



Faculté des sciences de l'ingénierie
Département d'informatique

THÈSE

Présentée en vue de l'obtention du diplôme de
Doctorat Ès Sciences

Multimédia mining Reconnaissance des formes dans une vidéo

Option : Système d'informations et de connaissance

Par

Mr. Brahim FAROU

Directeur de Thèse :

Hamid Seridi Prof. Université 8 mai 1945 - Guelma

DEVANT LE JURY

Président :

Labiba Souici-Meslati

Prof. Université Badji Mokhtar - Annaba

Examineurs :

Nadir Farah

Prof. Université Badji Mokhtar - Annaba

Abdelhani Boukrouche

Prof. Université 08 MAI 45 - Guelma

Abdelkrim Amirat

Prof. Université Mohamed-Chérif Messaadia - Souk Ahras



Faculté des sciences de l'ingénierie

Département d'informatique

THÈSE

Présentée en vue de l'obtention du diplôme de

Doctorat Ès Sciences

Multimédia mining Reconnaissance des formes dans une vidéo

Option : Système d'informations et de connaissance

Par

Mr. Brahim FAROU

Directeur de Thèse :

Hamid Seridi Prof. Université 8 mai 1945 - Guelma

DEVANT LE JURY

Président :

Labiba Souici-Meslati

Prof. Université Badji Mokhtar - Annaba

Examineurs :

Nadir Farah

Prof. Université Badji Mokhtar - Annaba

Abdelhane Boukrouche

Prof. Université 08 MAI 45 - Guelma

Abdelkrim Amirat

Prof. Université Mohamed-Chérif Messaadia - Souk Ahras

DEDICACE

À Mes Parents,

À Mes Frères : Dhia Eddine, Houssam Eddine et Zakarya

À Mon Épouse Halima,

À Ma Petite Fille Chérie Maria,

À Mon neveu Mohamed Wassim,

À Ma tante Sonia, son époux Hamid et leurs enfants,

À Mes Oncles : Djamel, Salim, Diafer et leurs familles,

À Toute Ma Famille,

À Tous Mes Amis,

REMERCIEMENTS

Tout d'abord, je remercie notre clément Dieu qui m'a donné la puissance, le courage et la détermination pour finaliser cette thèse.

J'adresse mes plus vifs remerciements au professeur Hamid SERIDI, mon directeur de thèse, pour sa rigueur scientifique et son amitié sincère dont il a fait preuve le long de ce travail. J'apprécie la confiance qu'il m'a témoignée et les conseils avisés qu'il m'a prodigués. Votre connaissance pointue des problématiques de recherche actuelles, votre philosophie, et votre résonnement unique m'ont permis de comprendre les fondements de la science et de faire aboutir ce travail. Merci également de m'avoir encouragé dans les moments les plus difficiles, sur le plan scientifique autant qu'humain.

Je souhaite exprimer toute ma reconnaissance aux professeurs Labiba SOUICI-MESLATI et Nadir FARAH de l'université d'Annaba, au professeur Abdelhani BOUKROUCHE de l'université de Guelma et au professeur Abdelkrim AMIRAT de l'université de Souk-Ahras pour l'honneur qu'ils m'ont fait en acceptant la charge d'examiner ce travail et de participer au Jury de soutenance.

Je tiens à exprimer ma profonde gratitude à Monsieur Yacine LAFIFI pour ses directives sur le plan technique et administratif. Je le remercie pour sa disponibilité, ses encouragements, son soutien moral, sa sympathie et son amitié sincère.

Un grand merci au professeur Hassina SERIDI pour ces conseils judicieux et son aide précieuse

Je remercie vivement mes parents et frères qui m'ont beaucoup soutenu durant les moments difficiles.

Je tiens à exprimer ici toute ma reconnaissance à Halima ALLAL, ma conjointe, qui m'a réconforté, motivé, entouré et qui a particulièrement supporté mon caractère à travers toutes les difficultés que j'ai rencontrées durant cette thèse.

Un très grand merci aux collègues du laboratoire LabSTIC qui ont contribué, chacun à sa manière, par une bonne ambiance et par des discussions techniques aussi bien que philosophiques.

ملخص

من أجل تحسين السلامة والأمن في الأماكن العامة والخاصة، اتخذ كبار المسؤولين خطوات جذرية من خلال تنصيب المئات من كاميرات المراقبة في المدن الكبرى. و امام هذا العدد المتزايد من الكاميرات التي انبثقت من الرغبة في الحصول على أنظمة المراقبة و انخفاض تكلفة معدات التصوير، أصبحت إدارة محطات المراقبة و الرصد عبئاً ثقيلاً. في الواقع ، إنه من الصعب جداً، بل من المستحيل وضع عامل بشري وراء كل كاميرا. بالإضافة إلى ذلك، الطبيعة البشرية تجعل العمال يكفون في وجود المهام المتكررة مما يؤدي إلى انخفاض في اليقظة و ظهور عدة ثغرات أمنية. في هذا السياق، سوف نقوم بتصميم نظام مراقبة ذكي قادر على استغلال البنية التحتية القائمة و الكامرات الموجودة لمساعدة الانسان في أداء المهام المتكررة و المملة للمراقبة. إن الهدف من هذا النظام هو توفير حل شامل لمشكل المراقبة يكون مبني على أساس برامج ذكية و مستقلة تعمل على كشف، تصنيف و تتبع الأجسام المتحركة.

لقد اعتمدنا في الكشف عن الأجسام المتحركة على طريقة جديدة تتمثل في التعاون بين نموذج خليط جاوس و مراقبي الخلفية. من أجل التقليل من التباينات المحلية و التغيرات الآنية في الاضاءة، اقترحنا عدة استراتيجيات. لقد قمنا كخطوة اولى بتجميع تأثيرات الاضاءة في عنصر واحد و ذلك بتحويل فضاء تمثيل الألوان من RGB إلى HSV . للتغلب على مشاكل التغيرات المحلية، قمنا بتقسيم الصورة إلى مناطق متعددة و تعيين لكل منطقة مراقب خلفية الذي يقوم بالكشف عن التغيرات المتواجدة في مجال المنطقة المكلف بها من خلال إجراء مقارنة بين الرسم البياني المخزن في العمليات السابقة و الرسم البياني الخاص بالصورة التي يتم معاملةها. لمعالجة المعلومات آتياً، إستبدلنا خوارزمية EM بخوارزمية أخرى اقل تكلفة في العمليات الحاسوبية و تتمثل في خوارزمية K-means . يتم الكشف عن التغيرات المفاجئة في الضوء و حذفها عن طريق عملية طرح بين الصورة التي يجري تجهيزها و نموذج للخلفية شيد على أساس التدرجات. لمعالجة مشكلة الظلال التي تميل إلى أن تصنف على أنها جزء من الأجسام المتحركة اقترحنا عملية للكشف و إزالة الظلال الناتجة عن الأجسام المتحركة. الأسلوب المقترح يعتمد على التحليل الطيفي القائم على أساس ترشيح متكون من ثلاثة قياسات في الفضاءات RGB ، HSV و YUV . لقد استخدمنا كذلك مرشح Canny للفصل بين أنواع الظلال ذات الصلة بالأشياء المتحركة. الخطوة الأخيرة هي التحليل السلوكي للأجسام المتحركة في الفيديو. لقد اقترحنا نهجاً جديداً يجمع بين التصنيف، التعلم و الرصد لتحقيق أقصى قدر من الأداء. التعلم اون لاين زاد من الوثوقية في المصنف من خلال آلية متأقلمة و ديناميكية لإدارة التحول الذي قد يطرأ على شكل الأجسام المتحركة. تتكون وحدة التتبع من مرشح للجسيمات الذي يستخدم تقدير كالمان للتنبؤ بالمواقع المحتملة في المراحل القادمة للأجسام المتحركة. لقد اخترنا كوسيلة للتصنيف، خوارزمية KNN الذي يسمح بإجراء تعلم اون لاين لاعدادات المصنف. كما قمنا بتطبيق استراتيجية جديدة لضمان اتساق زمني و هيكلية و ذلك للحصول على افضل أداء لنظام التتبع. إن النتائج المتحصل عليها في جميع المناهج المقترحة مرضية إلى حد كبير بالمقارنة مع ما هو متداول في المجال.

الكلمات المفتاحية : الوسائط المتعددة البحث، المراقبة بالفيديو، التعرف على الأنماط، رؤية الآلة، التفاعل الآلي البشري، الكشف و المراقبة.

Abstract

In order to improve safety and security in public and private places, senior responsible have taken radical steps by implementing hundreds of surveillance cameras in major cities. Faced with this growing number of emerged cameras by a desire to acquire CCTV systems and the low cost camera equipment, management of monitoring stations has become a heavy burden. Indeed, it is very difficult, even impossible to place a human operator behind each camera. In addition, human nature grows weary in the presence of repetitive tasks making a decrease in attention and security vulnerabilities. It is in this context that we will design an intelligent video surveillance system able to exploit existing infrastructure and to assist human operators in repetitive and tedious tasks of monitoring. The target is to provide an intelligent software solution, independent and in real time to detect, classify and track moving objects. The detection of moving objects is based on a collaboration between the Gaussian Mixture Model and the background observers. To minimize local variations and instant brightness variations, we proposed several strategies. We first channeled the brightness effect in a single component by a transformation from RGB to the HSV color space. To overcome the problems of local variations, we segmented the image into several regions and each region is assigned to an observer which can detect changes by making a comparison between the histogram stored in the previous frames and that calculated in the frame being processed. To deal with the real time constraint, we replaced the EM algorithm with a new update process based on the K-means algorithm. Sudden changes in light are detected and removed by performing a subtraction operation between the image being processed and a background model constructed on the basis of gradients. To deal with the problem of shadows which tend to be classified as part of moving objects, we proposed a process to detect and eliminate shadows generated by moving objects. The proposed method uses a spectral analysis based on a filter composed of three metrics in the RGB color space, HSV and YUV. A canny filter was also used to separate between attached and cast shadows. The last step is a behavioral analysis of moving objects through video frames. We proposed a new approach that combines the classification, learning and tracking to achieve maximum performance. Online learning has increased the reliability of the classifier through an adaptive and dynamic mechanism for managing postures change of moving objects. The tracking module is a Particle Filters which uses the Kalman estimations to predict the next possible locations of moving objects. We chose as a classification method K nearest neighbors algorithm that allows by design, the ability to perform an online adaptive learning of classifier parameters. We have also implemented a strategy to ensure both a temporal and structural coherence to maximize the performance of the tracking system. The obtained results in the proposed methods are largely satisfactory compared to what exists in the state of the art.

Key-words: Multimedia-mining, CCTV, pattern recognition, machine vision, human machine interaction, detection and trackin

Résumé

Dans le but d'améliorer la sûreté et la sécurité dans les endroits publics et privées, les hauts dirigeants ont pris des mesures radicales en implantant des centaines de caméras de surveillance dans les grandes villes. Devant ce nombre croissant des caméras émergées par une volonté d'acquiescer des systèmes de vidéosurveillance et par le faible coût matériel des caméras, la gestion des stations de surveillance est devenue un lourd fardeau. En effet, il est très difficile, voire impossible, de placer un opérateur humain derrière chaque caméra. En plus, la nature humaine se lasse devant les tâches répétitives générant ainsi une baisse d'attention et des failles de sécurités. C'est dans ce cadre que nous allons concevoir un système de vidéosurveillance intelligent capable d'exploiter l'infrastructure existante de caméras et de seconder les opérateurs humains dans les tâches répétitives et fastidieuses de surveillance. L'objectif du système est d'offrir une solution logicielle intelligente, autonome et en temps réel qui permet de détecter, de classer et de suivre les objets en mouvement. La détection des objets en mouvement est basée sur un travail collaboratif entre les mixtures de gaussiennes et les observateurs de l'arrière-plan. Afin de minimiser les variations locales et les variations instantanées de luminosité, nous avons proposé plusieurs stratégies. Nous avons d'abord canalisé l'effet de la luminosité dans un seul composant par une transformation de l'espace de couleur RGB vers l'espace de couleur HSV. Pour surmonter les problèmes de variations locales, nous avons segmenté l'image en plusieurs régions ou chaque région est assignée à un observateur qui peut détecter des changements dans cette région en effectuant une comparaison entre l'histogramme mémorisé dans les frames précédents et celui calculé dans le frame en cours de traitement. Pour faire face à la contrainte de temps réel, nous avons remplacé l'algorithme EM par un nouveau processus de mise à jour basé sur l'algorithme K-means. Les variations soudaines de lumière sont détectées et supprimées par la réalisation d'une opération de soustraction entre l'image en cours de traitement et un modèle de l'arrière-plan construit sur la base des gradients. Pour faire face au problème des ombres qui ont tendance à être classées comme faisant partie des objets mobiles, nous avons proposé un processus qui permet de détecter et de supprimer les ombres portées générées par les objets en mouvement. La méthode proposée utilise une analyse spectrale basée sur un filtre composé de trois métriques dans les espaces de couleur RGB, HSV et YUV. Un filtre de canny a été également utilisé pour séparer entre les ombres portées et les ombres propres liées aux objets. La dernière étape est une analyse comportementale des objets en mouvement à travers les frames des vidéos. Nous avons proposé une nouvelle approche qui combine la classification, l'apprentissage et le suivi pour aboutir à une performance maximale. L'apprentissage en ligne a permis d'augmenter la fiabilité du classifieur grâce à un mécanisme adaptatif et dynamique pour la gestion du changement de postures des objets en mouvement. Le module de suivi est un filtre à particule qui utilise les estimations de Kalman pour prédire les prochains emplacements possibles des objets en mouvement. Nous avons choisi comme méthode de classification l'algorithme K plus proches voisins qui permet par conception, la possibilité d'effectuer un apprentissage en ligne adaptatif des paramètres du classifieur. Nous avons également mis en point une stratégie pour assurer à la fois une cohérence temporelle et structurelle afin de maximiser les performances du système de suivi. Les résultats obtenus dans toutes les méthodes proposées sont largement satisfaisants par rapport à ce qui existe dans l'état de l'art.

Mots-Clés : multimédia-mining, vidéosurveillance, reconnaissance des formes, vision par machine, interaction homme-Machine, détection et suivi.

Table des matières

ملخص	i
Abstract	ii
Résumé	iii
Table des matières	iv
Table des figures	x
Liste des tableaux	xii
Liste des abréviations	xiii
Introduction Générale	1
Introduction	1
Motivations	2
Domaine d'utilisation	3
Difficultés et problématiques	4
Contexte et objectifs	5
Solutions envisagées	5
Organisation de la thèse	7
État de l'Art	
Chapitre 1 Extraction de l'arrière-plan	9
1.1 Introduction	9
1.2 Modélisation par une seule gaussienne	9
1.3 Mixture de gaussiennes	11
1.4 Estimation de la densité basée sur un noyau	12
1.5 Estimation de la densité basée sur un noyau adaptatif	14
1.6 Modélisation récursive	15
1.7 Différence de frames	16
1.8 Soustraction par image de référence	17
1.9 Apprentissage sub-spatial basé ACP (SL-PCA)	18
1.10 Filtrage prédictif	20
1.11 Filtre de Kalman	21
1.12 Dictionnaire de mots visuels	21

1.13	Séparateurs à vaste marge	22
1.14	Les séparateurs à vastes marges pour la régression (SVR)	23
1.15	Les séparateurs à vastes marges pour la description de données	24
1.16	Simple gaussienne généralisée	26
1.17	Mixture de gaussienne généralisée	26
1.18	Conclusion	27
Chapitre 2 Détection et extraction d'ombre		28
2.1	Introduction	28
2.2	L'ombre	30
2.3	Propriété spectrale de l'ombre	31
2.4	Propriétés géométriques des ombres	33
2.5	Modélisation de la luminosité	34
2.6	Modélisation de l'ombre	34
2.7	Caractéristiques utilisées pour la détection de l'ombre	35
2.7.1	Intensité	35
2.7.2	Chromaticité	35
2.7.3	Les propriétés physiques	36
2.7.4	La géométrie	37
2.7.5	Les textures	37
2.7.6	Caractéristiques temporelles	38
2.8	Conclusion	39
Chapitre 3 Reconnaissance des formes		40
3.1	Introduction	40
3.2	Contexte et problèmes	41
3.3	Principe des RDFs	42
3.4	Les enjeux dans les RdFs	43
3.5	Description générale du système de reconnaissance	43
3.5.1	Prétraitement	43
3.5.2	Extraction des primitives (caractéristiques)	44
3.5.3	Apprentissage et Classification	45
3.6	Les approches de classification	46
3.6.1	Les approches statistiques	46
3.6.2	Les approches structurelles	47
3.6.3	Les approches hybrides	48
3.7	Nature des formes traitées	48
3.8	Méthodes de reconnaissance des formes	48
3.8.1	Réseaux de neurones	49
3.8.1.1	Le Perceptron de ROSENBLATT	49
3.8.1.2	Adaline de Windrow–Hoff	50
3.8.1.3	Le modèle de KOHONEN	50
3.8.1.4	Le Perceptron multicouche	50
3.8.1.5	Le modèle de HOPFIELD	50
3.8.2	Méthodes à noyau	51
3.8.3	Les arbres de décision	52
3.8.4	K plus proches voisins	53
3.8.5	K plus proches voisins flous	54

