : 05-06505)* (2005).

[72] R. Perkins et J. Harris. « Traffic Conflict Characteristics : Accident Potential at Intersections ». In : *Highway Research Record* 225 (1968), p. 35–43.

[73] Anonymous. *The swedish traffic conflict technique*. 2005. URL : `http://www.tft.lth.se/rapporter/Conflict1.pdf`.

[74] W. Hu et al. « Traffic accident prediction using 3d model based vehicle tracking ». In : *IEEE Transactions on Vehicular Technology* 53 (3 2004), p. 677–694.

[75] N. Saunier et T. Sayed. « A Probabilistic Framework for the Automated Analysis of the Exposure to Road Collision ». In : *Transportation Research Record : Journal of the Transportation Research Board* 2083 (2008).

[76] M. Khayesi et A. Amekudzi. « Kingdon's multiple streams model and automobile dependence reversal path : the case of Curitiba, Bresil ». In : *Journal of Transport Geography* 19 (2011), p. 1547–1552.

[77] E. Greenberg. « Toward a new urbanist transportation agenda ». In : *Congress for the New Urbanism* (2005), p. 5–7.

[78] H. Hediyeh, T. Sayed et K. Ismail. « Automated analysis of pedestrian crossing speed behavior at scramble-phase signalized intersections using computer vision techniques ». In : *International journal of sustainable transportation* 8 (5 2014), p. 382–397.

[79] R. Passant et al. « Automated analysis of walking behavior : A case study from Qatar ». In : *Transportation Research Board 94th Annual Meeting* (2015), p. 1–18.

[80] J.Kang, I.Cohen et G. Medioni. « Object Reacquisition Using Invariant Appearance Model ». In : *International Conference on Pattern Recognition* (2004), p. 759–762.

[81] M. Mason et Z. Duric. « Using Histograms to Detect and Track Objects in Color Video ». In : *Applied Imagery Pattern Recognition Workshop* (2001), p. 154–159.

[82] H.Sidenbladh. « Detecting human motion with support vector machines ». In : *International Conference on Pattern Recognition* (2004), p. 188–191.

[83] F. Xu et K.Fujimura. « Human detection using depth and gray images ». In : *Conference on Advanced Video and Signal Based Surveillance* (2003), p. 115–121.

[84] I. Bouchrika et al. « Using Gait Features for Improving Walking People Detection ». In : *International Conference on Pattern Recognition* (2010), p. 3097–3100.

[85] A. Shashua, Y. Gdalyahu et G. Hayun. « Pedestrian detection for driving assistance systems : Single-frame classification and system level performance ». In : *Intelligent Vehicles Symposium* (2004), p. 1–6.

[86] S. Belongie, J. Malik et J. Puzicha. « Matching shapes ». In : *International Conference on Computer Vision* (2001), p. 454–461.

[87] S. Belongie, J. Malik et J. Puzicha. « Shape matching and object recognition using shape texts ». In : *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (4 2002), p. 509–522.

[88] G. Mori et J. Malik. « Estimating human body configurations using shape context matching ». In : *European Conference on Computer Vision* (2002), p. 666–680.

[89] N.Dalal. « Finding People in Images and Videos ». Phd thesis, National Polytechnic School of Grenoble, 2006.

[90] N. Dalal et B. Triggs. « Estimating human body configurations using shape context matchingHistograms of Oriented Gradients for Human Detection ». In : *Conference on Computer Vision and Pattern Recognition* (2005), p. 886–893.

[91] H. .Eng. et al. « A bayesian framework for robust human detection and occlusion handling using human shape model ». In : *International Conference on Pattern Recognition* (2004), p. 257–260.

[92] I. P. Alonso. et al. « Combination of feature extraction methods for svm pedestrian detection ». In : *Transactions on Intelligent Transportation Systems* 8 (2 2007), p. 292–307.

[93] I. Munder et D. Gavrila. « An experimental study on pedestrian classification ». In : *Transactions on Pattern Analysis and Machine Intelligence* 28 (11 2006), p. 1863–1868.

[94] C. Dai., Y. Zheng et X. Li. « Pedestrian detection and tracking in infrared imagery using shape and appearance ». In : *Computer Vision and Image Understanding* (2007), p. 288–299.

[95] P. Viola, M. J. Jones et D. Snow. « etecting pedestrians using patterns of motion and appearance ». In : *International Conference on Computer Vision* (2003), p. 734–741.

[96] K. Mikolajczyk, C. Schmid et A. Zisserman. « Human Detection Based on a Probabilistic Assembly of Robust Part Detectors ». In : *European Conference on Computer Vision* (2004), p. 69–82.

[97] Q. Zhu et al. « Fast Human Detection Using a Cascade of Histograms of Oriented Gradients ». In : *Conference on Computer Vision and Pattern Recognition* (2006), p. 1491–1498.

[98] M. Bertozzi et al. « Pedestrian localization and tracking system with Kalman filtering ». In : *Inteligent Vehicles Symposum* (2004), p. 584–589.

[99] T. Tung et T. Matsuyama. « Human Motion Tracking using a ColorBased Particle Filter Driven by Optical Flow ». In : *Workshop on Machine Learning for Vision-based Motion Analysis* (2008).

[100] L. Guo et al. « Pedestrian Tracking Based on Camshift with Kalman Prediction for Autonomous Vehicles ». In : *International Journal of Advanced Robotic Systems* 13 (120 2016), p. 292–307.

[101] S. Munder, C. Schnorr et D. M. Gavrila. « Pedestrian Detection and Tracking Using a Mixture of View Based Shape Texture Models ». In : *IEEE Transactions on Intelligent Transportation Systems* 9 (2 2008), p. 333–343.

[102] K. C. Li, H. Wang et J. Chiu. « Pedestrian tracking system by using human shape prior model ». In : *International Conference on Automation Science and Engineering* (2014).

[103] T. Kailath. « The Divergence and Bhattacharyya Distance Measures in Signal Selection Communication Technology ». In : *IEEE Transactions on In Communications* 15 (1 1967), p. 52–60.

[104] Z. Jiang et al. « Multiple Pedestrian Tracking using Colour and Motion Models ». In : *International Conference on Digital Image Computing : Techniques and Applications* (2010), p. 328–334.

[105] P. Perez et al. « Color based probabilistic tracking ». In : *European Conference on Computer Vision* (2002), p. 661–675.

[106] A. Andriyenko, K. Schindler et S. Roth. « Discrete-Continous Optimization for multiTarget Tracking ». In : *International Conference on Computer Vision and Pattern Recognition* (2012), p. 1926–1933.

[107] J. Kang, I. Cohen et G. Medioni. « Object Reacquisition using Invariant Appearance Model ». In : *International Conference on Pattern Recognition* (2004), p. 759–762.

[108] S. P. Hoogendoorn. « Pedestrian modelling by optimal con trol and differential games ». In : *German-Dutch-Finnish Seminar on Traffic Engineering* (2004), p. 153–172.

[109] A. Azim et O. Aycard. « Multiple pedestrian tracking using Viterbi data association ». In : *Intelligent Vehicles Symposium* (2010), p. 706–711.

[110] A. Derbel et al. « Comparative study between color texture and shape descriptors for multi-camera pedestrians identification ». In : *International Conference on Image Processing Theory and Tools and Applications* (2012), p. 313–318.

[111] N. Mansouri et al. « New Strategy Based on Spatiogram Similarity Association for Multi-Pedestrian Tracking ». In : *International Conference on Image Processing Theory and Tools and Applications* (2014), p. 1–6.

[112] C. Conaire, N. O'Connor et A. Smeaton. « An improved spatiogram similarity measure for robust object localization ». In : *International Conference on Acoustics and Speech and and Signal Processing* (2007), p. 1069–1072.

[113] R. Fisher. *CAVIAR : Context Aware Vision using Image based Active Recognition*. 2017. URL : `http://homepages.inf.ed.ac.uk/rbf/CAVIAR/`.

[114] K. Bernardin, A. Elbs et R. Stiefelhagen. « Multiple Object Tracking Performance Metric and Evaluation in a Smart Room Environment ». In : *International Workshop on Visual Surveillance* (2006).

[115] Z. Hu et Y. Tsai. « Homography-Based Vision Algorithm for Traffic Sign Attribute Computation ». In : *IEEE Transactions on In Communications* 24 (6 2009), p. 385–400.

[116] A. Toor et al. « Real world walking speeds of young pedestrians ». In : *SAE 2001 Transactions Journal of Passenger Cars Mechanical Systems V110 6* (2001), p. 11.

[117] T. Fugger et al. « Analysis of pedestrian gait and perception reaction at signal controlled crosswalk intersections ». In : *Transportation Research Record* 1705 (2000), p. 20–25.

[118] M. .Shokrolah et Sh.Morris. « Vision-based pedestrian behavior analysis at Intersections ». In : *Journal of Electronic Imaging* 25 (5 2016), p. 051203–051203.

[119] X. Jiang et al. « Analyses of pedestrian behavior on mid-block unsignalized crosswalk comparing Chinese and German cases ». In : *Advances in Mechanical Engineering* 17 (11 2015), p. 1–7.

[120] E. Avineri, D. Shinar et D. Susilo. « Pedestrians' Behaviour in Cross Walks : The effects of Falling and Age ». In : *Accident Analysis and Prevention journal* 44 (1 2012), p. 30–34.

[121] A. E. Patla. « Understanding the roles of vision in the cont rol of human locomotion ». In : *Gait and Posture* 5 (1997), p. 54–69.

[122] Y. Makihara et al. « Gait based age estimation using a whole generation gait database ». In : (2011).

[123] S. E. Choi et al. « Age estimation using a hierarchical classifier based on global and local facial features ». In : *Pattern Recognition* 44 (6 2011), p. 1262–1281.

[124] S. Zheng et al. « Learning universal multiview age estimator using video context ». In : (2011).

[125] Y. Makihara, H. Mannami et Y. Yagi. « Gait analysis of gender and age using a large-scale multiview gait database ». In : (2010).

[126] D. Zhang, Y. Wang et B. Bhanu. « Age classification base on gait using HMM ». In : (2010).

[127] J. T. Todd et al. « The perception of human growth. Scientific American ». In : *Pattern Recognition* 24 (2 1980), p. 132–144.

[128] T. Wu, P. K. Turaga et R. Chellappa. « Age estimation and face verification across aging using landmarks ». In : *IEEE Transactions on Information Forensics and Security* 7 (6 2012), p. 1780–1788.

[129] I. Pitanguy. « Defining and measuring aging parameters ». In : *Applied Mathematics and Computation* 78 (2-3 1996), p. 217–227.

[130] J. M. Guo, Y. M. Liou et H. Nguyen. « Human face age estimation with adaptive hybrid features ». In : *International Conference on System Science and Engineering* (2011).

[131] L. J. Hong et al. « Human face age estimation with adaptive hybrid features ». In : *International Conference on Machine Learning and Cybernetics* (2013).

[132] Z. Song et al. « Learning universal multiview age estimator using video context ». In : *International Conference on Computer Vision* (2013).

[133] H. M. fard et al. « Human Age Group Estimation Based On Anfis Using The HOG And LBP Features ». In : *Electrical and Electronics Engineering : An International Journal* 2 (1 2013), p. 21–29.

[134] M. Pietikainen et J. Heikkila. « Image and Video Description with Local Binary Pattern ». In : *International Conference on Computer Vision and Pateren Recognition* (2011).

[135] Y. Liang et al. « A Hierarchical Framework for Facial Age Estimation ». In : *Mathematical Problems in Engineering* 2014 (2014), p. 1–8.

[136] V.Tamil, V. Selvi et K.Vani. « An Efficient Age Estimation System Based on Multi Linear Principal Comonent Analysis ». In : *Journal of Computer Science* 7 (10 2011), p. 1497–1504.

[137] D. Blanke et P. Hageman. « Comparison of gait of young men and elderly men ». In : *Physical Therapy* 69 (2 1989), p. 144–148.

[138] K. Ostrosky et al. « A comparison of gait characteristics in young and old subjects ». In : *Physical Therapy* 74 (7 1989), p. 637–644.

[139] N. Boulgouris et Z. Chi. « Human gait recognition based on matching of body components ». In : *Pattern Recognition* 40 (6 2007), p. 1763–1770.

[140] G. Huang et Y. Wang. « Gender classification based on fusion of multiview gait sequences ». In : *Asian Conference on Computer Vision* (2007).

[141] L. Lee et W. Grimson. « Gait analysis for recognition and classification ». In : *International Conference on Automatic Face and Gesture Recognition* (2002).

[142] X. Li et al. « Gait components and their application to gender recognition ». In : *Transaction on Systems and Man and Cybernetics* 38 (2008), p. 145–155.

[143] S. Yu et al. « A study on gait based gender classification ». In : *IEEE Transaction on Image Processing* 18 (8 2009), p. 1905–1910.

[144] J. Han et B. Bhanu. « Individual recognition using gait energy image ». In : *Transaction on Pattern Analysis and Machine Intelligence* 28 (2 2006), p. 316–322.

[145] Y. Makihara et al. « Gait recognition using a view transformation model in the frequency domain ». In : *European Conference on Computer Vision* (2006).