3.8.6	Les modèles de Markov cachés	54
3.8.7	AdaBoost	54
3.9	Conclusion	55
Chapitre 4	Suivi des objets en mouvement	56
4.1	Introduction	56
4.2	Représentation de l'objet	57
4.2.1	Les Points	57
4.2.2	La forme géométrique	58
4.2.3	La silhouette et le contour des objets	58
4.2.4	Les modèles de forme articulés	58
4.2.5	Les modèles basés sur le squelette	58
4.3	Les caractéristiques d'apparence des objets	58
4.3.1	Densités de probabilité de l'apparence de l'objet	58
4.3.2	Les modèles d'apparence	59
4.3.3	Les modèles d'apparence actifs	59
4.3.4	Les modèles d'apparence à multiples vues	59
4.4	Sélection des caractéristiques pour le suivi	60
4.4.1	Les couleurs	60
4.4.2	Les contours	61
4.4.3	Le flux optique	61
4.4.4	Les textures	61
4.5	Le suivi des objets en mouvement	62
4.5.1	Le suivi de Point	62
4.5.1.1	Les méthodes déterministes	63
4.5.1.2	Les méthodes statistiques	64
4.5.2	Le suivi à base de noyau	66
4.5.2.1	Appariement de modèle	67
4.5.2.2	Les modèles d'apparence basés densités	67
4.5.2.3	Les modèles d'apparence à multiples vues	68
4.5.2.4	Le suivi de silhouette	68
4.6	Conclusion	70
Proposition d'un Système de Vidéosurveillance		71
Chapitre 5	Détection des formes mobiles dans une vidéosurveillance	72
5.1	Introduction	72
5.2	La vidéosurveillance	73
5.2.1	Définition	73
5.2.2	Présentation	73
5.2.3	Historique	73
5.2.4	Objectifs majeurs	73
5.2.5	Domaines d'application	74
5.3	Le document vidéo	75
5.4	Caractéristiques d'une vidéo	75
5.4.1	Nombre de frames par seconde	75
5.4.2	Entrelacement	76

5.4.3	Résolution d’affichage	76
5.4.4	Proportion	77
5.5	Analyse du contenu de la vidéo	77
5.6	Modélisation du Contenu Vidéo	77
5.7	Les formats de fichiers vidéo numériques	79
5.8	Les images de références	79
5.9	Architecture du système proposée	81
5.9.1	Initialisation	81
5.9.1.1	Calibrage de la caméra	82
5.9.1.2	Adaptation des caractéristiques de la scène	82
5.9.1.3	Initialisation des modèles des objets	82
5.9.2	Capture spatio-temporelle de l’information	83
5.9.3	Détection des objets en mouvement	83
5.9.4	Détection et suppression des ombres portées	83
5.9.5	Classification, Reconnaissance et suivi	84
5.9.6	Analyse comportementale	84
5.10	Conclusion	85
Chapitre 6	Analyse Spatio-Temporelle	86
6.1	Panoramas	86
6.2	Travaux similaires	87
6.3	Taxonomie proposée	89
6.4	Motivations	89
6.5	Proposition d’un système d’extraction d’arrière-plan	89
6.6	Acquisition	90
6.6.1	Prétraitement	91
6.6.2	Segmentation	91
6.7	Traitement	92
6.7.1	Le modèle de mixture de gaussiennes	92
6.7.1.1	Initialisation	93
6.7.1.2	Mise à jour des paramètres	93
6.7.1.3	Extraction de l’arrière-plan	95
6.7.2	Les observateurs de l’arrière-plan	96
6.7.2.1	Bhattacharyya distance	97
6.7.2.2	Divergence de Kullback-Leibler	97
6.7.2.3	Divergence de Jensen-Shannon	97
6.8	Détection	99
6.8.1	Post-traitement	99
6.8.1.1	Application des opérations morphomathématiques	100
6.8.1.2	Élimination des changements soudains de luminosité	101
6.8.2	Localisation des objets en mouvement	103
6.9	Conclusion	103
Chapitre 7	Analyse Spectrale	105
7.1	Introduction	105
7.2	Motivations	105
7.3	Taxonomie	106
7.4	Évaluation générale et commentaires	106

7.5	Architecture proposée	107
7.6	Détection de contours : Filtre de Canny	108
7.6.1	Principe général	108
7.6.2	Les étapes de l’algorithme de Canny	108
7.6.2.1	Le lissage	108
7.6.2.2	Le gradient	109
7.6.2.3	Suppression des valeurs non maximales	109
7.6.2.4	Seuillage	109
7.6.2.5	Détection de contour par décalage	110
7.7	Dilatation	110
7.8	Conversion de l’espace de couleur	111
7.9	Détection et élimination des ombres	112
7.10	Conclusion	114
Chapitre 8 Analyse comportementale		115
8.1	Introduction	115
8.2	Suivi des objets en mouvement	116
8.3	Classification et reconnaissance des formes	117
8.3.1	Détection des formes	117
8.3.2	Apprentissage automatique	118
8.3.2.1	Esperance-Maximisation (EM)	119
8.3.2.2	L’autoapprentissage	119
8.3.2.3	Le co-apprentissage	119
8.4	Travaux similaires	120
8.5	Analyse intrinsèque du comportement	120
8.6	Propositions et Architectures	121
8.6.1	Extraction des caractéristiques	122
8.6.1.1	Calcul du centre de gravité	123
8.6.1.2	Calcul de l’histogramme de couleur	123
8.6.1.3	Extraction de la silhouette de l’objet	123
8.6.2	Classification	124
8.6.3	Apprentissage	124
8.6.4	Suivi d’objets	125
8.6.4.1	Modélisation et initialisation	125
8.6.4.2	Evolution linéaire du vecteur d’état	126
8.6.4.3	Mesure de l’observation	126
8.6.4.4	Statut du suivi d’un objet	126
8.7	Cohérence spatio-temporelle	127
8.8	Cohérence structurelle	127
8.9	Conclusion	128
Chapitre 9 Expérimentations et résultats		129
9.1	Introduction	129
9.2	Procédé de validation	129
9.3	Vérité-terrain	130
9.4	Corpus de données	130
9.5	Évaluation quantitative	132
9.5.1	Notations standards	132

TABLE DES MATIÈRES

9.5.2	Métriques d'évaluation	132
9.5.3	Mesures ajoutées pour l'extraction de l'arrière-plan	133
9.5.4	Mesures ajoutées pour la détection des ombres	134
9.5.5	Modalité d'utilisation de la vérité-terrain	134
9.6	Résultats relatifs à l'extraction de l'arrière-plan	135
9.6.1	Configuration des paramètres	135
9.6.2	Résultats qualitatifs	135
9.6.3	Résultats quantitatifs	139
9.7	Résultats relatifs à la détection des ombres	144
9.7.1	Résultats qualitatifs	144
9.7.2	Résultats quantitatifs	152
9.8	Résultats relatifs à la classification et le suivi	154
9.9	Conclusion	156
	Conclusion Générale	157
	Bibliographie	161
	Publications de l'auteur	185

Table des figures

1.1	Modèle réursive	15
2.1	Résultats du suivi des objets en mouvement avec et sans la suppression de l'ombre	29
2.2	Représentation des types d'ombres	30
2.3	Représentation des caractéristiques géométriques de l'ombre	33
3.1	Taxonomie des méthodes de reconnaissance des formes	42
3.2	Schéma général d'un Système de Reconnaissance des formes	44
5.1	Structure d'une vidéo	78
5.2	Architecture du système de vidéosurveillance proposée	81
6.1	Taxonomie des systèmes d'extraction de l'arrière-plan	88
6.2	Architecture du module de détection et d'extraction des formes en mouvement	90
6.3	Exemple de modélisation par une mixture de gaussiennes	96
6.4	Résultat de l'extraction de l'arrière-plan	96
6.5	L'effet de la variation locale sur les performances des GMM	98
6.6	Résultat des opérations morphomathématiques sur les objets	101
6.7	Changement soudain de la lumière	102
6.8	Opération d'étiquetage et de cadrage des composantes connexes	103
7.1	Architecture du système proposé	107
7.2	Processus de Canny	108
7.3	Resultats de la detection de countours	110
8.1	Processus dynamique et adaptatif de suivi	122
8.2	Schéma du processus d'apprentissage et de décision	125
9.1	Résultats de la soustraction de l'arrière-plan dans la vidéo B7	135
9.2	Résultats de la soustraction de l'arrière-plan dans les vidéos A1 et A3	136
9.3	Résultats de la soustraction de l'arrière-plan dans les vidéos A5 et A6	137
9.4	Résultats de la soustraction de l'arrière-plan dans les vidéos D3-1 et D4-2	138
9.5	Résultats de la soustraction de l'arrière-plan dans les vidéos D1-1, D1-2, D1-3 et D2-2	139
9.6	Résultats de l'évaluation des performances en termes de Precision	141
9.7	Résultats de l'évaluation des performances en termes de Rappel	141
9.8	Résultats de l'évaluation des performances en termes F-mesure.	142
9.9	Résultats de la détection de l'ombre dans la vidéo D4-2 de la base DBD	145
9.10	Résultats de la détection de l'ombre dans la vidéo D3-2	146

TABLE DES FIGURES

9.11	Résultats de la détection de l'ombre dans la vidéo D1-3	147
9.12	Résultats de la détection de l'ombre dans la vidéo D2-1	148
9.13	Résultats de la détection de l'ombre dans la vidéo D1-1	149
9.14	Résultats de la détection de l'ombre dans la vidéo A5	150
9.15	Résultats de la détection de l'ombre dans la vidéo A2	151
9.16	Résultats du suivi sur des scènes monocible	154
9.17	Résultats du suivi sur des scènes multicibles	155

Liste des tableaux

5.1	Description des principaux formats vidéo	80
9.1	Corpus de tests détaillé	131
9.2	Résultats obtenus par les critères SE, AC, SP et PT	140
9.3	Résultats obtenus par les critères Recall, Precision et F-Score	143
9.4	Évaluations comparatives des résultats obtenus	152
9.5	Comparaison de la méthode proposée avec la méthode LR	153
9.6	Le temps de traitement calculé en millisecondes par image	153

Liste des Abréviations

ACI	Analyse en Composante Indépendante
ACP	Analyse en Composantes Principales
AKDE	Adaptive Kernel Density Estimation
BG	BackGround
CART	Classification And Regression Tree
CCTV	Closed Circuit Television Network
CODECS	Codeurs/Décodeurs
DVD	Digital Video Disc
DVR	Digital Video Recorder
EM	Espérance-Maximisation
EM	Expectation Maximisation
FAR	False Alarm Ratio
FCM	Fuzzy C-Means
FPM	Fuzzy Pattern Matching
FPS	Frames Per Second
GGMM	Generalized Gaussian Mixture Model
GMM	Gaussian Mixture Model
HSL	Hue Saturation Lightness
HSV	Hue Stauration Value
KDE	Kernel Density Estimation
KNN	k-Nearest Neighbors
KPPV	k-Plus Proches Voisins
LCD	Liquid Crystal Display
LMM	Longueur de Message Minimum
ML	Maximum Likelihood
MoG	Mixtures of Gaussians
NTSC	National Television System Committee

PAL	Phase Alternating Line
RdF	Reconnaissance des Formes
RGB	Red Green Blue
SECAM	SÉquentiel Couleur À Mémoire
SGG	Single General Gaussian
SL-PCA	Sequential Logistic Principal Component Analysis
SVDD	Support Vector Data Description
SVM	Support Vector Machines
SVR	Support Vector Regression
TFT	Thin-Film Transistor
VBR	Variable Bit Rate
VT	Vérité-Terrain

Introduction Générale

Introduction

Les systèmes de vidéosurveillance ont été utilisés depuis longtemps comme un outil de contrôle et de surveillance. Ils jouent un rôle de plus en plus important dans la surveillance de sites sensibles ou de lieux publics et privés. L'histoire de la vidéosurveillance a passé par trois principales générations appelées 1GSS, 2GSS et 3GSS.

La première génération des systèmes de vidéosurveillance existait entre les années 60 jusqu'aux années 80 et ils étaient basés sur un traitement analogique de l'acquisition, la transmission et le traitement de l'information. Le rôle principal de ces systèmes est d'amplifier les sens de vision de l'être humain par la transmission des vidéos de plusieurs caméras qui surveillent plusieurs endroits dans une seule chambre (chambre de contrôle) et sur un seul moniteur. Parmi les problèmes majeurs, dans un tel système, on trouve : le besoin d'une très grande bande passante pour le transfert de l'information et qui s'accroît de manière exponentielle avec le nombre de caméras ; la difficulté d'archiver et de retrouver en cas de problème un événement à cause de la quantité énorme de vidéo et enfin la fragilité des observateurs humains qui ne peuvent malheureusement pas rester figer avec une très grande concentration sur le moniteur pendant une longue période.

La deuxième génération est venue pour pallier un peu les problèmes de la première génération en utilisant des systèmes hybrides basés à la fois sur le signal analogique et numérique. Cette hybridation a permis d'introduire quelques opérations qui jouaient un peu le rôle d'un assistant aux agents de surveillance.

La troisième génération, apparue depuis les années 2000, est totalement basée sur un système digital allant de la caméra vers les systèmes de traitement de l'information. Elle a permis : une acquisition et un prétraitement au niveau des capteurs ; une communication via les réseaux mobiles et les réseaux internet ; et enfin une masse de stockage énorme grâce aux méthodes de compression et aux partages des données sur plusieurs serveurs. L'atout principal de cette génération est l'utilisation de caméras hyper sophistiquées pour réaliser plusieurs traitements de l'image telles que la discrétisation du signal analogique vers un signal numérique compréhensible par la machine et une compression des données de bas niveau. Les caméras équipées de systèmes embarqués ont pu aller encore plus

loin en permettant par exemple de détecter les mouvements et les visages. Cette troisième génération a été conçue pour permettre un déclenchement des alarmes en temps réel et une recherche efficace des informations hors ligne. Et c'est pour cette raison que le défi des systèmes de troisième génération était d'offrir des systèmes intelligents capable de générer en temps réel des alarmes traitant des événements très complexes et de sauvegarder ces derniers pour une éventuelle recherche d'information. La construction de tels systèmes nécessite des algorithmes rapides robustes et efficaces pour la détection, la classification, l'analyse et le suivi des activités des objets en mouvement. Depuis la dernière décennie, un nombre très important de recherches ont été consacrées pour le développement des algorithmes dédiés aux systèmes de vidéosurveillance intelligents.

Motivations

Dans le cadre de la sécurisation des endroits publics, les responsables ont mis en place une infrastructure constituée de centaines de caméras de surveillance dans les grandes villes. Les infrastructures existantes des systèmes de surveillance actuels offrent seulement la possibilité de capturer, de sauvegarder et de distribuer les vidéos afin de les traiter manuellement par des policiers ou par des agents spécialistes dans la détection des anomalies et des comportements suspects. Devant ce nombre croissant de caméras, il est devenu très difficile, voire impossible, de placer un opérateur humain derrière chacune d'elles. En plus, la demande et le besoin d'outils d'analyse automatique des données récupérées ont fortement augmenté. En effet, la répétitivité de la tâche et le faible nombre d'évènements ou de situations anormales entraînent un épuisement et baisse l'attention des agents de sécurité. Donc, un agent ne pourra pas rester devant des écrans de surveillance sans aucune faille. Le faible coût matériel des caméras a fortement contribué dans la volonté d'acquisition des systèmes de vidéosurveillance intelligents et de les utilisés dans une large gamme d'applications à travers des fonctionnalités telles que la reconnaissance et le suivi automatique d'objets, l'interprétation des comportements et l'indexation d'évènements particuliers.

La détection des objets en mouvement est la principale tâche dans les systèmes de traitement de la vidéo. Elle permet de séparer les objets en mouvement des objets stationnaires de l'arrière-plan. Cela ne permet pas uniquement de se focaliser sur les traitements de haut niveau, mais également de diminuer le temps d'exécution de manière considérable. Les techniques les plus communément utilisées pour la détection des objets en mouvement sont : l'extraction de l'arrière-plan, les modèles statistiques, la différence temporelle et le flot optiques. À cause de la nature dynamique des environnements tels que le changement de la luminosité, l'effet de l'ombre, la présence de poussière, le mouvement des branches d'arbres et le mouvement des caméras, la segmentation des objets est devenue une tâche difficile et un problème significatifs qui nécessitent une prise en

main de grand intérêt pour la construction d'un système de vidéosurveillance robuste.

L'étape de classification et de reconnaissance des objets nécessite généralement des classes définies préalablement telles que les humains, les animaux, les véhicules, etc. Il est nécessaire de séparer entre les objets en mouvement détecté pour le suivi et pour l'analyse de comportement. On distingue deux techniques pour la classification des objets en mouvements à savoir : les méthodes basées sur la forme de l'objet et les méthodes basées sur la nature du mouvement de l'objet. La première utilise les informations spatiales des objets tandis que la deuxième utilise les caractéristiques temporelles pour la classification des objets.

L'étape suivante dans l'analyse des vidéosurveillances est le suivi des objets qui permet de définir un critère de correspondance temporelle parmi les autres objets en mouvement détectés. Cette procédure produit une identification temporelle des régions segmentées et génère un ensemble d'information concernant l'objet tel que : la région de placement, la trajectoire de l'objet, la vitesse de déplacement ainsi que la direction. Cette étape produit une sortie qui sera généralement utilisée comme un support pour l'analyse des activités de haut niveau.

La dernière étape dans les systèmes de vidéosurveillance intelligents est de reconnaître les comportements des objets et générer une interprétation sémantique pour chaque comportement. En d'autres termes, c'est une classification temporelle des activités de l'objet selon un étiquetage de référence.

Domaine d'utilisation

Un système de vidéosurveillance offre aux agents responsables de la surveillance des informations de haut niveau pour simplifier la prise de décision dans un temps très réduit et pour offrir, en cas d'accident, une plateforme de recherche effective des vidéos enregistrés pour un traitement hors ligne. Parmi les secteurs susceptibles d'utiliser les systèmes de vidéosurveillance intelligente :

La sécurité publique et commerciale :

- La surveillance des banques, les magasins de stockage, les aéroports, les musées, les stations de service, les propriétés privées et les parkings pour la prévention des crimes ;
- La surveillance des autoroutes et des rails pour la détection des accidents ;
- La surveillance des maisons, des magasins et des forêts pour la détection des incendies ;
- L'observation des activités des personnes âgées et handicapées pour garantir un traitement médical en urgence ;
- Les contrôles d'accès aux endroits sécurisés.

Exploitation des données :

- Mesurer la circulation des voitures, l'encombrement des piétons et des voitures et parfois estimer les performances dans le domaine sportif ;
- Calculer des statistiques pour étudier la quantité de population présente dans les magasins et les Park de jeux ;
- Extraire des statistiques des activités sportives ;
- Compter les espèces menacées dans les réserves nationales ;
- Journaliser les tâches de maintenances des installations dans les usines.

Maintien de l'ordre public :

- Mesurer la vitesse des véhicules dans les autoroutes ;
- Détecter les infractions des feux rouges et le dépassement de la ligne continu.

Sécurité militaire :

- La surveillance des frontières ;
- Mesurer la circulation des réfugiés ;
- Fournir une région de sécurité aux alentours des bases militaires ;
- Assister les commandes et le contrôle de bataille.

Les utilisations des algorithmes de détection, de suivi et de reconnaissance des objets en mouvement ne sont pas limitées uniquement à la vidéosurveillance. En effet, plusieurs domaines bénéficient pleinement de ces avancements technologiques dans la réalité virtuelle, la réalité augmentée, la compression de vidéo et l'interaction homme-machine.

Difficultés et problématiques

La réalisation d'un système de vidéosurveillance performant est régie par de nombreuses contraintes d'ordre techniques, sémantiques et environnementales. Les contraintes techniques sont principalement liées aux matériels utilisés tels que la résolution, le taux de rafraîchissement et l'emplacement de la caméra. Les contraintes environnementales sont liées à la nature de la scène et aux changements climatiques de l'environnement. Les contraintes sémantiques se focalisent sur le contexte dans lequel le système de vidéosurveillance est implanté. La présence de toutes ces contraintes engendre des problèmes à plusieurs niveaux et à des degrés divers.

Dans la pratique, le matériel ainsi que ses caractéristiques sont imposés par un réseau de caméras déjà existant. Pour cela, les algorithmes proposés doivent faire face aux déformations des objets, aux occlusions, à la fusion d'un ou plusieurs objets engendrée par les emplacements des caméras, à la résolution, au taux de rafraîchissement déjà fixé, aux problèmes de luminosité, de camouflage, etc.

Contexte et objectifs

Nous allons concevoir un système de vidéosurveillance intelligent capable d'exploiter l'infrastructure existante de caméras et de seconder les opérateurs humains dans les tâches répétitives et fastidieuses de surveillance. L'objectif du système est d'offrir une solution logicielle intelligente, autonome et en temps réel en garantissant une certaine flexibilité, afin d'être facilement déployable dans d'autres configurations possibles. En plus des tâches de base citées précédemment, le système doit être capable de fournir des statistiques sur l'état de l'environnement pour une éventuelle analyse sémantique des événements détectés par le système.

Solutions envisagées

Nous présentons dans cette thèse un système de vidéosurveillance intelligent qui permet de détecter, de classer et de suivre les objets en mouvement. Le système utilise en entrée des vidéos en couleur capturée à partir d'une caméra fixe.

Dans le système proposé, les objets en mouvement sont détectés en utilisant une approche basée sur l'extraction de l'arrière-plan qui fonctionne dans les environnements intérieurs et extérieurs. Nous présentons également d'autres systèmes de détection des objets en mouvement tels que la différence temporelle et l'estimation de la densité basée noyaux pour une éventuelle étude comparative. Nous avons proposé un module d'extraction de l'arrière-plan basé sur la coopération entre les mixtures de gaussiennes et les observateurs de l'arrière-plan. Nous avons d'abord fait un passage de l'espace de couleur RGB vers l'espace HSV. Cette conversion a permis de canaliser la luminosité dans le seul composant (V) et de diminuer les problèmes causés par les variations instantanées de la luminosité. Pour surmonter les problèmes de variations locales, l'image initiale de la vidéo est segmentée en plusieurs régions ou chaque région est assignée à un observateur qui peut détecter des changements dans cette région en effectuant une comparaison entre l'histogramme mémorisé dans les frames précédents et celui calculé dans le frame en cours de traitement. Le calcul de la similarité est effectué via le coefficient de Bhattacharyya et seuls les segments qui ont subi un changement significatif sont mis à jour par le GMM. Pour faire face à la contrainte de temps réel, nous avons remplacé l'algorithme EM par un nouveau processus de mise à jour basé sur l'algorithme K-means. Pour prendre en compte les variations soudaines de lumière, nous avons proposé une détection basée sur les gradients. La détection et la suppression de ces variations s'effectuent par une différence entre l'image du modèle et l'image en cours de traitement.

Les pixels de l'avant-plan détectés contiennent des bruits générés par des erreurs d'acquisition, des mouvements non significatifs tels que les branches d'arbres, le changement de luminosité et la présence d'une similitude entre l'objet en mouvement et l'arrière-plan.

Pour éliminer ces pixels isolés, des filtres morpho-mathématiques tels que l'érosion et la dilatation sont utilisés. Après l'étape de post-traitement, les pixels sont regroupés et étiquetés par un algorithme d'étiquetage pour créer des composantes connexes de mouvement. À la base de ces composantes, une autre étape de filtrage est effectuée afin d'éliminer les segments de petite taille. Dans l'étape suivante, un cadrage est effectué sur les composantes connexes susceptibles de représenter un objet en mouvement. Chaque cadre sera représenté par : les coordonnées, la zone dans laquelle il se trouve, le centre de gravité et l'histogramme de couleur qui seront utilisés dans les étapes suivantes.

Les algorithmes de détection et d'extraction des objets en mouvement ont généralement un inconvénient majeur, les ombres ont tendance à être classées comme faisant partie du premier plan conduisant ainsi à une confusion entre l'objet et son ombre. Effectivement, les ombres partagent les mêmes habitudes de déplacement et ont une ampleur similaire au changement d'intensité que celle des objets mobiles, ce qui influence sur le rendement des systèmes de vidéosurveillance. La complexité de ce problème rend les post-traitements conventionnels impuissants face à ce genre de situation. Pour cela, nous avons proposé un module qui permet de détecter et de supprimer les ombres portées engendrées par les objets en mouvement. La méthode proposée utilise une analyse spectrale pour détecter les ombres en se basant sur les caractéristiques chromatiques de trois espaces de couleurs. Nous avons proposé un filtre construit sur la base de trois métriques sur les espaces de couleur RGB, HSV et YUV. Un filtre de Canny a été également utilisé pour séparer entre les ombres portées et les ombres propres liées aux objets.

Après l'extraction des objets en mouvement et le regroupement des pixels en un ensemble de composantes connexes et la suppression des ombres portées, un processus de suivi est appliqué afin d'analyser le comportement des objets en mouvement. Pour aboutir à un système de suivi performant, nous avons proposé une combinaison entre la classification, l'apprentissage et le suivi. L'apprentissage en ligne, avec les données étiquetées et les données non étiquetées, a permis d'augmenter la fiabilité du classifieur grâce à un mécanisme adaptatif et dynamique pour la gestion du changement de postures des objets en mouvement. Le module de suivi est un filtre à particule qui utilise les estimations de Kalman pour prédire les prochains emplacements possibles des objets en mouvement. Nous avons choisi comme méthode de classification l'algorithme K plus proches voisins qui permet par conception, la possibilité d'effectuer un apprentissage en ligne adaptatif des paramètres du classifieur. Afin d'alléger le processus de suivi, nous avons utilisé les caractéristiques géométriques comme un critère de discrimination des objets en mouvement. Ces primitives se résument dans le calcul du centre de gravité, l'histogramme de couleur, la taille de la composante connexe et l'utilisation des contours de l'objet. Nous avons également mis en point une stratégie pour assurer à la fois une cohérence temporelle et une cohérence structurelle afin de maximiser les performances du système de suivi. La cohérence structurelle est assurée par l'analyse des histogrammes de contour tandis que

l'analyse temporelle est assurée par un travail collaboratif entre le module de détection, le classifieur et le module de suivi.

Organisation de la thèse

Cette thèse est divisée en neuf chapitres et s'organise de la façon suivante :

Le chapitre 1 est une étude approfondie des méthodes d'extraction de l'arrière-plan selon plusieurs aspects et sous différents angles.

Le chapitre 2 dresse un état de l'art sur les méthodes les plus communément utilisées pour la détection et la suppression des ombres.

Le chapitre 3 présente une collection des méthodes utilisées pour la classification et la reconnaissance des formes en général et plus particulièrement les classifieurs utilisés dans les systèmes de vidéosurveillance.

Le chapitre 4 est une présentation détaillée des méthodes utilisées dans l'état de l'art pour le suivi des objets en mouvement.

Le chapitre 5 décrit l'architecture générale de notre système de détection, de reconnaissance et de suivi des formes mobiles dans une vidéo. Il présente également le contexte général des systèmes de vidéosurveillance ainsi que les caractéristiques d'un document vidéo.

Le chapitre 6 est une analyse spatiotemporelle dans laquelle nous avons d'abord proposé une nouvelle taxonomie des méthodes d'extraction de l'arrière-plan ensuite, nous avons proposé un processus de détection et d'extraction des objets en mouvement.

Le chapitre 7 traite le problème des ombres portées engendrées par les objets en mouvement. Il présente les différents techniques proposés afin de détecter et de supprimer l'ombre.

Le chapitre 8 est une analyse comportementale qui contient les algorithmes proposés pour la classification, la reconnaissance et le suivi des objets mobiles dans la scène.

Le chapitre 9 est dédié à l'expérimentation. Il présente les différentes bases vidéo utilisées, les métriques d'évaluation proposées et les résultats obtenus par rapport aux méthodes proposées. Il contient également une interprétation approfondie des résultats qualitatifs et quantitatifs obtenus et une étude comparative avec quelques systèmes présentés dans l'état de l'art.

Nous allons terminer cette thèse par une conclusion et quelques perspectives dans le domaine de la vision en général et particulièrement les systèmes de vidéosurveillance.

Première partie

État de l'Art

1

Extraction de l'arrière-plan

1.1 Introduction

Différentes applications telles que la vidéo surveillance, la capture de mouvement et la multimédia ont recours à la modélisation de l'arrière-plan avant d'entamer la détection des objets en mouvement. L'une des manières d'obtenir l'arrière-plan est de prendre une image de l'arrière-plan dépourvu d'objet en mouvement. Cependant, l'existence ou la disponibilité de cet arrière-plan n'est pas toujours possible. De plus, l'arrière-plan supposé statique peut subir des modifications importantes par l'ajout ou la suppression d'un objet ou tout simplement par un changement d'éclairage.

Les difficultés rencontrées dans ce type de modélisation se regroupent dans deux principaux axes : les problèmes causés par la dynamique de l'arrière-plan et ceux causés par les changements de luminosité. Plusieurs méthodes ont été proposées afin de résoudre les problèmes liés à l'extraction et à la modélisation de l'arrière-plan. Nous allons dans ce qui suit présenter les méthodes les plus communément utilisées.

1.2 Modélisation par une seule gaussienne

La modélisation par une gaussienne a été initialement proposée par Wren et al. en 1997 [WRE 97]. Le modèle est basé sur l'utilisation d'une fonction de densité de probabilité gaussienne sur les n dernières valeurs des pixels. Pour cela, les auteurs se sont basés sur l'hypothèse que les pixels d'arrière-plan sont indépendamment distribués selon une

distribution gaussienne. De cette façon, la méthode a permis de prendre en compte les faibles variations d'intensité lumineuse considérées comme étant des bruits de mesures dans le modèle.

Le principe de la méthode est très simple, sachant qu'une gaussienne est caractérisée par une moyenne et une variance, cette dernière peut facilement modéliser l'arrière-plan. Donc, la moyenne de la distribution permet de caractériser la couleur dominante prise par le pixel et la variance permet de caractériser la variabilité autour de cette valeur. Le modèle propose également une mise à jour régulière des paramètres de moyenne et de variance pour prendre en compte les changements de luminosité. La mise à jour s'effectue de façon récursive par le biais du paramètre α appelé le taux d'apprentissage ou régulateur de la vitesse d'adaptation. L'arrière-plan et sa variance sont estimés, pour chaque image, selon les équations 1.1 et 1.2 suivantes :

$$\mu_{t+1} = (1 - \alpha)\mu_t + \alpha X_t \quad (1.1)$$

$$\sigma_{t+1}^2 = (1 - \alpha)\sigma_t^2 + \alpha(X_{t+1} - \mu_{t+1})(X_{t+1} - \mu_{t+1})^T \quad (1.2)$$

Où X_{t+1} est la valeur du pixel courant, μ_t est la valeur ancienne de la moyenne, σ_t est l'ancienne valeur de la variance et α est le taux d'apprentissage du modèle.

La détection des objets de l'avant-plan est assurée en fonction de la position du pixel par rapport à la moyenne. Donc, l'image de l'avant-plan est construite en choisissant les pixels d'arrière-plan comme étant ceux suffisamment éloignés de la valeur moyenne. Généralement, la valeur de la variance est directement exploitée dans la décision de la façon suivante :

$$\frac{|X_{t+1} - \mu_{t+1}|}{\sqrt{\sigma_{t+1}}} < 2.5 \quad (1.3)$$

Afin d'améliorer ce modèle, Francois et al. [FRA 99] ont exploité la représentation HSV à la place de l'espace de couleur RGB. L'utilisation de l'espace HSV est plus robuste dans les changements d'illumination progressive, car il permet de séparer entre les intensités lumineuses concentrées dans l'unique composant V et entre les propriétés chromatiques représentées par les deux composants H et S. En plus, l'espace de couleur HSV permet de traiter partiellement le problème de camouflage. Les travaux proposés par Zhao et al. [ZHA 02b ; ZHA 02a] ont également montré que l'utilisation de l'espace de couleur HSV donne une meilleure performance dans la présence de l'ombre et des changements d'illumination progressive. Le modèle est très simple à implémenter, consomme une faible mémoire et a une très grande vitesse d'exécution.

Les améliorations sur la modélisation par une simple gaussienne n'ont cessé de croître ces dernières décennies [HON 03 ; KAM 05 ; ABD 06 ; CHE 07 ; PAN 08 ; RAM 09 ; LOP 11] pour converger vers la perfection et la liste n'est pas exhaustive. Cependant, ce qu'on peut remarquer à travers toutes ces contributions est que dans la majorité d'entre elles, les amé-

liorations ne traitent qu'une seule problématique. Ainsi, les modèles proposés sont limités aux objectifs des auteurs dans des environnements bien spécifiques et avec des contraintes bien définies. En plus, les systèmes basés sur une modélisation par une seule gaussienne ne sont performants que dans les environnements fermés avec des changements modérés de luminosité.

1.3 Mixture de gaussiennes

L'utilisation d'une seule gaussienne pour la modélisation d'un pixel n'est pas pratique dans la mesure où l'arrière-plan subit une dynamique instance. En effet, la variance des pixels devient tellement rapide que la gaussienne n'arrive pas à mémoriser les anciens états du pixel. Cela nécessite une migration vers la multimodalité afin de mieux gérer la nature dynamique de la plupart des environnements. Le premier modèle des mixtures de gaussienne a été initialement proposé par Friedman et Russel [FRI 97] pour la surveillance du trafic routier. Ils proposent de modéliser chaque pixel de l'image avec une mixture de trois gaussiennes.

L'initialisation du modèle et l'estimation des paramètres sont effectués à l'aide d'un algorithme de type Expectation Maximisation (EM) . Pour plus de vitesse dans l'exécution du modèle, une version récursive de l'algorithme EM est proposée, dont les approximations ont été identifiées par Power et al. [POW 02]. Cependant, la version la plus communément utilisée a été proposée par Stauffer et Grimson [STA 99b] qui ont généralisé cette idée en modélisant les récentes valeurs de couleur prises par chaque pixel avec une mixture de K gaussiennes.

Pour une éventuelle utilisation en temps réel, Ils proposent également de remplacer l'algorithme EM par l'algorithme K-means pour la mise à jour des paramètres des gaussiennes. Cette méthode décrit la probabilité d'observer la valeur d'un pixel P_t dans le temps t suivant l'équation 1.4 suivante :

$$P(P_t) = \sum_{i=1}^k w_{i,t} \cdot \eta(P_t, \mu_{i,t}, \Sigma_{i,t}) \quad (1.4)$$

Où

k : est le nombre de gaussiennes associées à chaque pixel.

$w_{i,t}$: Le poids calculé pour la $i^{\text{ème}}$ gaussienne dans le temps t .

$\mu_{i,t}$: La moyenne de la $i^{\text{ème}}$ gaussienne dans le temps t .

La version originale des mixtures de gaussiennes présente plusieurs avantages. En effet, elle peut travailler sans avoir recours à stocker un ensemble important de données d'entrée dans le processus en cours. La multi modalité du modèle permet de traiter les

environnements multimodaux et les environnements avec des changements d'éclairages graduels. Malgré cela, ce modèle présente certains inconvénients : le nombre de gaussiennes doivent être prédéterminé, la nécessité d'une bonne initialisation, la dépendance des résultats sur la vraie loi de distribution qui peut être non gaussienne, le problème de la reprise en cas de défaillances et le besoin d'une suite d'image ne contenant aucun objet en mouvement pour l'initialisation des paramètres.

1.4 Estimation de la densité basée sur un noyau

L'estimation basée sur un noyau KDE (Kernel Density estimation) est une technique d'estimation de la densité de probabilité à partir d'un ensemble d'échantillons sans aucune hypothèse sur la forme de la distribution dont ils sont issus.