[146] M. S. Nixon, T. N. Tan et R. Chellappa. « Human Identification Based on Gait ». In : *International Series on Biometrics* (2005).

[147] M. Hu, Y. Wang et Z. Zhang1. « Maximisation of mutual information for gait based soft biometric classification using Gabor features ». In : *IET Biometrics* 1 (1 2012), p. 55–62.

[148] A. A. Kale et al. « Identification of humans using gait ». In : *IEEE Transactions on Image Processing* 13 (9 2004), p. 1163–1173.

[149] K. Benz et al. « A Preliminary Study of Gait based Age Estimation Techniques ». In : *Article of APSIPA Annual Summit and Conference* (2015).

[150] M. S. Giaquinto, E. Ciotolaa et F. Margutti. « Gait in the water : A comparison between young and elderly subjects ». In : *Disability and Rehabilitation* 29 (9 2007), p. 727–730.

[151] R. V. Briani1 et al. « Comparison of temporal and kinetic walking parameters among young people and falling and nonfalling elderly persons ». In : *Revista Brasileira de Geriatria e Gerontologia* 18 (4 2015), p. 761–768.

[152] M. Cunado, S. Nixon et J. N. Carter. « Automatic extraction and description of human gait models for recognition purposes ». In : *Computer Vision and Image Understanding* 90 (1 2003), p. 1–41.

[153] A. Derbel et al. « Comparative study between spatio/temporal descriptors for pedestrians recognition by gait ». In : *International Conference on Image Analysis and Recognition* (2014).

[154] H. Iwama et al. « The OUISIR gait database comprising the large population dataset and performance evaluation of gait recognition ». In : *IEEE Transactions on Information Forensics and Security* 7 (5 2012), p. 1511–1521.

[155] P. Philips et al. « Comparative study between spatio/temporal descriptors for pedestrians recognition by gait ». In : *International Conference on Automatic Face and Gesture Recognition* (2002).

[156] L. Wang et al. « Silhouette analysis based gait recognition for human identification ». In : *IEEE Transaction on Pattern Analysis and Machine Intelligence* 25 (12 2003), p. 1505–1518.

[157] A. R. Chowdhury. *Human Identification at a Distance*. 2001. url : `http://www.umiacs.umd.edu/labs/pirl/hid/data.html`..

[158] S. Sivaraman et M. M. Trivedi. « Looking at Vehicles on the Road : A Survey of Vision Based Vehicle Detection and Tracking and Behavior Analysis ». In : *IEEE Transactions On Intelligent Transportation Systems* 14 (4 2013), p. 1773–1795.

[159] D. Surender et al. « A Survey on Vehicle Detection based on Vision ». In : *Modern Applied Science* 9 (12 2015), p. 118–124.

[160] G. Wang, D. Xiao et J. Gu. « Review on Vehicle Detection based on Video for Traffic Surveillance ». In : *International Conference on Automation and Logistics Qingdao China* (2008).

[161] S. Sato et al. « Multilayer Lidar based Pedestrian Tracking in Urban Environments ». In : *IV'10 and Intelligent Vehicules Symposium* (2010).

[162] C. Azevedo et al. « Automatic Vehicle Trajectory Extraction by Aerial Remote Sensing ». In : *Social and Behavioral Sciences* 111 (2014 2014), p. 849–858.

[163] X.Mao et al. « Amplitude modulated Laser Radar for Range and Speed Measurement in Car Applications ». In : *IEEE Transaction in Intelligent Transportation Systems* 13 (1 2012), p. 408–413.

[164] J. Antony et M. Suchetha. « Vision Based Vehicle Detection : A Literature Review ». In : *International Journal of Applied Engineering Research* 11 (5 2016), p. 3128–3133.

[165] A. Angel et al. « Methods of analyzing traffic imagery collected from aerial platforms ». In : *IEEE Transactions on Intelligent Transportation Systems* 4 (2 2003), p. 99–107.

[166] z. Sun, G. Bebis et R. Miller. « On road vehicle detection : A review ». In : *Pattern Analyses Machine Intelligent* 28 (5 2006), p. 694–711.

[167] A. Jazayeri et al. « Vehicle detection and tracking in car video based on motion model ». In : *IEEE Transaction on Intelligent Transportation System* 12 (2 2011), p. 583–595.

[168] P. Viola et M. Jones. « Rapid object detection using a boosted cascade of simple features ». In : *IEEE Computer Vision Pattern Recognition Conference* (2001).

[169] A. Fossati, P. Schonmann et P. Fua. « Real time vehicle tracking for driving assistance ». In : *Machine Vision Application* 22 (2 2010), p. 439–448.

[170] C. Hoffmann. « Fusing multiple 2D visual features for vehicle detection ». In : *IEEE Intelligent Vehicule Symposium* (2006).

[171] T. Gandhi et M. M. Trivedi. « Parametric ego motion estimation for vehicle surround analysis using an omnidirectional camera ». In : *Machine Vision Application* 216 (2 2010), p. 85–95.

[172] C. Idler et al. « Real time vision based multi target tracking with particle filters in automotive applications ». In : *IEEE Intelligent Vehicule Symposium* (2006).

[173] D. Kasper et al. « Object oriented Bayesian networks for detection of lane change maneuvers ». In : *IEEE IV* (2011).

[174] W. Liu et al. « Rear vehicle detection and tracking for lane change assist ». In : *IEEE Intelligent Vehicule Symposum* (2007).

[175] S. Sivaraman et M. M. Trivedi. « Active learning for on road vehicle detection : A comparative study ». In : *Machine Vision Application, Special Issue Car Navigation and Vehicle Systems* (2011).

[176] J. Cui et al. « Vehicle localization using a single camera ». In : *IEEE IV* (2010).

[177] A. Haselhoff et A. Kummert. « A vehicle detection system based on Haar and triangle features ». In : *IEEE Intelligent Vehicule Symposum* (2009).

[178] S. Sivaraman et M. Trivedi. « A general active learning framework for on road vehicle recognition and tracking ». In : *IEEE Transaction in Intelligent Transportation Systems* 11 (2 2010), p. 267–276.

[179] W. C. Chang et C. W. Cho. « Real time side vehicle tracking using parts based boosting ». In : *IEEE Int. Conf. SMC* (2008).

[180] A. Haselhoff et A. Kummert. « An evolutionary optimized vehicle tracker in collaboration with a detection system ». In : *Int. IEEE Conf. ITSC* (2009).

[181] S. Sivaraman et M. M. Trivedi. « Real time vehicle detection using parts at intersections ». In : *IEEE Intelligent Transportation Systems Conference* (2012).

[182] S. Teoh et T. Brunl. « Symmetry based monocular vehicule detection system ». In : *Machine Vision Application* 23 (5 2012), p. 831–842.

[183] M. Cheon et al. « Vision based vehicle detection system with consideration of the detecting location ». In : *IEEE Transaction in Intelligent Transportation Systems* 13 (3 2012), p. 1243–1252.

[184] R. Wijnhoven et P. de With. « Unsupervised subcategorization for object detection : Finding cars from a driving vehicle ». In : *IEEE ICCV Workshops* (2011).

[185] Y. Zhang, S. Kiselewich et W. Bauson. « Legendre and Gabor moments for vehicle recognition in forward collision warning ». In : *IEEE ITSC* (2006).

[186] S. Toral, F. Barrero et M. Vargas. « Development of an embedded vision based vehicle detection system using an ARM video processor ». In : *ITSC'2011 and 14th International Conference on Intelligent Transportation Systems* (2011).

[187] T. Machida et T. Naito. « GPU and CPU cooperative accelerated pedestrian and vehicle detection ». In : *ICCV'2011 and International Conference on Computer Vision ICCV Workshops* (2011).

[188] K. Philip et al. « Performance analysis and visualisation in tennis using a lowcost camera network ». In : *ACMM'2010 and Multimedia Grand Challenge Track at ACM Multimedia* (2010).

[189] S. Agarwal et D. Roth. *UIUC Image Database for Car Detection*. 2002. URL : `http://cogcomp.cs.illinois.edu/Data/Car/`.

[190] URL : `http://pascallin.ecs.soton.ac.uk/challenges/VOC/databases.html#TUD`.

[191] C. Vondrick et al. « Visualizing Object Detection Features ». In : *International Conference on Computer Vision* (2013).

[192] C. Caraffi, T. Vojir et J. Sochman. « A system for real-time detection and tracking of vehicles from a single car-mounted camera ». In : *15th International Intelligent Transportation Systems* (2012).

[193] J. Cooper. « Experience with Traffic Conflicts in Canada with Emphasis on 'Postencroachment Time' Techniques ». In : *International Calibration Study of Traffic Conflict Techniques* (1983).

[194] C. Hupfer. « Deceleration to Safety Time (DST) a Useful Figure to Evaluate Traffic Safety ». In : *ICTCT and Department of Traffic Planning and Engineering* (1997).

[195] D. Vernon et D. Ecvision. *research roadmap v4.0.* 2004. url : `http%20://www.ecvision.org/,%202004`.

[196] A. Ziani, C. Motamed et J. C. Noyer. « Temporal reasoning for scenario recognition in video-surveillance using Bayesian networks ». In : *IET Computer Vision, Special Issue on Visual Information Engineering* 2 (2 2008), p. 99–107.

[197] T. Xiang et S. Gong. « Beyond tracking : Modelling activity and understanding behaviour ». In : *International Journl in Computer Vision* 67 (2006), p. 21–51.

[198] L. Zhao et al. « Video behavior analysis using topic models and rough sets (applications notes) ». In : *IEEE Computer Intelligent Magazine* 8 (2013), p. 56–67.

[199] H. Bui, Q. Phung et S. Venkatesh. « Hierarchical hidden Markov models with general state hierarchy ». In : *National Conf. Artificial Intelligence* (2004).

[200] H. Bui, Q. Phung et S. Venkatesh. « Hierarchical hidden Markov models with general state hierarchy ». In : *International Conference on Pattern Recognition* (2000).

[201] M. Brand et V. Kettnaker. « Discovery and segmentation of activities in video ». In : *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (8 2000), p. 844–851.

[202] G. Sanrom et al. « unified approach to the recognition of complex actions from sequences of zone-crossings ». In : *Image and Vision Computing* 32 (5 2014), p. 363–78.

[203] T. Hospedales, S. Gong et T. Xiang. « Video behavior mining using a dynamic topic model ». In : *International Journal in Computer Vision* 98 (3 2012), p. 303–323.

[204] X. Wang et X. Ma. « Grimson, W. Unsupervised activity perception in crowded and complicated scenes using hierarchical bayesian models ». In : *IEEE Transaction in Paterner Analysis* 31 (4 2009), p. 539–555.

[205] A. Howell et H. Buxton. « RBF network methods for face detection and attentional frames ». In : *Neural Processing Letters* 15 (3 2002), p. 197–211.

[206] C. Tessier. « Towards a commonsense estimator for activity tracking ». In : *AI Springer Symposum* (2003).

[207] V. Vu, F. Bremoud et M. Thonnat. « Automatic video interpretation : a novel algorithm for temporal scenario recognition ». In : *International Joint Conference on Artificial Intelligence* (2003).

[208] C. D. C. P. L. Maigat. « Chronicle recognition improvement using temporal focusing and hierarchization ». In : *International Joint Conference on Artificial Intelligence* (2007).

[209] F. Zhang et al. « Star Skeleton for Human Behavior Recongition ». In : *International Conference on Audio, Language and Image Processing* (2012).

[210] Y. Chen et al. « Abnormal behavior detection by multi-svm-based Bayesian network ». In : *International Conference on Information Acquisition* (2007).

[211] T. Wang et h. Snoussi. « Detection of abnormal events via Optical flow feature analysis ». In : *Sensors* 15 (4 2015), p. 7157–7171.

[212] J. Candamo et al. « Understanding Transit Scenes :A Survey on Human Behavior Recognition Algorithms ». In : *IEEE Transaction in Intelligent Transportation Systems* 11 (1 2009), p. 206–224.

[213] G. Medioni et al. « Event Detection and Analysis from Video Streams ». In : *IEEE Transaction in Paterner Analysis* 23 (8 2001), p. 873–889.

[214] N. Oliver, B. Rosario et A. entland. « A bayesian computer vision system for modeling human interactions ». In : *IEEE Trans. Pattern Anal. Mach. Intell.* 22 (2000), p. 831–843.

[215] Z. Zhang, T. Tan et K. Huang. « An extended grammar system for learning and recognizing complex visual events ». In : *IEEE Trans. Pattern Anal. Mach. Intell.* 33 (2011), p. 240–255.

[216] Y. Zeng et K. Xiang. « Applying behavior recognition in road detection using vehicle sensor networks ». In : *Proceedings of the International Conference on Computing, Networking and Communication* (2012).

[217] e. a. H. T. Thien. « Traffic Behavior Recognition Using the Pachinko Allocation Model ». In : *Sensors* (2015).

[218] K. Ismail, T. Sayed et N. Saunier. « Automated Collection of Pedestrian Data Using Computer Vision Techniques ». In : *Transportation Research Board Annual Meeting* (2009).

[219] « Organisation de Cooperation, "Zero tue sur la route : un systéme sur, des objectifs ambitieux" ». In : *Paris Organisation de Cooperation et de Developpement economiques* (2008).