Pour faire face aux milieux dynamiques, le mouvement de la caméra, les agitations des arbres et le mouvement de l'eau, Elgammal et al. [ELG 00] ont proposé d'estimer la fonction de densité de probabilité pour chaque pixel en utilisant l'estimateur de noyau K pour les N récents échantillons des valeurs d'intensité prises consécutivement dans des intervalles de temps I selon l'équation 1.5 :

$$P(X_t) = \frac{1}{N} \sum_{i=1}^N K(X_t - X_i) \quad (1.5)$$

Où K est la fonction d'estimation du noyau qui est prise comme une fonction gaussienne normale $N(0, \Sigma)$. X est un vecteur de couleur de dimension $d = 3$. Donc, la fonction de densité de probabilité est déterminée par l'équation suivante :

$$P(X_t) = \frac{1}{N} \sum_{i=1}^N \frac{1}{2\pi^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} \exp^{-\frac{1}{2}(X_t - X_i)^T \Sigma^{-1} (X_t - X_i)} \quad (1.6)$$

La largeur du noyau est estimée à partir de la valeur médiane m issue de la différence entre les images consécutives. Elgammal et al. [ELG 00] supposent que les canaux de couleur sont totalement indépendants avec différentes largeurs de bande du noyau, la largeur de bande de la fonction du noyau est la suivante :

$$\Sigma = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix}$$

Donc, la fonction de densité de probabilité peut être écrite selon l'équation 1.7 suivante :

$$P(X_t) = \frac{1}{N} \sum_{i=1}^N \prod_{c=1}^d \frac{1}{\sqrt{2\pi\sigma_c^2}} \exp - \left(\frac{(X_{t,c} - X_{i,c})^T}{2\sigma_c^2} \right) \quad (1.7)$$

La détection des objets en mouvement est assurée par l'équation suivante :

$$P(X_t) < T \quad (1.8)$$

T étant un seuil fixé de manière empirique. Le pixel est considéré comme premier plan si l'équation 1.8 est vraie. Sinon le pixel est classé comme faisant partie de l'arrière-plan.

Pour que le modèle reste fonctionnel, une mise à jour des paramètres est requise. Pour cela, Elgammal et al. [ELG 00] ont proposé deux modèles de l'arrière-plan pour la mise à jour des paramètres. Le premier modèle est un modèle de court terme. Il permet une adaptation très rapide pour permettre les détections sensibles. Ce modèle utilise les N valeurs des échantillons de l'arrière-plan les plus récentes. L'échantillon est mis à jour en utilisant un mécanisme de maintenance sélective à base de la classification des objets en mouvement. Le second modèle est qualifié de long terme, c.-à-d. qu'il permet de modéliser les représentations les plus stables de l'arrière-plan en s'adaptant d'une façon très lente. Ce modèle se compose de N échantillons des valeurs de pixel prises dans des fenêtres de temps beaucoup plus importantes. L'échantillon est mise à jour avec un mécanisme de maintenance non sélective.

Basée sur les deux concepts cités précédemment, la détection des objets en mouvement se décide par l'application d'une intersection entre le modèle long et le modèle court. Cette combinaison permet de supprimer les fausses détections. Cependant, l'évènement rare dans l'arrière-plan reste toujours un bruit perturbateur vu qu'aucun modèle ne permet de l'expliquer. Heureusement, ce problème ne va pas persister dans le temps à cause de l'influence du modèle long qui permet une adoption du phénomène après un certain bout de temps.

Malgré les multiples avantages (La multi modalité, l'adaptation aux changements rapides) du KDE, ce modèle présente plusieurs inconvénients majeurs dont le plus important est le besoin des N échantillons pour le processus de décision et de mise à jour qui oblige le système à conservés dans la mémoire une très grande quantité d'information. La complexité algorithmique (égale a $o(N^2)$) et le cout très élevé en temps de calcul est un autre problème qui permet de limiter les utilisations du KDE surtout lorsqu'il s'agit d'une application en temps réel.

Pour résoudre ces problèmes, différentes améliorations ont été proposées. Les premières propositions se focaliser sur la modification de la fonction du noyau elle-même soit par le changement de la bande passante [TAV 05], soit par le remplacement de la gaussienne par une fonction rectangulaire [TAN 07a], par une fonction dérivative [CVE 06] ou par une fonction polynomiale [WIT 07]. Certaines améliorations proposent de diminuer le nombre d'échantillons en déterminant une bonne taille de la mémoire tampon d'image [IAN 05], en utilisant un plan d'échantillonnage diversifié [MAO 04] ou en utilisant un plan d'échantillonnage séquentiel basé sur la loi de Monte-Carlo [TAN 07c]. La

réduction du temps de traitement a pris aussi une part des améliorations proposées par le biais d'une maintenance récursive [TAV 06c] ou une maintenance sélective [TAV 05 ; TAV 06c].

Malheureusement, ces améliorations non pas permis un grand gain de performance. Pour cela, des propositions hybrides ont vu le jour pour pallier les problèmes non encore résolus. Parmi ces approches nous citons : l'utilisation des champs aléatoires de Markov [SHE 05], la théorie des graphes [MAH 06], la différence temporelle [HU 06] et l'ACP [AKM 07].

1.5 Estimation de la densité basée sur un noyau adaptatif

L'estimation de la densité basée sur un noyau adaptatif (AKDE) est un algorithme qui utilise trois étapes pour la modélisation de l'arrière-plan à savoir l'apprentissage, la classification et la mise à jour. Le modèle de l'arrière-plan est généré dans l'étape d'apprentissage. En utilisant quelques frames, les valeurs des pixels sont utilisées pour estimer la probabilité qu'un pixel fasse partie du modèle de l'arrière-plan. Cette phase d'initialisation nécessite que les frames utilisés soient totalement dépourvus des objets en mouvement.

Dans les approches basées noyau, la bande passante du noyau est l'unique paramètre utilisé. Théoriquement, si le nombre de l'ensemble d'apprentissages augmente sans limites alors la densité estimée converge vers la densité réelle indépendamment de la valeur de la bande passante du noyau [DUD 12]. Dans un tel système, on génère pour chaque pixel un modèle non paramétrique et un classifieur déjà entraîné. Les critères de classification sont générés à partir des probabilités estimées pour chaque pixel, devenu un arrière-plan, après le passage de plusieurs nouveaux frames. Dans l'étape de classification, chaque pixel est classé en tant qu'objet en mouvement ou arrière-plan en se basant sur la probabilité calculée selon l'équation suivante :

$$P_t(X_t) = \frac{1}{N2\pi|\Sigma|^{\frac{1}{2}}} \sum_{i=1}^N \exp\left(-\frac{1}{2}(X_t - X_i)^T \Sigma^{-1} (X_t - X_i)\right) \quad (1.9)$$

Où X_t est un vecteur de caractéristiques du pixel dans le temps t et les X_i sont leurs valeurs dans la séquence d'apprentissage. Σ est une matrice symétrique positive non diagonale représentant la bande passante du noyau. N est le nombre de frames utilisés pour l'apprentissage du modèle de l'arrière-plan.

À cause des limitations engendrées par la taille de la mémoire et la puissance de calcul, une fine partie des ensembles d'apprentissages de l'arrière-plan seront sauvegardés. Grâce à cette technique, l'estimation de la densité du noyau non paramétrique sera totalement dépendante du choix de la bande passante du noyau. Le nombre de trames de fond

nécessaires pour l'apprentissage des modèles probabilistes est l'un des principaux inconvénients des approches basées sur l'estimation de densité du noyau pour la modélisation de l'arrière-plan.

1.6 Modélisation récursive

Soit X_t le vecteur de chromaticité d'un pixel à l'instant t . L'estimation non paramétrique du modèle de l'arrière-plan qui suit avec précision la distribution multimodale peut être reformulée en termes de filtrage récursif [TAV 06c] selon l'équation suivante :

$$M_t^B(x) = [1 - \beta_t] \cdot M_{t-1}^B(x) + \alpha_t \cdot H_\Delta(x - x_t) \quad (1.10)$$

avec

$$\sum_x H_\Delta = 1 \quad \text{et} \quad \sum_x x \times H_\Delta(x) = 0 \quad (1.11)$$

Où M_t^B représente le modèle non normalisé de l'arrière-plan à l'instant t . X est une valeur entière comprise entre 0 et 255. $H(\cdot)$ est un noyau local avec une bande passante Δ centrée dans X_t utilisé pour la mise à jour du modèle de l'arrière-plan. Les paramètres α_t et β_t représentent respectivement le taux d'apprentissage et le taux d'oubli planifier.

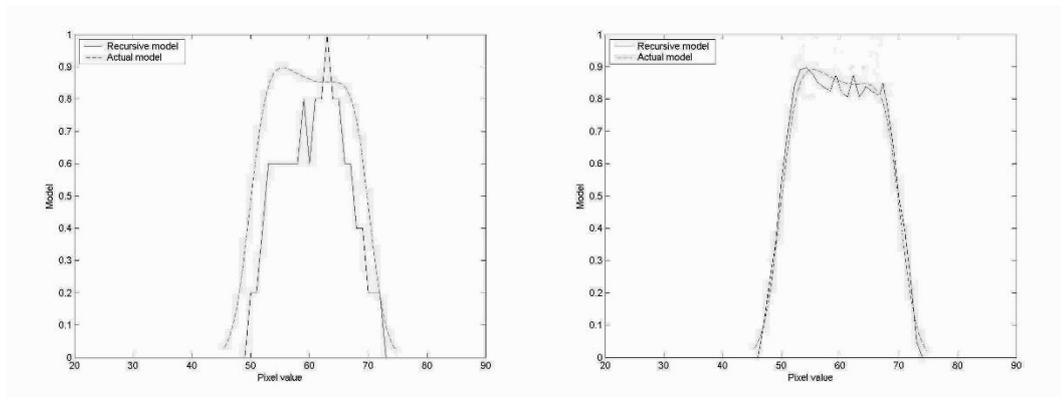


FIGURE 1.1 – Modèle récursive (a) Le modèle après 10 frames (b) le modèle après 467 frames [TAV 14]

La figure 1.1 représente un exemple sur l'utilisation de la modélisation récursive sur une fonction de densité de probabilité générée de manière aléatoire pour générer la population. Le modèle ainsi formé converge avec chaque nouveau frame introduit.

Afin d'accélérer la convergence du modèle, Tavakkoli et al. [TAV 14] ont proposé un algorithme d'apprentissage du modèle de l'arrière-plan basé sur l'historique des valeurs des pixels. Ce processus a permis de rendre la convergence de l'apprentissage adaptatif encore plus rapide. Le taux d'apprentissage change constamment en utilisant le taux

suivant :

$$\alpha_t = \frac{1 - \alpha_0}{h(t)} + \alpha_0 \quad (1.12)$$

avec

$$\alpha_0 = \frac{1}{256 \times \alpha_\theta} \quad \text{et} \quad h(t) = t - t_0 + 1 \quad (1.13)$$

Où α_t est le taux d'apprentissage à l'instant t et α_0 représente un petit taux cible. α_θ est la variance du modèle. $h(t)$ est une fonction monotone croissante. t_0 exprime l'instant dans lequel un changement soudain est appliqué sur l'arrière-plan.

1.7 Différence de frames

Le calcul de la différence entre deux frames est la méthode la plus intuitive pour la détection des objets en mouvement. Les approches basées sur la différence entre images ne sollicitent aucune modélisation de l'arrière-plan pour détecter les régions de mouvement. Donc, une simple analyse de la variation temporelle de l'intensité lumineuse des pixels est suffisante pour confirmer l'existence d'un objet n'appartenant pas à l'arrière-plan.

La manière la plus évidente consiste à prendre deux images successives et d'appliquer une différence absolue. Pour traiter les bruits parasites, l'application d'un seuillage est nécessaire pour déterminer les changements dans la scène observée. Cette méthode dépend fortement de la vitesse de mouvement des objets, une très grande vitesse nécessite un seuil extrêmement élevé.

Dans la pratique, on essaye plutôt de prendre les images sur un intervalle de temps qui peut s'étaler de 3 jusqu'à 7 frames selon l'équation 1.14.

$$\Delta_t = |I_{t+d} - I_t| \quad (1.14)$$

I_{t+d} étant l'image en cours de traitement et I_t l'image de référence. d est la distance en frame entre les deux images précédentes. Pour détecter les objets en mouvement, un masque binaire est appliqué selon un seuil prédéfini selon l'équation suivante.

$$I_{foreground}(i, j) = \begin{cases} 1 & \text{si } \Delta_t > T \\ 0 & \text{sinon} \end{cases} \quad (1.15)$$

En plus de la simplicité d'implémentation et le temps réduit dans l'exécution, cette méthode présente l'avantage d'avoir une très grande capacité de s'adapter avec les changements de luminosité. Cependant, elle est moins robuste face aux bruits. En plus, les résultats de la détection dépendent fortement du choix d'une bonne méthode de seuillage.

Dans la littérature, on trouve plusieurs propositions pour améliorer cette méthode. Kameda et al. [KAM 96] proposent une double différence à partir de trois images consé-

cutives. Les résultats définitifs est obtenus par un ET logique entre les deux masques résultants. Spagnolo et al. [SPA 06] ont proposé une amélioration significative en ajoutant un modèle de l'arrière-plan. Cette méthode hybride utilise le même principe que Kameda [KAM 96] à laquelle les auteurs ont ajouté un modèle de fond pour raffiner les résultats obtenus. Migliore et al. [MIG 06] ont essayé de résoudre le problème d'ouverture par l'application d'une double différence. La première différence est appliquée entre deux frames successifs et la deuxième différence entre le résultat de la première et une image de l'arrière-plan.

1.8 Soustraction par image de référence

Par définition, l'arrière-plan est une image I_F dans laquelle la scène est totalement dépourvue d'objet n'appartenant pas de nature à l'image. Pour pouvoir extraire les objets qui ne font partie du décor, on applique une segmentation qui consiste à étudier chaque pixel de l'image et à les comparer avec l'image de référence I_F . La décision est prise par rapport un seuil T fixé a priori de manière empirique. Si la différence est supérieure à un seuil T , le pixel est classé en tant qu'objet en mouvement, sinon il est classé en tant qu'arrière-plan. Le résultat de cette opération est une image binaire I_B exprimée par l'équation suivante :

$$I_{B,t}(x,y) = \begin{cases} 1 & \text{si } |I_t(x,y) - I_F| \geq T \\ 0 & \text{sinon} \end{cases} \quad (1.16)$$

La difficulté dans un tel système réside dans l'obtention de l'image initiale utilisée comme référence. En plus, le modèle doit contenir un mécanisme de mise à jour robuste pour pouvoir surmonter les problèmes liés aux changements de luminosité. L'estimation de l'image de référence I_F peut être effectuée de différentes façons. La plus intuitive consiste à représenter cette dernière par une moyenne estimée de façon incrémentale à l'aide de l'équation suivante.

$$I_{F,t}(x,y) = \alpha I_{F,t-1} + (1 - \alpha) I_t(x,y) \quad (1.17)$$

α est le taux d'apprentissage qui permet de contrôler la vitesse de mise à jour de l'image de l'arrière-plan. Si α est suffisamment élevé, le modèle est très stable et ne prend que très peu en compte la nouvelle image, tandis qu'une valeur très faible permettra une adaptation très rapide de l'arrière-plan.

Pour pallier un peu le choix du paramètre α , une alternative consiste à utiliser les n derniers échantillons pour estimer la valeur médiane des intensités des pixels (équation 1.18). De cette façon, on prend comme image de référence, une image capable en même

temps de représenter l'ancien et le nouvel état du système.

$$I_{F,t}(x,y) = \text{Median}\{I_{t-n+1}(x,y), I_{t-1}(x,y), I_t(x,y)\} \quad (1.18)$$

L'inconvénient majeur de cette approche est la nécessité de conserver en mémoire les images précédentes pour l'estimation. Une implémentation récursive a été proposée dans [MCF 95] où la valeur médiane est incrémentée de 1 si le pixel est supérieur à sa valeur, et décrétementée de 1 dans le cas échéant. L'inconvénient est une lente adaptation et demande par conséquent une longue période d'apprentissage. De manière générale, l'incrémenta-tion (ou la décrémentation) de la valeur médiane s'effectue à l'aide d'une constante notée c selon l'équation 1.19 :

$$I_{F,t}(x,y) \begin{cases} I_{F,t}(x,y) + c & \text{si } I_t(x,y) > I_{F,t-1}(x,y) \\ I_{F,t}(x,y) - c & \text{si } I_t(x,y) < I_{F,t-1}(x,y) \\ I_{F,t}(x,y) & \text{sinon} \end{cases} \quad (1.19)$$

1.9 Apprentissage sub-spatial basé ACP (SL-PCA)

L'apprentissage subspatial permet d'offrir une bonne plateforme pour résoudre les problèmes liés aux changements de luminosité, car il permet de prendre en considération les informations spatiales.

L'idée principale consiste à prendre un échantillon de N images et d'appliquer une ACP afin de dégager un vecteur de valeurs propres. Oliver et al. [OLI 00] proposent de calculer une moyenne I_μ et une matrice de covariance σ à partir des enchan-tions d'arrière-plan prisent dans des intervalles de temps non consécutifs. Ensuite, une diagonalisation est appliquée sur la matrice de covariance pour obtenir une base de vecteurs propres v et une matrice diagonale M comportant les valeurs propres associées.

$$M = v \cdot \sigma \cdot v^T \quad (1.20)$$

Afin de réduire la dimension de l'espace, seuls les vecteurs propres associés aux K plus grandes valeurs propres sont conservés ($K < N$) sont conservés dans l'ACP. Une fois l'apprentissage terminé, une approximation de l'image en court de traitement est effectuée en utilisant l'image moyenne I_μ et la somme des pondérations des vecteurs propres. Les ob-jets en mouvement sont extraits en calculant la distance entre l'image d'entrée et l'image reconstruite à partir de sa projection dans l'espace de dimension réduite selon l'équation 1.21 et 1.22.

$$|I_t - I_P| > T \quad (1.21)$$

avec

$$I_p = v (I_t - I_\mu) v^T + I_\mu \quad (1.22)$$

Où I_t est l'image en court de traitement, I_p est l'image projetée et T est un seuil fixe défini empiriquement. Dans ce modèle, la fonction de distribution de probabilité de l'arrière-plan est très robuste. Cependant, ce n'est pas le cas pour les objets en mouvement, car ces derniers n'ont aucune contribution dans la création du modèle.