[220] M. A. Belin. « Public road safety policy change and its implementation : Vision Zero a road safety policy innovation ». Phd thesis, Stockholm, Karolinska Institutet, 2012.

[221] L. Lonero, K. Clinton et D. Sleet. « Behavior change interventions in road safety ». In : *Injury and violence prevention : behavioral science theories, methods and applications* (2006).

[222] P. Smets. « Jeffrey's rule of conditioning generalized to belief functions ». In : *IDIA, Universite Libre de Bruxelles* (2013).

[223] N. P. et al. *Reseaux Bayeisien*. Eyrolles,

[224] M. P. et al. *Rapport mondial sur la prevention des traumatismes dus aux accidents de la circulation*. Geneve, Organisation mondiale de la Santé, 2004.

[225] P. Jacobsen. « Safety in numbers : more walkers and bicyclists, safer walking and bicycling ». In : *Injury Prevention* 9 (2003), p. 205–209.

[226] A. Ahouandjinou. « Reconnaissance de scénario par les Modéles de Markov Caches Credibilistes. Application a l'interprétation automatique de séquences vidéos medicales ». Thèse de doctorat, Université du Littoral Cote d'Opale-ULCO, 2014.

[227] L. Cannon. *The Cost of Urban Traffic Congestion in Canada*. Transport Canada, Environmental Affairs, 2012.

[228] P. Dollar, S. Belongie et P. Perona. « The Fastest Pedestrian Detector in the West ». In : *the British Machine Vision Conference* (2010).

[229] J. N. Morris et A. E. Hardman. « Walking to health ». In : *Sports Medicine* 23 (5 1997), p. 306–332.

[230] OCDE. *Pitons : Sécurité, espace urbain et santé*. Forum International des Transports. Rapport de recherche, 2011.

[231] N. Benlahrech, A. L. Ruyet et C. Liveboardon. *La mobilite des personnes agees. Analyse des enquetes menages deplacements. Rapport d'etude*. CERTU, 2001.

[232] A. S. Heam et M. Dejeammes. *L'insécurité routière des piétons agés à travèrs le système Mobilité - Urbanisme - Réseau, Rapport d'étude - Insécurité routière des piétons agés*. CERTU, 2000.

[233] F. Papon. « Les modes oublies : Marche, bicyclette, cyclomoteur, motocyclette ». In : *Recherche Transports Sécurité* 56 (1 1997), p. 61–75.

[234] *Direction technique Territoires et ville Direction territoriale Centre Est, "Rapport : Piétons séniors et amenagement de la voirie en milieu urbain-Etat de l'art"*. Cerema, 2016.

[235] D. N. Lee, D. S. Young et C. M. McLaughlin. « A roadside simulation of road crossing for children ». In : *Ergonomics* 27 (12 1984), p. 1271–1281.

[236] J. D. Demetre et al. « Errors in young children's decisions about traffic gaps : Experiments with roadside simulations ». In : *British Journal of Psychology* 83 (1 1992), p. 189–202.

[237] A. Ziani. « Interprétation en temps réel de séquence vidéo par éxploitation des modéles graphiques probabilistes ». Thèse de doctorat, Université du Littoral Cote d'Opale-ULCO, 2010.

Annexe

I Acquisition de la base de données : CrossWalk Valenciennes (CWV)

Le système d'analyse comportementale des piétons lors de la traversée de la rue proposé est composé de plusieurs sous-systèmes tels que le sous-système de collecte de données des piétons, de trafic ainsi que l'étude de l'effet des aménagements urbains (cas de feux de circulation). Pour la validation de ce système nous avons fait l'acquisition d'une base de traversée de la rue intégrant plusieurs scénarios et englobant les différentes hypothèses d'interactions entre les piétons et les véhicules. L'acquisition de cette base est effectuée sur deux endroits différents : Aulnoy-lez-Valenciennes (Nord de la France) et La garde à Toulon (Sud de la France). Sur la première zone deux types de scénarios de tests sont filmées : les premières scénarios se composent de séquences vidéo pour la validation de l'étape de collecte de paramètres des piétons. Alors que les deuxièmes scénarios intègrent des séquences vidéo pour la validation de l'étape de collecte de paramètres de trafic. Ces scénarios sont filmés en respectant les protocoles présentés à la fin de cette annexe (respectivement scénarios de tests 1 et scénarios de tests 2). Les équipements et les matériels utilisés pour l'acquisition des ces scénarios sont aussi décrits dans ces protocoles. Les figures **??** et **??** présentent la zone de tournage ainsi que les dispositions des caméras.

L'acquisition de ces scénarios a nécessité la sollicitation de plusieurs candidats. Des candidats ont joué le rôle des piétons et d'autres de conducteurs et de co-pilotes (pour assurer la communication avec les conducteurs sans lui engendrer aucun risque). Les protocoles de tests sont distribués aux candidats en lui expliquant les taches à faire.

Néanmoins plusieurs bases pour la détection et le suivi des personnes et des véhicules sont disponibles (proposés par la communauté de vision par ordinateur) et qui ont été utilisées pour la validation des algorithme de détection et de suivi, l'acquisition de ces scénarios est justifiée par le fait que les bases existantes n'englobent pas les hypothèses de traversée de la rue ni les interactions et les conflits P-V.

II Acquisition des scénario de test à Toulon

Afin d'étudier l'effet des aménagements urbains et valider l'hypothèse que ces der-



FIGURE 5.16 – Présentation de la zone de tournage, Aulnoy-lez-Valenciennes, avec les dispositions des caméras (Vue 1) : Les caméras encerclées en rouge sont les caméras pour filmer le trafic gauche et doit et la caméra encerclées en vert est pour filmer les piétons.



FIGURE 5.17 – Présentation de la zone de tournage, Aulnoy-lez-Valenciennes, avec les dispositions des caméras (Vue 2) : Les caméras encerclées en rouge sont les caméras pour filmer le trafic gauche et doit et la caméra encerclées en vert est pour filmer les piétons.

niers facilitent l'action de traversée, nous avons suivi le comportement des piétons avant et après l'installation d'un feu de circulation. Après la prise de contact avec la mairie de

La Garde à Toulon, nous avons été informés que juste devant l'université de La Garde, la mairie envisage installer un nouveau feu (demande de tournage attachée à la fin de cette annexe). Une telle zone alors constitue un endroit d'étude très favorable pour notre objectif. Par conséquent nous avons filmé le comportement aléatoire des piétons et de véhicules sur cette zone avant et après l'installation de feux de circulation. Une présentation de la zone de tournage est présentée par la figure **??**.