La première limitation de ce modèle est que la taille de l'objet en mouvement doit être petite. En plus, l'objet ne doit pas apparaître figé dans le même emplacement pendant de longue période dans la phase d'apprentissage. Le deuxième problème concerne la mise à jour du modèle de l'arrière-plan, car dans la pratique, l'utilisation de l'ACP nécessite un calcul intensif. De plus, sans la présence d'un mécanisme d'analyse robuste, le risque d'absorption des objets en mouvement dans le modèle de l'arrière-plan devient très fréquent. Le dernier problème est lié à l'utilisation du modèle pour des images au niveau de gris à cause de la complexité de mise en œuvre dans les espaces de couleur multidimensionnels tels que RGB ou HSV. En plus, le modèle ne permet pas une représentation multimodale pour confronter les bruits causés par les changements de luminosité.

Pour remédier à toutes ces limitations, plusieurs approches ont été proposées dans cette dernière décennie. La solution pour la première limitation est d'essayer de corriger le problème de la taille des objets en mouvement. Pour cela, Xu et al. [XU 06 ; XU 08] proposent d'appliquer un algorithme qui permet de réduire l'influence des objets en mouvement sur le modèle par une compensation de l'erreur commise. Kawabata et al. [KAW 06] proposent d'estimer un arrière-plan varié contenant des objets en mouvement par le biais d'une projection optimale itérative en temps réel d'une scène dynamique. Une autre alternative proposée par Quivy et al. [QUI 11] dans laquelle le processus de projection et de reconstruction de la méthode SL-PCA a été entièrement remplacé par un générateur d'arrière-plan.

Pour résoudre le problème de temps d'exécution et de la mise à jour du système, Li et al. [LI 04b] proposent de réduire le temps d'exécution par une analyse en composantes principales incrémentale. Une version améliorée a été proposée par Skocaj et al. [SKO 08] dans laquelle un poids a été attribué pour chaque frame afin d'augmenter la robustesse du modèle.

Malheureusement, le modèle néglige la contribution des parties de l'image dans la construction du modèle de l'arrière-plan. Les travaux de Zhang et al. [ZHA 07a] ont permis de corriger ce problème par une pondération adaptative en attribuant un poids à chaque pixel dans l'étape de mise à jour du modèle. Toutes ces propositions ont permis de réduire le temps de mise à jour grâce à l'utilisation d'une décomposition en valeurs propres de la matrice de covariance.

Pour résoudre le problème de l'utilisation des niveaux de gris et du changement de lu-

mière, Wu et al. [WU 10] ont proposé une hybridation entre l'ACP et le modèle gaussien. Ce dernier offre une flexibilité par rapport au changement de luminosité et une meilleure description pour les pixels. Dans les travaux de Han et al. [HAN 07], les auteurs ont proposé un ACP incrémental 2D sur les trois canaux de l'espace de couleur RGB. Le système offre une très bonne gestion des variations temporelle de l'arrière-plan. Zhao et al. [ZHA 08] ont proposé une approche basée sur des blocs spatio-temporels pour remédier aux limitations des approches basées pixels. Le système offre une convergence rapide et une complexité moins importante. Cette performance est due à l'utilisation d'une autre variante du modèle ACP incrémentale basée sur une covariance candide libre. D'autres travaux ont proposé de résoudre le problème de luminosité par un apprentissage multi spatiale [DON 11], par une analyse en composante principale locale [DON 11] ou par une approche spatio-temporelle par boc.

1.10 Filtrage prédictif

La prédiction est une astuce largement utilisée dans le traitement de l'image et de la vidéo. Le principe de cette méthode est basé sur l'historique des anciennes valeurs prises par un pixel. Dans ce cas, un nouveau pixel est considéré comme faisant partie de l'arrière-plan si la valeur de ce dernier est proche de la prédiction. La prédiction de la valeur du pixel est donnée par l'équation 1.23.

$$X_t = - \sum_{k=1}^p \alpha_k \cdot X_{t-k} \quad (1.23)$$

Où

α_k : le coefficient de prédiction déterminé à partir de la covariance des valeurs de X_t .

X_t : la prédiction de la valeur du pixel à l'instant t .

p : le nombre d'échantillons les plus récents.

Toyama et al. [TOY 99] ont proposé une prédiction basée sur le filtre de Wiener. Le filtre de Wiener permet de construire une valeur estimée X_t de la valeur que l'on devrait observer à l'instant t . Ils ont également proposé une fonction pour calculer la distance entre la valeur du pixel est la prédiction X_t selon l'équation suivante.

$$E [e_t^2] = E [s_t^2] + \sum_{k=1}^p \alpha_k \cdot E [s_t s_{t-k}] \quad (1.24)$$

Alors, le pixel est considéré comme étant un objet en mouvement s'il satisfait l'équation suivante :

$$\frac{|X - X_t|}{4} < \sqrt{E [e_t^2]} \quad (1.25)$$

Cette méthode permet de gérer les problèmes d'arrière-plans non statiques, et les perturbations intervenant à intervalles réguliers.

1.11 Filtre de Kalman

Lorsque le système dynamique relatif au problème est représentable sous forme linéaire et qu'il est perturbé par un bruit supposé gaussien, le filtrage de Kalman est un moyen efficace qui utilise également une autre approche prédictive fournissant ainsi une solution optimale.

On trouve dans la littérature plusieurs variantes du filtre de Kalman pour la modélisation de l'arrière-plan. De nombreuses versions ont été proposées [RID 95 ; GAO 01 ; ZHO 03 ; LEI 10 ; NOO 11] et se différencient généralement par le choix des caractéristiques utilisées.

Notons X_t le vecteur d'état d'un pixel p à l'instant t . Ce vecteur décrit la valeur d'intensité lumineuse du pixel ainsi que sa dérivée temporelle et s'écrit $x_t = [I_t, \dot{I}_t]^T$. Chaque pixel de l'image est mis à jour récursivement selon l'équation 1.26

$$\begin{bmatrix} I_t \\ \dot{I}_t \end{bmatrix} = \begin{bmatrix} I_{t-1} \\ \dot{I}_{t-1} \end{bmatrix} + k \cdot \left(I_t - H \cdot A \cdot \begin{bmatrix} I_{t-1} \\ \dot{I}_{t-1} \end{bmatrix} \right) \quad (1.26)$$

avec A la matrice d'évolution décrivant la dynamique de l'arrière-plan et H la matrice d'observation décrivant la relation entre la mesure et l'état. Le gain K du filtre caractérise généralement le taux d'apprentissage et s'écrit $K = [\alpha \quad \alpha]^T$

1.12 Dictionnaire de mots visuels

Dans cette approche, on essaye de construire pour chaque pixel de l'image un modèle représenté par un ensemble de variables appelées mots visuels caractérisant son état actuel.

L'utilisation d'un dictionnaire de mots visuels a été initialement proposée dans [HAR 00]. Dans le modèle W4, proposé par Haritaoglu et al. [HAR 00], chaque pixel est caractérisé par un jeu de 3 valeurs, le minimum d'intensité (*Min*), le maximum d'intensité (*Max*) et la différence maximum d'intensité entre deux images consécutives (*Diff*). Ces valeurs sont initialement estimées durant une phase d'apprentissage et mises à jour régulièrement au cours du temps. Un pixel est considéré comme étant en mouvement s'il satisfait l'une des deux conditions suivantes :

$$|Min - I| > Diff, |Max - I| > Diff \quad (1.27)$$

On trouve également dans la littérature des extensions et des propositions similaires [KIM 04 ; ZHA 09 ; SHA 11] pour la modélisation de l'arrière-plan. Dans Kim et al. [KIM 04], le dictionnaire de mots visuel va prendre une allure de cylindre. Ce dernier est enrichi par la fréquence d'occurrence du mot, la durée maximale durant laquelle le mot n'a pas été sollicité pendant l'apprentissage, et le premier et le dernier accès au mot visuel. Ces valeurs sont utilisées pendant la période d'apprentissage pour construire le modèle. La classification s'effectue en calculant la différence de couleur et de luminosité. Si la valeur du pixel est contenue dans le cylindre alors il est considéré comme appartenant à l'arrière-plan.

1.13 Séparateurs à vaste marge

Les séparateurs à vaste marge sont une classe très spécifique d'algorithmes caractérisés par : l'utilisation de noyaux, l'absence de minima locaux, la faible densité de la solution et la capacité de contrôle obtenue en agissant sur la marge, ou sur d'autres quantités de dimension indépendante telle que le nombre de vecteurs de support.

Les séparateurs à vaste marge ont été inventés par Vapnik et al. en 1992 [VAP 98]. Cependant, toutes ces caractéristiques étaient déjà présentes et avaient été utilisées dans l'apprentissage de la machine depuis les années 1960. Pour la classification, les SVM travaillent en déterminant un hyperplan dans un espace de fonction dimensionnelle élevée pour séparer les données d'apprentissage en deux classes. Le meilleur hyperplan peut être dérivé en minimisant la marge qui représente la plus petite distance entre l'hyperplan et les données. L'utilisation de cette technique pour la modélisation de fond a été proposée par Lin et al. [LIN 02] qui proposent une version probabiliste des SVM vu que ces derniers génèrent uniquement des sorties binaires. La transformation des sorties est assurée par une fonction sigmoïdale selon l'équation suivante :

$$p\left(y = \frac{1}{f}\right) = \frac{1}{1 + \exp^{Af+B}} \quad (1.28)$$

Où y est une étiquette de classe binaire et f est un score de sortie de la fonction de décision des SVM. Les deux paramètres A et B sont adaptés en utilisant une estimation du maximum de vraisemblance de l'ensemble (f, y) , et dérivé en minimisant la fonction log-vraisemblance négative.

Pour éviter le sur apprentissage et d'en tirer la formation impartiale pour la minimisation, un ensemble de retenue est générée à partir des données en divisant respectivement chaque ensemble d'apprentissages en 80% utilisés pour la formation des SVM et 20% utilisés pour la minimisation des deux paramètres.

Dans le contexte de la modélisation de l'arrière-plan, Lin et al. [LIN 02] ont utilisé 100 images de taille 160×120 avec des arrière-plans connus. Chaque image est divisée en

blocs de taille 4×4 et deux caractéristiques ont été utilisé : le flux optique et la différence temporelle. Un bloc de l'image est classifié comme arrière-plan si la probabilité de sortie est supérieure à un certain seuil T fixé de manière empirique.

1.14 Les séparateurs à vastes marges pour la régression (SVR)

L'utilisation des Séparateurs à vastes marges étendue au cas de la régression constitue une alternative prometteuse pour la modélisation des systèmes complexes. Étant donné un ensemble de données d'apprentissage, le SVR adopte une fonction en spécifiant une limite supérieure sur une fraction des données d'apprentissage autorisées à se situer en dehors d'une distance ε à partir de l'estimation de la régression. Pour le cas de la régression linéaire, la variable à prédire y_i est un nombre réel. L'utilisation des SVM pour la régression consiste à utiliser l'hyperplan séparateur optimal pour prédire y .

Pour chaque pixel appartenant à la l'arrière-plan, un SVR distinct est utilisé pour le modéliser en tant qu'une fonction d'intensité. Pour classer un pixel donné comme arrière-plan ou non. Wang et al. [WAN 09] proposent de fournir la valeur de l'intensité de chaque pixel au SVR associé, ensuite d'appliquer un seuil pour la sortie du SVR.

Soit l'ensemble D de N paires de données d'un pixel p de l'image, ayant un vecteur de données X comme entrée et l'étiquette Y de ce vecteur comme sortie.

$$D = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}, i \in \{1, \dots, N\}, x_i \in \mathfrak{R}^p \text{ et } y_i \in \mathfrak{R} \quad (1.29)$$

Où x_i correspond à la valeur d'intensité du pixel p dans le frame i , et y_i correspond au degré de confiance du pixel p étant un pixel d'arrière-plan. Une fois l'apprentissage du SVR est achevé, le degré de confiance du pixel p dans un nouveau frame i , $f(x_i)$, est calculé selon la fonction de régression linéaire suivante :

$$f(x_i) = \sum_{i=1}^N (\alpha_i - \hat{\alpha}_j) k(x_i, x_j) + \varepsilon \quad (1.30)$$

Où $k(x_i, x_j)$ est une fonction noyau. Les paramètres α , $\hat{\alpha}$ et ε appelés multiplicateurs de Lagrange, sont obtenus par la résolution d'un problème d'optimisation en utilisant le procédé des multiplicateurs de Lagrange. Compte tenu du modèle de l'arrière-plan à base de SVR, l'intensité de chaque pixel dans un nouveau frame est utilisée pour la construction du vecteur d'entrée du SVR. La sortie de la SVR représente le degré de confiance qu'un pixel donné appartient à l'arrière-plan.

Donc, un pixel est étiqueté comme arrière-plan si son degré de confiance est compris

entre deux seuils S_L et S_H selon l'équation 1.31

$$I_f = \begin{cases} 0 & \text{si } S_L < f(x_i) < S_H \\ 1 & \text{sinon} \end{cases} \quad (1.31)$$

Le résultat de cette opération est une image binaire I_f des objets en mouvement. La mise à jour du modèle est effectuée selon l'algorithme d'apprentissage en ligne du SVR.

1.15 Les séparateurs à vastes marges pour la description de données

Les Séparateurs à vastes marges pour la description de données (SVDD) est une méthode qui permet de retrouver la frontière autour d'un ensemble de données. Les SVDD ont été largement utilisées dans le domaine la reconnaissance de l'écriture manuscrite, la détection et la reconnaissance de visage et la détection des anomalies. Tavakkoli et al. [TAV 06c] ont proposé de modéliser l'arrière-plan en utilisant les SVDD sur des vidéos contenant des arrière-plans presque stationnaires. La frontière d'un ensemble de données peut être utilisée pour détecter de nouvelles données n'appartenant pas à l'ensemble de départ.

Une description de données normale donne une frontière fermée autour de la donnée. L'utilisation d'un hyper sphère est la manière la plus simple pour représenter cette frontière. La sphère est caractérisée par un centre α et un rayon $R > 0$. La minimisation du volume de la sphère est assurée par la réduction de la valeur de R^2 tout en gardant la totalité des données d'apprentissage x_i à l'intérieur du sphère. De façon analogique au SVM, la fonction à minimiser est donnée par l'équation suivante.

$$F(R, \alpha) = R^2 \quad (1.32)$$

avec

$$\|x_i - \alpha\|^2 \leq R^2, \forall i \quad (1.33)$$

Pour permettre la possibilité de trouver les valeurs atypiques dans l'ensemble d'apprentissage, la distance entre la valeur x_i et le centre α ne doit pas être strictement inférieure à R^2 , tandis que les distances les plus larges sont pénalisées. L'introduction de la variable négligeable $\varepsilon_i \geq 0$ permet de changer le problème de minimisation selon l'équation 1.34 suivante :

$$F(R, \alpha) = R^2 + C \sum_i \varepsilon_i \quad (1.34)$$

avec $\|x_i - \alpha\|^2 \leq R^2 + \varepsilon_i, \forall i$

Le paramètre de confiance C permet de jouer le rôle d'un arbitre entre la simplicité du système et l'erreur commise par ce dernier. L'équation précédente peut être également écrite de la manière suivante :

$$L(R, \alpha, \alpha_i, \gamma_i, \varepsilon_i) = R^2 + c \sum_i \varepsilon_i - \sum \gamma_i \varepsilon_i - \sum \alpha_i \{R^2 + \varepsilon_i - (\|x_i\|^2 - 2\alpha \cdot x_i + \|\alpha\|^2)\} \quad (1.35)$$

Avec des multiplicateurs de Lagrange $\alpha_i \geq 0$ et $\gamma_i \geq 0$. L doit être minimisé en respectant les paramètres R, a, ε_i et maximiser suivant les paramètres α_i, γ_i . Avec la supposition que $0 \leq \alpha_i \leq C$, le multiplicateur de Lagrange γ_i peut être facilement supprimé. La résolution du problème d'optimisation nous ramène à l'équation suivante :

$$L = \sum_i \alpha_i (x_i \cdot x_i) - \sum_{i,j} \alpha_i \alpha_j (x_i \cdot x_j) \quad (1.36)$$

avec $0 \leq a_i \leq C, \forall \alpha_i$

Lorsqu'une nouvelle donnée satisfait l'équation 1.34, alors son multiplicateur de Lagrange est exprimé par $\alpha_i \geq 0$ sinon il est mis à zéro.

$$\begin{cases} \|x_i - \alpha\|^2 < R^2 & \text{si } \alpha_i = 0 \text{ et } \gamma_i = 0 \\ \|x_i - \alpha\|^2 = R^2 & \text{si } 0 < \alpha_i < C \text{ et } \gamma_i = 0 \\ \|x_i - \alpha\|^2 > R^2 & \text{si } \alpha_i = C \text{ et } \gamma_i = 0 \end{cases} \quad (1.37)$$

On remarque clairement que le centre de la sphère est une combinaison linéaire des objets x_i , donc nous avons besoin uniquement des données avec des $\alpha_i > 0$ pour la description des données. Et c'est pour cette raison qu'on les appelle des SVM de description de données.