FIGURE 5.18 – Présentation de la zone de tournage, La Garde à Toulon, avec les dispositions des caméras : Les caméras encerclées en rouge sont les caméras pour filmer le trafic gauche et doit et la caméra encerclées en vert est pour filmer les piétons.

Scénarios de tests 1

Objectif : Par ces scénarios, nous envisageons tester et valider le processus de détection et de suivi d'une personne et calculer sa vitesse durant les différentes phases de traversée de rue.

Dispositif général :

- Une caméra fixe
- 3 ou 4 candidats
- Un chronomètre pour calculer la durée de chaque phase (rapprochement, attente, traversée) afin qu'on les compare avec les durées calculées automatiquement.
- Une séquence pour chaque candidat et pour chaque scénario.

Scénario 1 : filmer une personne qui se rapproche latéralement du passage piéton (vue frontaux/parallèle) puis s'arrête au bord de la rue après il la traverse (vue frontale). Ce scénario est répété avec deux types de marches :

1.1) Marche normale
1.2) Marche rapide

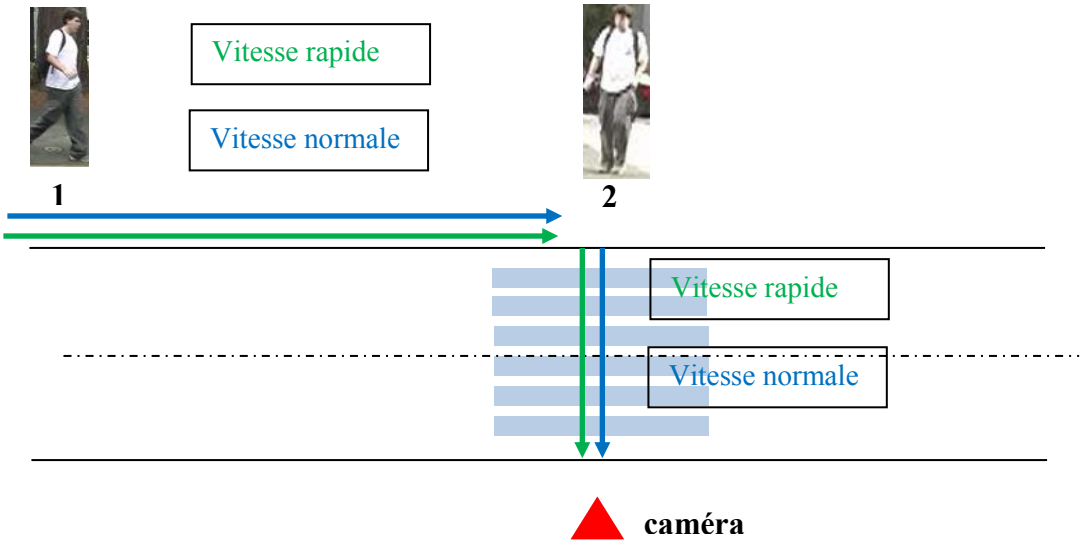


Fig1. Scénario 1

Scénario 2 : filmer une personne qui se rapproche latéralement du passage piéton (vue frontaux/parallèle) avec une vitesse normale constante, puis s'arrête au bord de la rue après il la traverse (vue frontale) avec une vitesse rapide. En fait pour ce scénario, on a une variation considérable entre la vitesse de marcher sur le territoire et la vitesse de traversée.

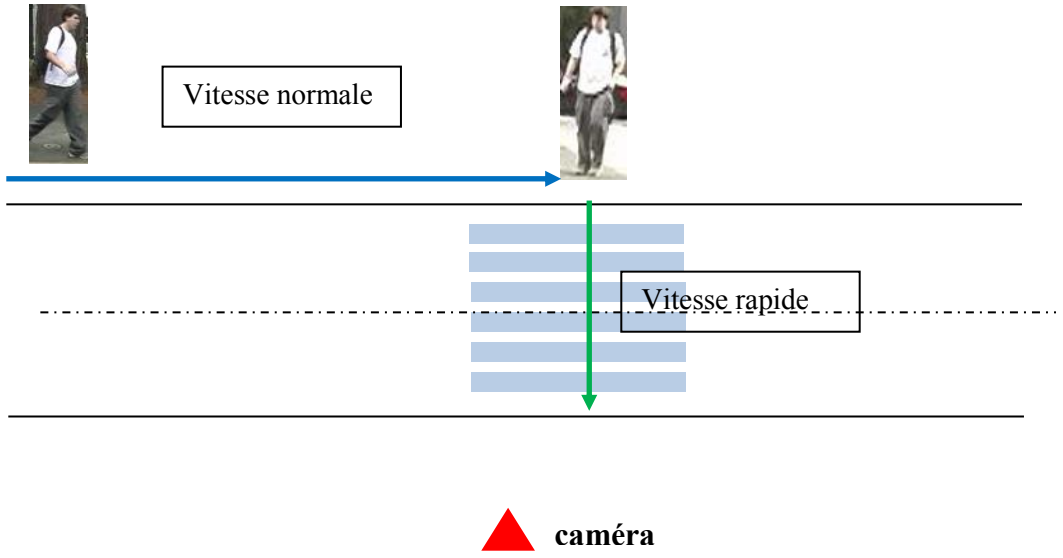


Fig2. Scénario 2

Scénario 3 : filmer une personne qui se rapproche latéralement du passage piéton (vue frontaux/parallèle) avec soit une vitesse quelconque soit une vitesse constante, puis s'arrête au bord de la rue après il la traverse (vue frontale) avec une vitesse quelconque. La personne change la direction de marcher sur le passage piéton :

3.1) Faisant un angle aigu (Φ_1) (avec une vitesse quelconque et avec une vitesse constante)

3.2) Faisant un angle obtu (Φ_2) (avec une vitesse quelconque et avec une vitesse constante)

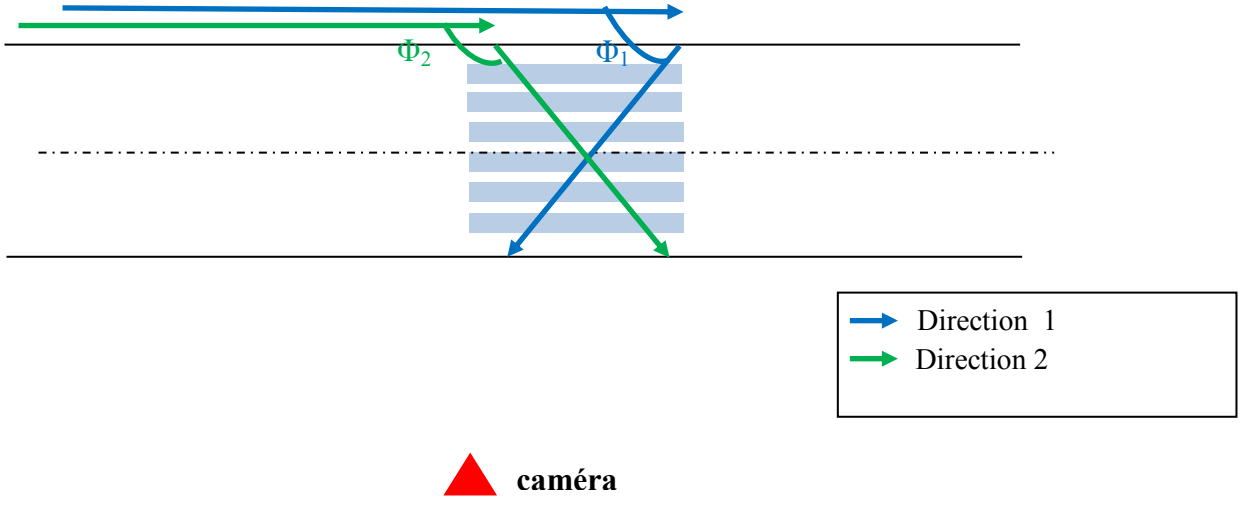


Fig 3. Scénario 3

Scénario 4 : filmer une personne qui se rapproche latéralement du passage piéton (vue frontaux/parallèle) avec une vitesse normale, puis s'arrête au bord de la rue pour faire une première période d'attente, après il traverse la première voie de la rue. Ensuite il fait une deuxième période d'attente avant de traverser la deuxième voie.

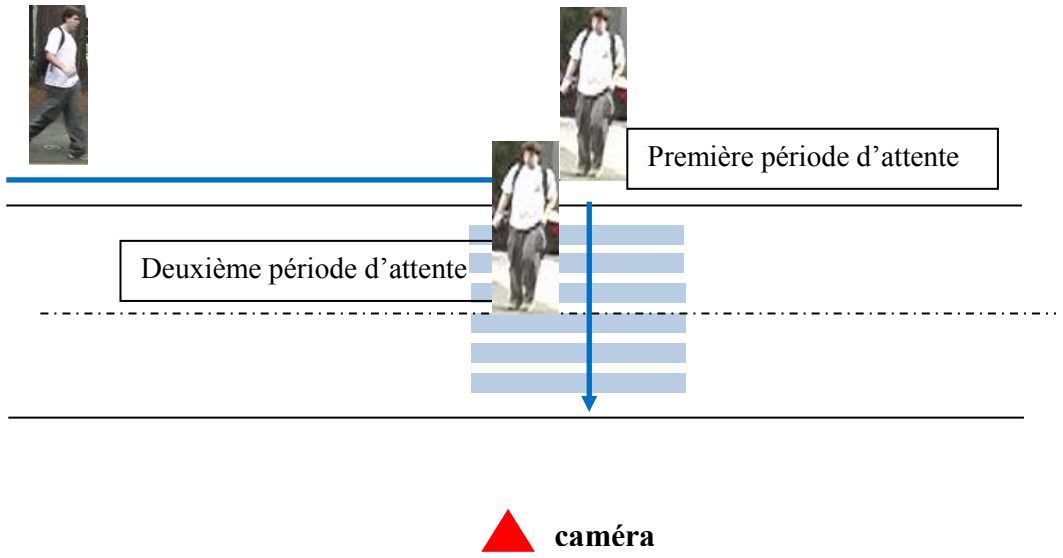


Fig 4. Scénario 4

Scénario 5: filmer une personne qui se rapproche latéralement du passage piéton (vue frontaux/parallèle) avec une vitesse normale. Puis il traverse la rue sans aucune période d'attente.

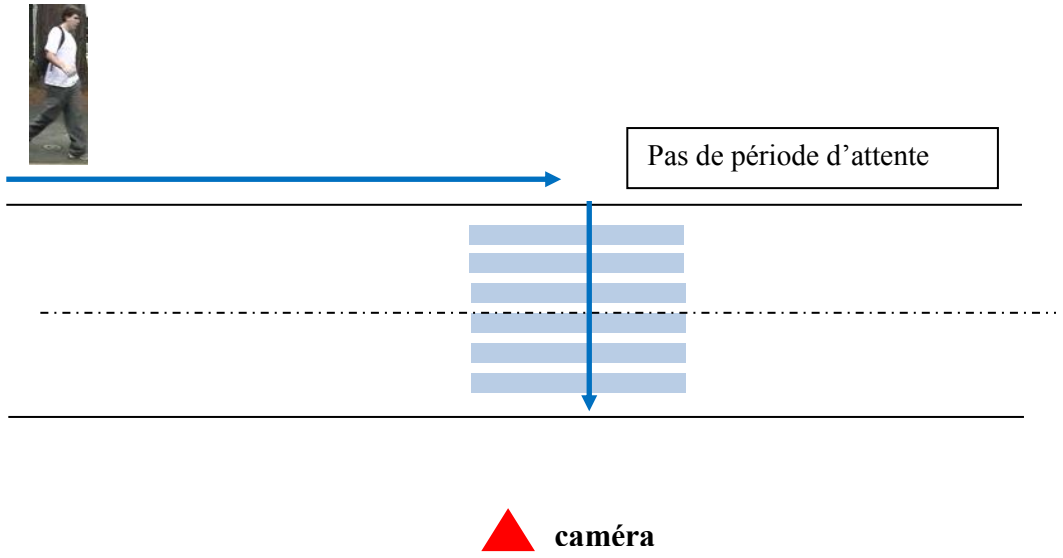
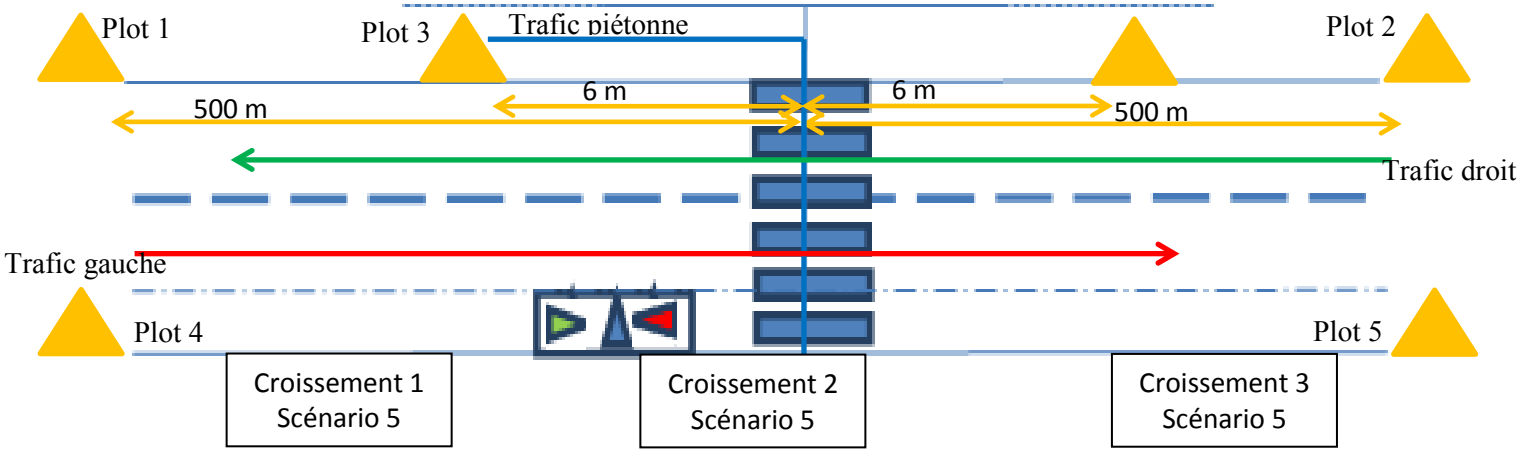