Pour tester un objet z , la distance vers le centre de la sphère est calculée. L'objet z sera accepté si la distance est inférieure ou égale au rayon de la sphère

$$\|z - \alpha\|^2 = (z \cdot z) - 2 \sum_i \alpha_i (z \cdot x_i) + \sum_{i,j} \alpha_i \alpha_j (x_i \cdot x_j) \leq R^2 \quad (1.38)$$

Tavakkoli et al. [TAV 06c] ont utilisé cette méthode pour construire une descriptive des frontières de chaque pixel dans l'apprentissage pour la construction du modèle de l'arrière-plan. Ce modèle sera utilisé plus tard pour la classification des pixels des nouveaux frames en tant qu'objet en mouvement ou arrière-plan.

Cette modélisation présente beaucoup d'avantages, la précision du modèle n'est pas limitée par la précision des fonctions de densité de probabilité estimées. En plus, le modèle consomme moins de mémoire par rapport aux autres méthodes non paramétriques. Cela est relié directement à la manière dont les SVDD traitent les frontières de classes

inconnues. Cependant, elle nécessite une optimisation des paramètres de Lagrange qui sont très coûteux en temps de calcul.

1.16 Simple gaussienne généralisée

Pour faire face aux divers problèmes liés aux changements dans l'arrière-plan et dans les ombres, Kim et al. [KIM 08] ont proposé de modéliser l'arrière-plan en utilisant un modèle gaussien généralisé (SGG). L'idée principale est de prendre en considération que la variance des pixels suit parfois la loi de Laplace au lieu d'une gaussienne. Dans les scènes avec de longues périodes statiques de l'arrière-plan, la variance des pixels est plus proche d'une modélisation Laplace qu'une modélisation gaussienne. Cependant, le modèle de Laplace n'est pas adéquat pour des environnements divers. La variation d'un pixel est définie selon l'équation 1.39 suivante :

$$P(X_t) = \frac{\rho\gamma}{2\Gamma\left(\frac{1}{\rho}\right)} e^{-(\gamma^\rho|x-\mu|^\rho)} \quad (1.39)$$

avec

$$\gamma = \frac{1}{\sigma} \left(\frac{\Gamma\left(\frac{3}{\rho}\right)}{\Gamma\left(\frac{1}{\rho}\right)} \right) \quad (1.40)$$

Où $\Gamma(\cdot)$ est une fonction gamma et σ^2 est la variance de la distribution. Dans l'équation 1.39, $\rho = 1$ représente une distribution Laplace tandis que $\rho = 2$ représente une distribution gaussienne. Les modèles sont décidés pour chaque pixel en calculant le coefficient d'aplatissement $g2$ des premiers m frames. Le coefficient d'aplatissement des distributions Laplace et gaussiennes sont respectivement 3 et 0. Les paramètres optimaux du modèle d'arrière-plan sont estimés par la maximisation de la vraisemblance de la valeur observée :

$$g2 = \frac{N \sum_{i=1}^N (x_i - \mu)^4}{\left(\sum_{i=1}^N (x_i - \mu)^2 \right)^2} - 3 \quad (1.41)$$

1.17 Mixture de gaussienne généralisée

Pour une meilleure segmentation dans les environnements contenant des bruits, Allili et al. [ALL 07] ont proposé un modèle basé sur les mixtures de gaussiennes généralisées (GGMM). Ce modèle dispose d'une plus grande flexibilité pour s'adapter avec la forme des données. En plus, il est moins sensible qu'une GMM pour la détection de nombre de classes et plus robuste dans la présence d'ombre dans la scène. Chaque pixel est caractérisé par son intensité dans l'espace couleur RGB. Ensuite, la probabilité d'observer la valeur du pixel courant est considérée dans le domaine multidimensionnel selon les

équations 1.42 et 1.43 :

$$P(X_t) = \sum_{i=1}^k \omega_{i,t} \cdot \eta(X_t, \mu_{i,t}, \sigma_{i,t}, \lambda_i) \quad (1.42)$$

avec

$$\eta(X_t, \mu_{i,t}, \sigma_{i,t}, \lambda_i) = \prod_{j=1}^d A(\lambda_i) \exp\left(-B(\lambda_i) \left|\frac{X_j - \mu_j}{\sigma_j}\right|^{\lambda_j}\right) \quad (1.43)$$

où

$$A(\lambda) = \frac{\left(\frac{\Gamma(\frac{3}{\lambda})}{\Gamma(\frac{1}{\lambda})}\right)^{\frac{1}{\lambda}}}{2\sigma\Gamma(\frac{1}{\lambda})} \text{ et } B(\lambda) = \left(\frac{\Gamma(\frac{3}{\lambda})}{\Gamma(\frac{1}{\lambda})}\right)$$

Où K est le nombre de distributions attribuées à chaque pixel, $\omega_{i,t}$ est le poids associé à la $i^{\text{ème}}$ gaussienne dans le temps t , $\mu_{i,t}$ est la moyenne, $\Sigma_{i,t}$ est l'écart type et η est la fonction de densité de probabilité. De façon identique à une simple gaussienne généralisée, le paramètre λ permet de désigner le modèle utilisé. Avec $\lambda_i = 0$ la distribution est de type gaussienne, si $\lambda_i = 3$ alors la distribution est de type Laplace. La mise à jour des paramètres de la GGMM se fait de la même manière que dans un modèle GMM. Le nombre de gaussiennes est calculé pour chaque frame par la minimisation du critère LMM (longueur minimale des messages).

1.18 Conclusion

Nous avons présenté dans ce chapitre un aperçu détaillé sur les techniques de base utilisées pour la modélisation de l'arrière-plan. Ces méthodes offrent un support pour la détection et l'extraction des objets en mouvement dans les systèmes de vidéosurveillance grâce à un appartement entre les modèles de l'arrière-plan et le frame en cout de traitement pour détecter les régions susceptible de contenir des objets d'intérêt. Les enjeux relatifs à la nature elle-même de l'environnement à inciter les chercheurs à développer davantage de méthodes pour faire face à la complexité des environnements sous surveillance, telles que les changements de luminosité et les différentes conditions climatiques. Récemment, les techniques de détection et d'extraction des objets en mouvement se sont orientées vers des modèles plus complexes qui s'adaptent et prennent en compte la multimodalité de l'arrière-plan. Les modèles statistiques tels que les modèles de mixture de lois sont parmi les méthodes qui ont eu un très grand intérêt dans la littérature grâce à leurs capacités de s'adapter avec les changements de l'environnement.

2

Détection et extraction d'ombre

2.1 Introduction

Au cours des dernières années, les systèmes de vision sont largement utilisés dans divers domaines, le nombre de ces systèmes est en augmentation, et leurs évolutions dans le domaine de l'intelligence artificielle ont permis la détection et le suivi automatique des objets en mouvement.

Dans les systèmes de vision par ordinateur il est souvent nécessaire de différencier entre l'objet et son ombre. Par conséquent, la détection de l'ombre est utile dans de nombreuses applications, y compris : l'interprétation de scènes, la segmentation d'image, la reconnaissance et le suivi d'objet.

Pour détecter un objet en mouvement, un système de vision estime le fond, puis obtient une image différentielle en soustrayant le fond estimé à partir de l'image courante. Cependant, l'ombre portée par l'objet mobile est également détectée. L'ombre pose des enjeux majeurs et rend la tâche de détection et de reconnaissance de la forme exacte de l'objet difficile et parfois impossible. Par conséquent, la précision de la détection d'un objet en mouvement et l'acquisition de sa forme exacte par l'élimination des ombres ont un grand effet sur la performance des étapes ultérieures telles que le suivi, la reconnaissance, la classification et l'analyse de l'activité.

L'ombre a une dynamique similaire de l'objet générateur. En plus, les méthodes d'extraction de l'arrière-plan ont du mal à classer les pixels d'ombre comme arrière-plan puisqu'en réalité ces derniers ne font pas partie de l'arrière-plan. C'est pour ces raisons que

l'identification de l'ombre est essentielle dans la reconnaissance d'objet et cela à la fois pour la vidéo et les images fixes. Donc, il est nécessaire, dans les systèmes de vision automatisée, de prévoir des mécanismes pour le suivi d'objets dans le champ de vision. Dans le suivi d'objets en mouvement, les ombres peuvent être classées comme des objets en raison de leurs caractéristiques visuelles. De ce fait, une mauvaise classification des ombres peut entraîner une fusion ou une déformation des objets causant d'importantes confusions pour les étapes ultérieures.

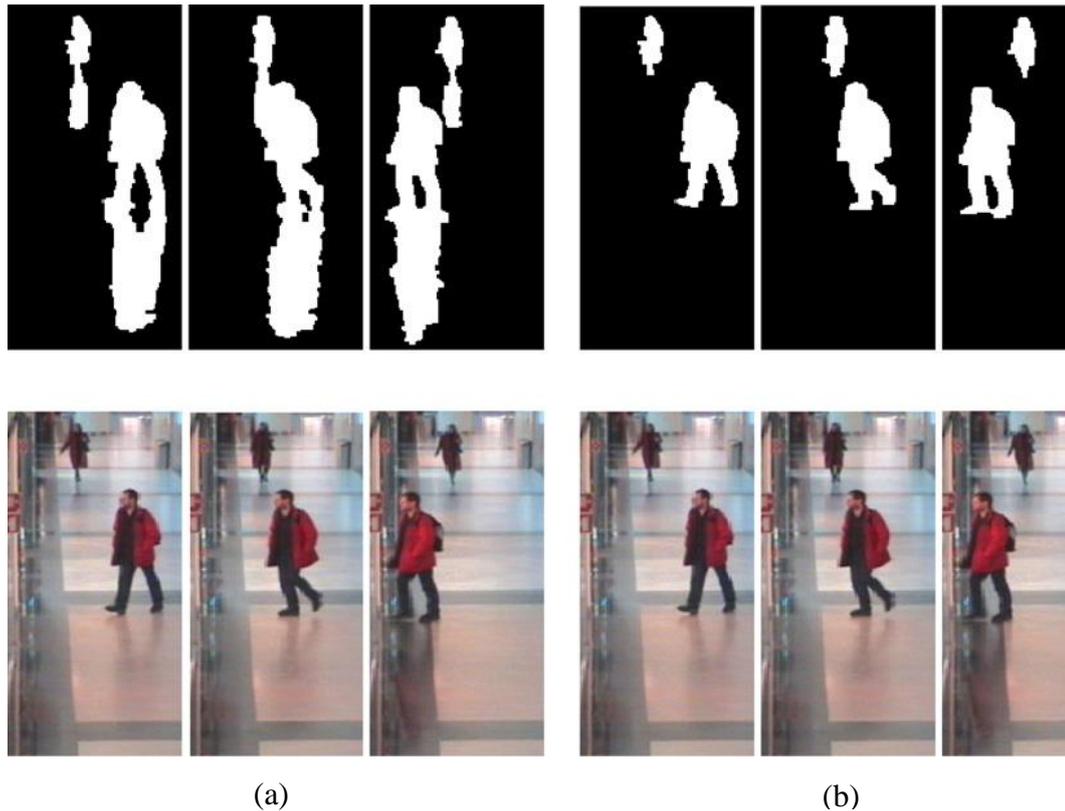


FIGURE 2.1 – Résultats du suivi des objets en mouvement avec et sans la suppression de l'ombre. (a) extraction sans suppression d'ombre. (b) extraction avec suppression d'ombre.

Les images de la figure 2.1 montrent un cas où la trajectoire correcte du suivi ne peut être obtenue lorsque les ombres ne sont pas supprimées. Le masque binaire montre clairement que les deux objets ont carrément fusionné pour ne créer qu'un seul objet. Donc, sans l'élimination de l'ombre les limites ne sont pas fiables, entraînant un problème ou une perte dans le suivi des objets.

Parce que l'ombre est principalement reliée à l'objet et la vitesse de l'ombre est la même que celle de l'objet, il est impossible à discriminer ceux-ci à l'aide de flux optique. En outre, il est très difficile à discriminer l'ombre à l'aide des informations sur la couleur ou luminosité parce que la couleur ou la texture de l'ombre peuvent changer en fonction de l'arrière-plan. Cependant, la chromaticité de l'ombre est fixe par rapport à celle du

fond, et la luminosité de l'ombre est généralement inférieure à celle de l'arrière-plan.

Lorsque les objets d'intérêt ont une forme bien définie, plusieurs classificateurs sophistiqués peuvent être utilisés pour segmenter les objets directement à partir de l'image. Ces techniques fonctionnent bien pour des objets avec des contours bien définis, mais sont difficiles à mettre en œuvre pour les objets à contour flexible.

2.2 L'ombre

Lorsque la lumière rencontre un objet opaque elle est diffusée, absorbée ou réfléchi, mais à l'arrière de cet objet la lumière ne passe pas, c'est une zone d'ombre. Donc, on peut dire qu'une ombre est une zone sombre créée par l'interposition d'un objet opaque entre une source de lumière et la surface sur laquelle se réfléchit cette lumière. Cette ombre prend une forme d'une silhouette sans épaisseur qui dépend de l'intensité de la source et de son emplacement par rapport à l'objet.

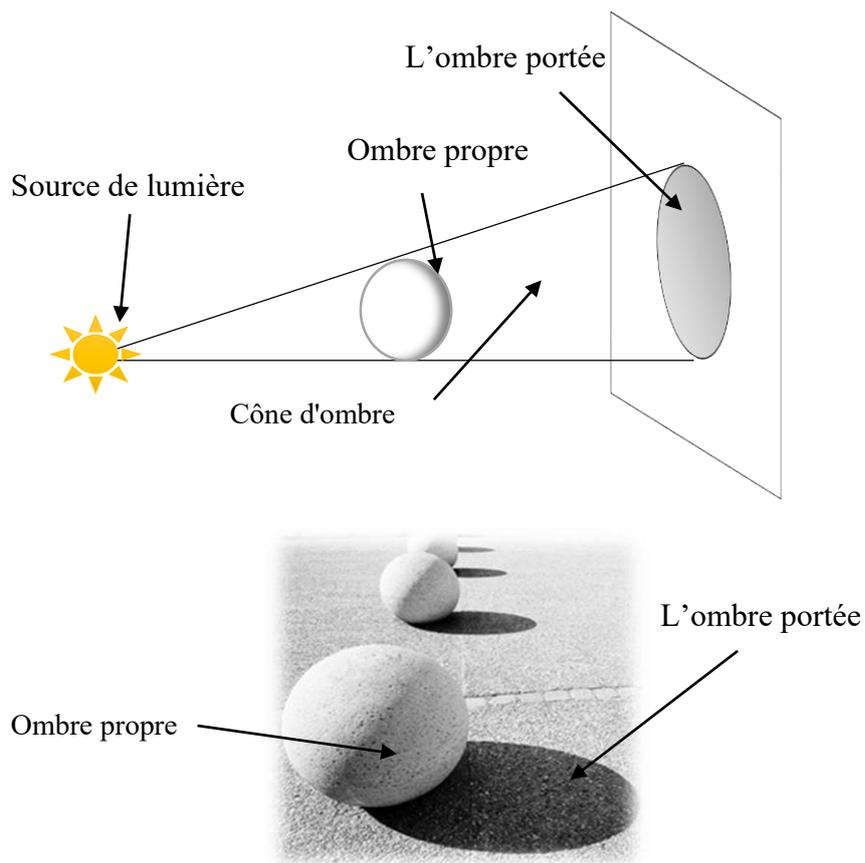


FIGURE 2.2 – Représentation des types d'ombres

On distingue quatre types de zones d'ombres :

- **L'ombre propre (Attached shadow)** est une région de l'objet qui ne reçoit pas de lumière. Elle se situe derrière l'objet, dans la zone où la lumière issue de la source

n'arrive pas.

- **Le cône d'ombre (Umbra)** est une région de l'espace où les rayons de lumière issus de la source ne passent pas puisqu'ils sont arrêtés par l'objet.
- **L'ombre portée (Cast Shadow)** est une région d'un écran, placé derrière un objet par rapport à la source de lumière et qui ne reçoit pas de rayon. La taille et la forme de l'ombre portée dépendent de la forme, de la taille et de la position de l'objet par rapport à la source, mais dépendent aussi de l'emplacement et de l'inclinaison de l'écran.
- **La pénombre (Penumbra)** est une zone frontière qui apparaît entre la partie éclairée et la partie dans l'ombre.

Dans la plupart des systèmes de vision tels que la vidéosurveillance, la reconnaissance des formes et le suivi des objets en mouvement, la détection des ombres propre et moins importante et parfois négligée vues que l'objectif principal de ces systèmes est de segmenter l'image en deux classes : objet en mouvement et arrière-plan. Dans de telles situations, il est préférable de considérer l'ombre propre comme faisant partie de l'objet.

2.3 Propriété spectrale de l'ombre

Pour décrire l'apparence spectrale d'une surface dans l'ombre, on doit d'abord considérer les physiques de génération des couleurs. L'apparence d'une surface est le résultat de l'interaction entre la lumière, les propriétés de réflexion de la surface et la réponse du mécanisme chromatique. La chromaticité est composée de trois filtres de couleur.

Plusieurs travaux ont été effectués pour modéliser l'interaction entre la lumière et la surface de l'objet [SHA 85 ; FOR 03]. La radiance de la lumière $L_r(\lambda, \vec{P})$, réfléchi dans un point \vec{P} dans une surface à trois dimensions est exprimée par l'équation suivante :

$$L_r(\lambda, \vec{P}) = L_a(\lambda) + L_b(\lambda, \vec{P}) + L_s(\lambda, \vec{P}) \quad (2.1)$$

Où les termes de réflexion de la lumière $L_a(\lambda), L_b(\lambda, \vec{P}), L_s(\lambda, \vec{P})$ expriment respectivement la réflexion ambiante, la réflexion du corps et la réflexion de la surface. λ étant la longueur d'onde.

L'expression de l'éclairage ambiant est supposée contenir toutes les lumières réfléchies indirectement sur la surface et ne varie dans aucun cas avec la géométrie de l'environnement. S'il n'existe aucun éclairage direct sur la surface à cause d'un objet qui se situe entre la source de lumière et la surface alors le rayonnement de la lumière réfléchie est exprimé par l'équation 2.2.

$$L_{r,ombre}(\lambda, \vec{P}) = L_a(\lambda) \quad (2.2)$$

Cette expression permet de représenter l'intensité de la lumière réfléchie sur un point d'ombre. Soit $S_R(\lambda)$, $S_G(\lambda)$ et $S_B(\lambda)$ les sensibilités spectrales respectives du rouge, vert et bleu dans les capteurs d'une caméra couleur. Les composantes de couleur de l'intensité réfléchie dans un point (x, y) dans une image à deux dimensions sont exprimées par l'équation suivante :

$$C_i(x, y) = \int_k E(\lambda, x, y) S_{C_i}(\lambda) d\lambda \quad (2.3)$$

Où $C_i \in R, G, B$ sont les réponses de capteurs, $E(\lambda, x, y)$ est le rayonnement de l'image dans le point (x, y) et $S_{C_i}(\lambda) \in \{S_R(\lambda), S_G(\lambda), S_B(\lambda)\}$. L'intervalle de la somme est déterminé par $S_{C_i}(\lambda)$, qui est non nul sur un intervalle de longueurs d'onde limitée K . Comme le rayonnement de l'image est proportionnel au rayonnement de la scène [FOR 03], les mesures d'un capteur pour une position d'un pixel (x, y) représentant un point \vec{P} affecté directement par la lumière sont exprimées pour un vecteur de couleur $\vec{C}(x, y) = (R, G, B)$ par l'équation suivante :

$$C_i(x, y) = \int_k \alpha \left(L_a(\lambda) + L_b(\lambda, \vec{P}) + L_s(\lambda, \vec{P}) \right) S_{C_i}(\lambda) d\lambda \quad (2.4)$$

Où α est un facteur qui exprime la proportionnalité entre l'éclat et l'irradiation. Pour un point dans l'ombre, les mesures sont exprimées pour un vecteur de couleur $\vec{C}_{ombre}(x, y) = (R_{ombre}, G_{ombre}, B_{ombre})$ par l'équation 2.5

$$C_{ombre}(x, y) = \int_k \alpha L_a(\lambda) S_{C_i}(\lambda) d\lambda \quad (2.5)$$

Il en résulte que chacune des trois composantes de couleur RGB, si elle est positive et non nulle, diminue lorsqu'on passe d'une région éclairée à une région ombre, soit $R_{ombre} < R$, $G_{ombre} < G$, $B_{ombre} < B$. La lumière ambiante peut avoir des caractéristiques spectrales différentes à l'égard de la lumière directe [GER 86]. Dans le cas d'une scènes en plein air, la lumière diffusée du ciel diffère dans la composition spectrale par rapport à la lumière directe du soleil [NAD 02].

Dans la plupart des approches proposées, les auteurs essayent d'éviter les procédures de calibrage et les calculs qui dépendent de la caméra [FIN 02], de manière à proposer des algorithmes de segmentation qui peuvent être appliqués même lorsqu'aucun contrôle sur les conditions d'imagerie et la scène n'est possible. Les travaux proposés dans [FUN 01 ; SCH 02a ; PIN 01] supposent que la lumière ambiante est proportionnelle à la lumière directe occluse. Une autre situation est l'existence de plusieurs surfaces de réflectances conduisant à une interrélflexion. Les effets locaux générés par la réflexion interobjet sont extrêmement difficiles à analyser dans les scènes complexes du monde réel [FOR 03]. Les travaux proposés par Salvador et al. [SAL 04] montrent que la méthode proposée est robuste à l'hypothèse mentionnée ci-dessus, même dans les scènes où l'effet d'illumination mutuelle pourrait être non négligeable. Une deuxième propriété spectrale des ombres peut

être dérivée en prenant en considération l'invariance des couleurs photométriques. L'invariance des couleurs photométriques peut être considérée comme des fonctions qui décrivent la configuration de couleur de chaque point de l'image actualisant une ombre, les ombres et les fortes lumières. Les travaux de [GEV 99] ont démontrés que ces fonctions sont invariant par rapport : aux changements de la direction d'observation, à l'orientation de la surface de l'objet et par rapport aux changements des conditions d'éclairage. Soit F l'un des invariants de couleurs photométriques mentionnées ci-dessus. $F_{lumière}$ est la valeur d'un point exposé à la lumière, F_{ombre} est la valeur du même point dans l'ombre. Donc selon les hypothèses citées précédemment $F_{lumière} = F_{ombre}$

2.4 Propriétés géométriques des ombres

L'aspect géométrique d'une ombre dépend des objets et de la disposition de la scène. Cependant, il est possible d'identifier certaines caractéristiques géométriques des ombres (les frontières de l'ombre) sans aucune connaissance de la structure de l'objet ou de la scène. Les frontières de l'ombre peuvent être classées en quatre catégories [SAL 04] : les lignes générées par l'ombre (AB), des lignes d'ombre (DE), les lignes d'occlusion (CD), et les lignes d'ombre cachées (CE). Ces lignes sont représentées par la figure 2.3.

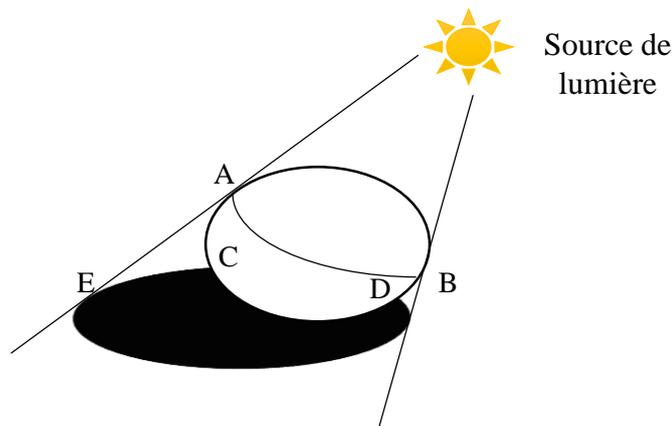


FIGURE 2.3 – Représentation des caractéristiques géométriques de l'ombre

La ligne générée par l'ombre permet de séparer la surface éclairée et la surface non éclairée d'un objet. Si la position de l'observateur est alignée avec la direction de la source lumineuse, cette ligne se cale sur les contours d'un objet. Les lignes d'ombre sont les projections de la ligne générée par l'ombre dans la direction des rayons lumineux. La ligne d'occlusion permet de séparer un objet de son ombre portée. Une ligne d'ombre cachée est une ligne d'ombre correspondant à une ligne générée par l'ombre non visible.

2.5 Modélisation de la luminosité

Lorsqu'une caméra capture l'ombre dans une scène, c'est la quantité de lumière réfléchie de la surface qui est enregistrée. La lumière réfléchie d'un point (x, y) de l'image à l'instant t peut être modélisé selon l'équation suivante :

$$\psi_t(x, y) = \xi_t(x, y) \rho_t(x, y) \quad (2.6)$$

Où $\rho_t(x, y)$ est la réflectance de la surface de l'objet et $\xi_t(x, y)$ est l'illumination. $\xi_t(x, y)$ est une fonction de la direction $L_{x,y}$ de la source de lumière par rapport à la surface normale de l'objet $N_{x,y}$ l'intensité directe de la lumière C_p et la lumière ambiante C_A reçue dans le point de coordonnées (x, y) . L'expression de l'illumination est modélisée selon l'équation 2.7 suivante :

$$\begin{cases} C_p N_{x,y} \cdot L_{x,y} + C_A & \text{aucun objet(ombre inclu)} \\ \lambda_{x,y} C_p N_{x,y} \cdot L_{x,y} + C_A & \text{pénombre} \\ C_A & \text{cône d'ombre} \end{cases} \quad (2.7)$$

Le modèle suivant est basé sur la loi de Lambert qui exprime l'effet de la luminosité avec et sans ombre. Le paramètre $\lambda_{x,y}$ compris entre 0 et 1 décrit la transition à l'intérieur de la zone de pénombre, et dépend fortement de la source de lumière et de la géométrie de la scène. En plus du modèle ci-dessus, les hypothèses générales suivantes, proposées par Stander et al.(Stander 1999)[STA 99a], sont prises par la plupart des méthodes de détection d'ombre existantes :

- **Hypothèse 1.** Le coefficient C_p de l'intensité de la source lumineuse est supposé élevé.
- **Hypothèse 2.** La caméra et l'arrière-plan sont statiques.
- **Hypothèse 3.** L'arrière-plan de la scène est un plan, et la position de la source lumineuse est suffisamment éloignée de l'arrière-plan.
- **Hypothèse 4.** La distance entre l'objet en mouvement et l'arrière-plan n'est pas négligeable par rapport à la distance entre la source de lumière et l'objet en question.

2.6 Modélisation de l'ombre

Soit $p(x, y)$ un pixel de l'arrière-plan. On suppose que le pixel p se trouve en dehors de la zone d'ombre à l'instant t et dans la zone d'ombre à l'instant $t + 1$. Si l'hypothèse 1 est vérifiée à l'instant $t + 1$, alors le paramètre C_p a une très grande valeur et la différence entre les expressions de luminosité aux instants $t, t + 1$ exprimée par l'équation 2.8 a

également une très grande valeur.

$$\xi(x, y) = \xi_{t+1}(x, y) - \xi_t(x, y) \quad (2.8)$$

On note que l'expression de la réflectance de l'arrière-plan ne change pas dans le temps i.e. $\rho_{t+1}(x, y) = \rho_t(x, y)$. Donc, si les deux premières hypothèses sont vérifiées, le résultat de la différence entre les deux équations donnera une valeur très grande dans la présence de l'ombre sur l'arrière-plan. Par conséquent, il suffit de définir un seuil sur chaque pixel de l'image pour déterminer si ce dernier est affecté par une ombre.

Plusieurs méthodes dans l'état de l'art utilisent ces hypothèses pour exprimer l'effet de l'ombre sur la surface de l'arrière-plan. Cependant, l'hypothèse 2, qui impose un arrière-plan est parfaitement statique dans le sens où aucun changement de la luminosité n'est toléré, est pratiquement impossible et l'utilisation de cette modélisation nécessite l'implémentation d'algorithmes supplémentaires pour stabiliser l'arrière-plan par rapport aux changements de luminosité.

2.7 Caractéristiques utilisées pour la détection de l'ombre

Nous allons présenter les caractéristiques les plus communément utilisées dans les systèmes basés sur l'estimation de l'arrière-plan pour la détection des objets en mouvement. Le bon fonctionnement de ces derniers nécessite un arrière-plan dépourvu de tout objet (ombre incluse) au moins dans la phase d'initialisation du modèle de l'arrière-plan. Nous rappelons que la problématique tourne autour des objets en mouvement et par conséquent les ombres en mouvement.

2.7.1 Intensité

La supposition la plus simple et la plus évidente pour la détection des ombres portées consiste à considérer que les régions affectées par l'ombre deviennent plus sombres, car la source de lumière ne passe pas [HSI 03]. Cependant, la région garde un certain niveau de luminosité grâce à la lumière ambiante. Cette supposition peut être utilisée pour prédire l'effet de l'ombre sur une région par la définition d'un intervalle d'intensité [LEO 07]. Cette supposition est nécessaire et non pas suffisante pour détecter l'ombre, mais peut être utilisée comme une étape préliminaire pour sélectionner les régions susceptibles d'être affectées par une ombre.

2.7.2 Chromaticité

La plupart des méthodes de détection des ombres basées sur les caractéristiques spectrales utilisent les couleurs comme critère de discrimination entre l'objet et son ombre

[HOR 99]. Dans ce type de système, les régions affectées par une ombre deviennent plus sombres, mais gardent toujours leur couleur. Cette supposition est visuelle dans notre vie quotidienne. Si on prend par exemple une surface ou un objet avec une couleur rouge ou bleu ou autre couleur sans ombre, ces derniers garderont toujours les mêmes couleurs sous l'ombre. En effet, la couleur devient plus sombre, mais elle ne change pas.

Pour bien détecter les ombres en suivant cette supposition, il est indispensable de bien choisir l'espace de couleur utilisé. Effectivement, l'ombre rend les objets plus sombres, donc il faut choisir un espace de couleur qui permet une meilleure séparation entre la luminosité et la chromaticité.

Plusieurs espaces colométriques ont été utilisés dans l'état de l'art pour séparer entre l'intensité et la chromaticité [CAV 05 ; CHE 10b ; CUC 03], les plus communément utilisés sont : HSV, HSL, YUV, YCbCr et RGB normalisé. Une combinaison entre ces derniers a fait également l'objet de plusieurs travaux afin de bénéficier de la puissance de chaque espace et d'éliminer le maximum d'inconvénances.

Les méthodes basées sur la chromaticité sont très simples à implémenter et se distinguent par rapport aux autres méthodes de détection d'ombre par une faible consommation d'espace mémoire. En plus, les méthodes basées sur la chromaticité sont les plus adéquates pour les traitements en temps réel à cause de leur faible temps de calcul.

Malgré les avantages multiples des méthodes basées sur la chromaticité, ces derniers ne tolèrent pas les bruits [NAD 04]. En effet, ces méthodes traitent l'information au niveau du pixel, ce qui rend les perturbations les plus faibles, nuisible sur leurs rendements. En plus de variations locales, ces modèles sont très sensibles aux changements de luminosité ayant une grande amplitude.

2.7.3 Les propriétés physiques

Les modèles linéaires supposent que les sources de luminosité produisent des lumières purement blanches ce qui n'est pas toujours le cas. En effet, la principale source de lumière est le soleil qui produit une lumière totalement blanche. Cependant, les objets sont illuminés au moins par le soleil et par la lumière réfléchi des autres objets tels que le ciel, la terre, les arbres, l'eau, etc.

Dans la plupart des cas, la lumière du soleil domine les autres lumières ce qui simplifie la création d'un modèle pour exprimer l'effet de la lumière sur les couleurs, mais ce n'est pas toujours le cas. Lorsque la lumière du soleil est bloquée par exemple par les nuages, l'effet des autres lumières réfléchies augmente. Ces derniers étant des sources de lumière de couleur différente que la lumière blanche, l'effet de la lumière sur les couleurs deviennent totalement différent.

Plusieurs travaux ont essayé de résoudre ce problème. Les travaux de Nadimi et al. [NAD 04] proposent un système prenant en charge plusieurs sources de lumière pour

mieux prédire l'effet de l'ombre sur les couleurs. D'autres travaux [HUA 09 ; MAR 08] ont essayé de généraliser le modèle physique pour prendre en considération plusieurs sources de lumière et cela à la fois dans les environnements intérieurs et extérieurs. Une autre variante de ces modèles tente de gérer le problème de diversité des sources de lumière par un apprentissage [JOS 08] effectuer sur l'effet de l'ombre sur les mêmes pixels en appliquant à chaque fois une source de lumière de couleur différente [LIU 07]. Les méthodes basées apprentissage ont montré leur efficacité par rapport aux autres méthodes [MAR 07]. Malheureusement, les propriétés spectrales lui font défaut et les limites des méthodes physiques se situent dans les objets qui ont une chromaticité similaire à l'arrière-plan.

2.7.4 La géométrie

Les méthodes basées sur la géométrie utilisent les informations telles que les contours, la taille ainsi que l'orientation pour la détection des ombres. En effet, chaque forme projette une allure d'ombre bien spécifique sous un certain angle de la source de lumière. Cette information permettra de séparer entre l'objet et son ombre [HSI 03].

Le principal avantage des caractéristiques géométriques est la possibilité de détecter les ombres en utilisant uniquement le frame en court de traitement et par conséquent, aucune prédiction a priori relative aux précédents frames n'est nécessaire pour aboutir au résultat [NIC 06]. Comme toutes méthodes, les modèles géométriques souffrent également de plusieurs inconvénients. Le principal inconvénient réside dans le besoin d'une forme géométrique bien précise et avec des contours bien définis, tel que les voitures, pour le bon fonctionnement du système. En plus, l'orientation ainsi que la direction de la source de lumière doivent être uniques et connues à priori et par conséquent chaque système traite un environnement bien défini.

Les travaux récents ont également montré que les méthodes basées sur la géométrie ne permettent pas de traiter des vidéos contenant, en même temps, plusieurs objets en mouvement [CHE 10a]. La conception de ces derniers ne permet pas de détecter les ombres multiples d'un seul objet dans le cas où l'environnement est éclairé par plusieurs sources de lumière.

2.7.5 Les textures

Les méthodes basées textures exploitent le fait que l'ombre change la chromaticité, mais n'affecte jamais les textures [JAV 02]. Cette caractéristique spécifique à la texture permet de séparer sans aucun problème entre les régions qui ont subi un changement à cause de la présence d'un objet et ceux affectées par la présence de l'ombre. En effet, les régions qui gardent leurs textures intactes ne seront pas prises en considération dans la

détection des objets en mouvement et par conséquent le problème de l'ombre ne se pose même pas.

La plupart des méthodes basées sur les textures comme caractéristiques de discriminations utilisent deux étapes [SAN 12]. La première étape consiste à sélectionner les pixels des régions qui peuvent être des régions d'ombre. La deuxième étape utilise une classification des régions candidates selon l'existence ou non d'une corrélation entre les textures pour décider si les régions sélectionnées sont des ombres ou ils font partie des objets en mouvement. La sélection des régions candidates est généralement basée sur les caractéristiques spectrales telles que la chromaticité. Une fois les régions candidates sélectionnées, le système calcule une fonction de corrélation entre les textures du frame en cour de traitement et un arrière-plan de référence supposé ne contenant aucune ombre. Si le résultat de la fonction est supérieur à un certain seuil fixé à priori (définissant la similarité), le système classe les pixels candidats en tant qu'ombre sinon ils seront classés comme objet en mouvement.