Fig 5. Scénario 5

Scénarios de tests 2

Objectif : Par ces scénarios nous envisageons tester et valider le processus de détection et suivi de véhicules vis à vis la variation de vitesse et le croisement de deux véhicules en premier lieu. En deuxième lieu, nous envisageons enregistrer quelques séquences qui englobent la totalité de la scène : piéton et véhicule en même temps avec quelques interactions.

Scène :



Equipements :

- 2 véhicules
- 5 plots
- Outil de mesure de distance (mètre)
- 3 caméras avec 3 pieds
- 4, 5 piétons
- Téléphone portable pour communiquer avec les 'co-pilotes' des voitures pour leur donner le signal de départ et leur rappeler les consignes à transmettre au conducteur. C'est ce même co-pilote qui aura pour fonction de sortir le bras de la fenêtre à l'instant précis où la voiture passe à hauteur des plots (pour déterminer de manière précise la position des véhicules en post production)

Scénario 1 : Test avec une variété de vitesses constantes de véhicules sans passage de piétons

Filmer un véhicule qui roule avec une vitesse constante : 20, 40, 50 km^{-1}. Chacun des conducteurs de véhicules passe une fois sur la voix 1 (champ de vision de la caméra droite) et une fois sur la voix 2 (champ de vision de la caméra gauche) sans passage de piétons.

- 2 fois pour le trafic gauche
- 2 fois pour le trafic droit

<u>Scénario 2 :</u> Test avec une vitesse constante de véhicules avec freinage sans passage de piétons

Filmer un véhicule qui roule avec une vitesse constante de 20 km^{-1} puis frêne juste près du passage piéton et s'arrête 5 secondes puis repart à la même vitesse de 20 km^{-1}. Chacun des conducteurs de véhicule passe une fois sur la voix 1 (champ de vision de la caméra droite) et une fois sur la voix 2 (champ de vision de la caméra gauche) sans passage de piétons

- 2 fois pour le trafic gauche
- 2 fois pour le trafic droit

<u>Scénario 3 :</u> Test avec vitesse constante de véhicules avec freinage et passage d'un seul piéton

Filmer un véhicule qui roule avec une vitesse constante de 40 km^{-1} puis frêne juste près du passage piéton pour laisser passer un piéton. Chacun des conducteurs de véhicule passe une fois sur la voix 1 (champ de vision de la caméra droite) et une fois sur la voix 2 (champ de vision de la caméra gauche) sans passage de piétons

- 2 fois pour le trafic gauche
- 2 fois pour le trafic droit

<u>Scénario 4 :</u> Test avec vitesse constante de véhicules avec freinage et passage de plusieurs piétons

Filmer un véhicule qui roule avec une vitesse constante de 40 km^{-1} puis s'arrête au passage piéton pour laisser passer des piétons dans les deux sens de traversée.

- 2 fois pour le trafic gauche
- 2 fois pour le trafic droit

<u>Scénario 5:</u> Test avec un croisement de deux véhicules sur deux voix différentes

Filmer 2 véhicules qui roulent chacun sur une voix (dans deux directions opposées) à une vitesse constante de 40 ou 50 km^{-1} et qui se croisent pour quelques instants.

- 3 fois : Une fois où les deux sont à 40 km^{-1}, une fois ou les deux sont à 50 km^{-1} et une fois l'un est à 40 km^{-1} et l'autre à 50 km^{-1}.

<u>Scénario 6 :</u> Test avec un croisement de deux véhicules sur la même voix

Filmer 2 véhicules qui roulent avec des vitesses constantes de 40 et 50 km^{-1} respectivement l'un derrière l'autre sur la même voix après un certain moment il y a doublement du véhicule arrière par rapport à celui en avance.

- 2 fois

République Française
Département du Var
Ville de La Garde

Jean-Louis MASSON
Maire de La Garde
1er Vice-Président du Conseil
Départemental du Var
Vice-Président de la Communauté d'Agglomération
Toulon-Provence-Méditerrannée
Chevalier de la Légion d'Honneur
Officier de l'Ordre National du Mérite

La Garde, le 15 décembre 2015

Courrier reçu le :
04 JAN. 2016
UFR STAPS

à

Monsieur Eric WATELAIN
Maître de conférences
Université de Toulon UFR STAPS
Avenue du B I M P
83130 LA GARDE

VISAS		
CSPM	DGAS	Jurid.

POLICE MUNICIPALE
JLM/RC/EC – n° 381/2015

Affaire suivie par :
Monsieur Rémy CIRINA - CSPM
04.94.08.98.20

OBJET : Demande d'autorisation

Monsieur,

Je fais suite à votre courrier du 4 décembre dernier par lequel vous sollicitez auprès de moi une autorisation pour filmer la RD 29, avenue du 1er B.I.M.P. et ainsi recueillir des informations dans le cadre d'un projet de thèse.

C'est bien volontiers que j'émets un avis favorable à votre demande.

Par ailleurs, je vous demande, si cela est possible, de bien vouloir partager les informations que vous seriez en mesure de recueillir car la municipalité est, depuis longtemps, engagée dans une démarche de sécurité routière et en cela, vous pourriez vous rapprocher de monsieur Jean-Pierre HASLIN, premier adjoint en charge des questions de sécurité et de circulation.

Vous souhaitant pleine réussite dans votre projet.

Veuillez agréer, Monsieur, l'assurance de ma considération distinguée.

Le Maire,

Jean-Louis MASSON

MAIRIE DE LA GARDE VAR

imprimerie municipale