Il est clair que le résultat, des systèmes basés texture, dépend fortement du choix de la fonction de corrélation. Pour cette raison, plusieurs travaux ont été effectués pour améliorer le système à ce niveau. Parmi les méthodes de corrélation les plus utilisées, nous citons : le gradient [SAN 10], les contours, la corrélation croisée [TIA 05], les champs de Markov [QIN 10], les transformations en cosinus discrètes [XU 05], les transformations orthogonales [ZHA 06], le filtre de Gabor [LEO 07], et la liste n'est pas exhaustive.

Les textures offrent la possibilité de détecter de façon très performante les ombres vu que ces dernières sont complètement immunisées contre les changements de luminosité et ne dépendent d'aucune couleur. En plus, les textures offrent une plateforme fortement distinctive. Cependant, les textures consomment du temps pour calculer, pour chaque pixel de l'image, la corrélation entre un pixel et ces voisins limitant ainsi leurs utilisations en temps réel (vidéosurveillance) et les systèmes embarqués (appareil photo numérique, robot, missile).

2.7.6 Caractéristiques temporelles

Les caractéristiques temporelles sont généralement utilisées comme un filtre supplémentaire [CAV 05] ou une étape de post-traitement [LIU 07] pour améliorer les résultats de la détection en gardant uniquement les pixels consistants au court du temps [NIC 06]. Les systèmes à base de caractéristiques temporelles utilisent le fait que les objets en mouvement et l'ombre partagent le même style de déplacement et par conséquent, les mêmes filtres de cohérence temporelle qui ont été appliqués aux objets peuvent être appliqués à l'ombre.

2.8 Conclusion

Nous avons présenté dans ce chapitre un état de l'art sur la détection et la suppression des ombres. Nous avons commencé par définir l'ombre et expliquer en détail les types d'ombre générés par un objet. Les caractéristiques de l'ombre rendent les systèmes de détection des objets en mouvement incapable de faire la différence entre l'objet et son ombre. Pour cela, plusieurs techniques ont été développées dans l'état de l'art pour remédier à ce problème. Nous avons également présenté le modèle de lumière vu que l'ombre est reliée directement à l'effet de la lumière sur les objets. Finalement, nous avons mis les projecteurs sur les caractéristiques techniques les plus utilisées pour la modélisation de l'ombre et qui permettent de donner un aperçu sur les différentes catégories de méthodes de détection et de suppression des ombres.

3

Reconnaissance des formes

3.1 Introduction

La reconnaissance des formes est une discipline qui permet l'étude des systèmes automatiques ou semi-automatiques capables de reconnaître les formes qu'on leur présente. Les mots reconnaissance et formes sont devenus un vocabulaire courant de tous les jours. Cette discipline n'a pas besoin de définition formelle dans le cas général. Dans une tentative d'imiter la nature, la reconnaissance des formes est le domaine qui permet aux machines d'hériter un peu des facultés de l'être humain. Cependant, ce dernier était et restera le système de reconnaissance de formes le plus parfait qui existe. En effet, la diversité des tâches de reconnaissance que nous pouvons accomplir sur des formes à grande variabilité est toujours impressionnante. Effectivement, nous distinguons, sans effort, toutes les formes visuelles et sonores.

Les progrès scientifiques et techniques nous ont permis d'essayer d'imiter certaines des facultés des humains à l'aide des machines. L'idée principale n'étant pas de remplacer complètement l'homme par une machine, mais de confier à une machine certains travaux de reconnaissance simples effectués par l'homme. Si nous prenons, par exemple, la détection des anomalies dans des milliers de chiffres de mesures, un être humain prendra énormément de temps pour pouvoir distinguer ensuite séparer les anomalies de ceux qui sont corrects. D'autres exemples, tels que chercher un mot dans un texte, détecter et facturer les produits dans un magasin ou encore surveiller des zones par le biais d'une caméra, ont montré l'inefficacité des humains lorsqu'il s'agit de traiter une masse importante de

données. Une machine peut effectuer ces travaux beaucoup plus rapidement que l'homme.

Sur le plan social, l'homme est souvent confronté à des tâches monotones, ennuyeuses, répétitives ou dangereuses. Il n'est pas intéressant pour une secrétaire de taper un texte à la machine à écrire, surtout lorsqu'elle n'en comprend pas le sens. Il en est de même d'un ouvrier qui, à longueur de journée, doit placer les mêmes vis dans les mêmes trous. Il peut être très dangereux pour un policier de désamorcer une bombe ou un robot peut effectuer des tâches similaires. Sur le plan économique, il peut être moins onéreux, par exemple, d'automatiser le contrôle de la qualité sur une chaîne de production ou le tri sur une chaîne de montage.

3.2 Contexte et problèmes

Les exemples précédents montrent la diversité des problèmes de reconnaissance des formes. En effet, cette discipline se trouve au croisement de beaucoup d'autres dont les principales sont :

- Les statistiques,
- La linguistique,
- La recherche opérationnelle,
- La théorie des communications,
- L'informatique,
- La biologie,
- L'optique,
- L'électronique.
- etc.

Pour aboutir à des applications aussi diverses que la robotique ou l'analyse d'images médicales, il n'y a malheureusement pas une théorie unifiée de reconnaissance des formes. Ceci est principalement dû à deux raisons. La première est la grande variété dans les concepts provenant des disciplines citées, les applications et les moyens techniques. Le second est le grand écart qui existe entre la théorie et la pratique de la reconnaissance des formes qui reste, jusqu'à aujourd'hui, non obturer. Les théories disponibles ne permettent pas de cerner tous les problèmes pour déduire des systèmes optimisés et pratique. Même au niveau des produits commercialisés, les entreprises se contentent des performances raisonnables, acceptables ou utilisables sans insister sur une rigueur absolue.

Malgré ces difficultés, il existe trois approches principales de la reconnaissance des formes : l'approche statistique (basé sur la théorie statistique), l'approche structurelle (syntaxique) et l'approche mixte. Donc, en fonction des paramètres utilisés par le système, cette dernière fera partie des méthodes statistiques, structurelles ou mixtes de classification. On peut, également, diviser ces méthodes en méthodes supervisées, non supervisées ou semi-supervisées selon le type d'information disponible sur le système. En effet, s'il

s'agit de données étiquetées alors ce sera forcément un apprentissage supervisé, dans le cas contraire un apprentissage non supervisé ou semi-supervisé est indispensable.

Dans le contexte de la détection des formes dans une vidéo, l'environnement subit des évolutions et des changements au cours du temps ce qui nécessite un apprentissage continu pour prendre en compte la nature évolutive de ces systèmes. On trouve les modèles incrémentaux parmi les méthodes les plus communément utilisées pour un apprentissage continu qui permettent d'utiliser toutes les informations acquises pour améliorer l'estimation des classes en y ajoutant de nouvelles configurations. Cependant, certaines formes, qui ont une présence temporelle de quelques frames dans la vie du système, peuvent devenir rapidement obsolètes et donc non représentatives de la classe. En plus, ces méthodes commencent à consommer énormément d'espace mémoire et prendront des temps d'apprentissage exponentiel à cause de la masse importante des formes incluses dans chaque cycle. L'apparition des méthodes dite dynamique a permis de résoudre ce type de problème grâce à un mécanisme performant qui permet de supprimer les formes non représentatives des caractéristiques actuelles des classes après leur évolution.

La figure suivante présente une taxonomie des méthodes de reconnaissance de formes sous leur différent aspect.

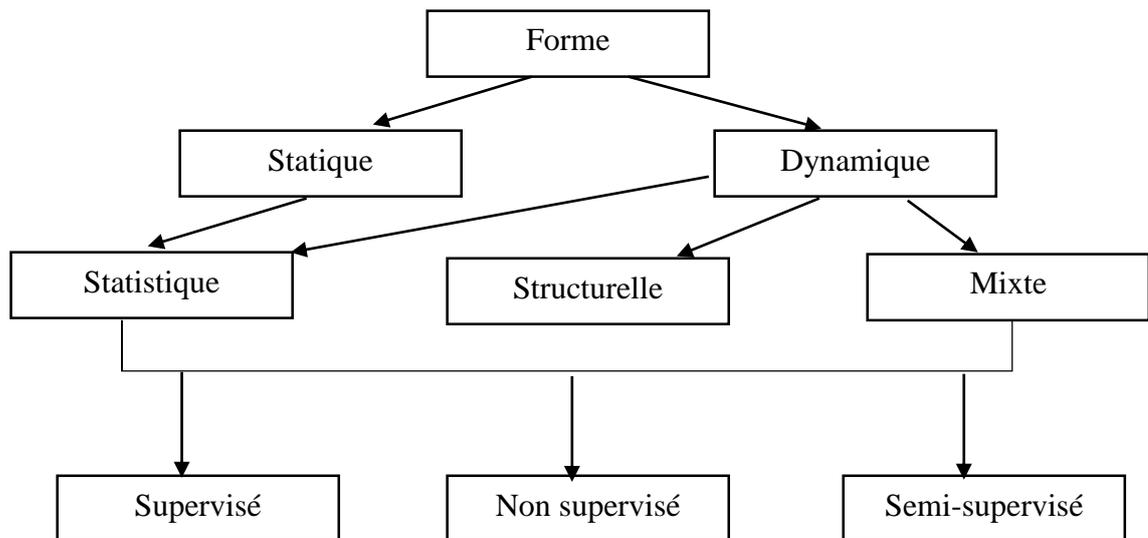


FIGURE 3.1 – Taxonomie des méthodes de reconnaissance des formes

Nous allons présenter dans ce qui suit une synthèse des méthodes de reconnaissance des formes qui existe dans la littérature.

3.3 Principe des RDFs

Le rôle d'un système de reconnaissance des formes est de reconnaître une forme parmi d'autres et de pouvoir la classer par le biais d'un classifieur. La classification se résume

dans le calcul d'une fonction d'appartenance de la nouvelle forme par rapport à toutes les classes. Ensuite, essayer d'attribuer à cette dernière, la classe la plus représentative.

On trouve dans la littérature deux aspects selon l'ensemble d'apprentissages à savoir les méthodes paramétriques et les méthodes non paramétriques. Dans le premier aspect, les données d'apprentissage sont supposées indépendantes et ont la même distribution de probabilité [DUB 90]. Cependant, ce n'est pas toujours le cas dans la pratique. Tandis que les méthodes non paramétriques se basent sur les fonctions d'appartenance des classifieurs.

Dans la littérature, on trouve trois manières pour réaliser ces fonctions d'appartenance. La première manière se base sur la création des régions de décision après une étape d'apprentissages tels que les réseaux de neurones [RIP 96] et les SVM [VAP 13]. La deuxième est une estimation de la densité de probabilité telle que les KPPV [COV 67]. Tandis que la troisième utilise les métaheuristiques [DUM 86].

3.4 Les enjeux dans les RdFs

Pour tirer le maximum d'avantage d'une méthode de classification, le développeur doit avoir, à l'avance, une connaissance parfaite et complète du mode de fonctionnement du système. En effet, le système peut être exposé à des bruits générés soit par les capteurs ou par les données eux-mêmes. En plus, on ne peut prévoir tous les scénarios possibles lors de la mise en œuvre du système. Cependant, plusieurs approches proposent des solutions pour résoudre ces problèmes tels que la théorie du flou pour les données bruitées et un apprentissage continu pour que le système puisse les prendre en considération.

3.5 Description générale du système de reconnaissance

Généralement, un système de reconnaissance des formes se divise en deux principales phases à savoir l'apprentissage et la classification. Une troisième étape omise permet premièrement de corriger les bruits dans les données en entrées (prétraitement) et de minimiser les paramètres nécessaires pour établir l'espace de représentation (extractions des primitives).

3.5.1 Prétraitement

Le prétraitement est une suite d'opérations qui permet de réduire le volume des données à traiter et d'éliminer les bruits introduits par les outils d'acquisition. Cette phase utilise une suite de plusieurs algorithmes et filtres tels que la binarisation, la squelettisation, la détection des contours, le changement d'espace de couleur, la transformation de

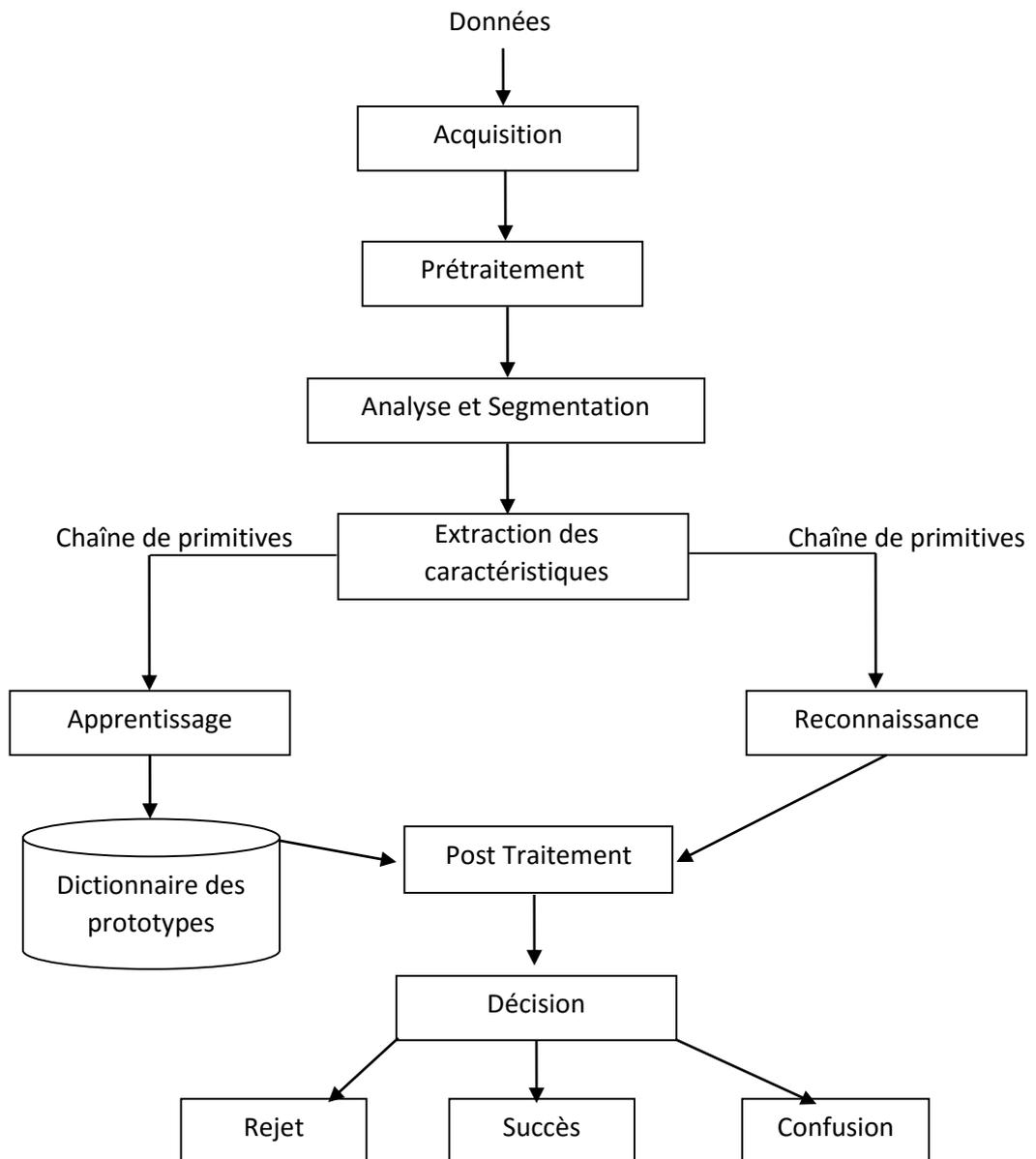


FIGURE 3.2 – Schéma général d'un Système de Reconnaissance des formes

fourrier, la transformation en ondelette, la transformation en cosinus discret, les filtres autorégressifs, les transformations géométriques et affines, le filtre médian, etc. Cette phase est très importante dans la réussite des classifieurs et le retraitement utilisé se diffère d'une application à une autre.

3.5.2 Extraction des primitives (caractéristiques)

Les systèmes de reconnaissance utilisent des vecteurs de caractéristique pour pouvoir reconnaître les formes ou chaque vecteur se comporte comme une signature d'une forme. La difficulté de cette étape est de déterminer quelles caractéristiques employées pour ob-

tenir un bon résultat de classification.

L'objectif de l'extraction et de la sélection des caractéristiques est d'identifier les caractéristiques qui sont importantes dans la discrimination de classes de formes. Cela signifie de trouver, si c'est possible, les primitives les plus robustes, les plus pertinentes (afin de diminuer la taille du vecteur de primitives) et les moins coûteuses en calcul. Le résultat de cette phase d'extraction de primitives est une séquence de symboles qui représente la forme et qui traduit, soit la présence ou l'absence (cas booléen) ou la valeur associée (cas réel) à la caractéristique concernée.

En fonction de l'objectif fixé et de la méthode d'extraction choisie, l'approche de l'extraction des primitives peut être systématique ou heuristique. La modélisation et le codage conduisent à une approche systématique dans la mesure où l'objectif fixé est la détermination d'une représentation complète de la forme, même de façon approximative.

Dans la modélisation, les primitives sont obtenues a posteriori, par le résultat de l'approximation, en ce qui concerne le codage, les catégories de primitives sont définies a priori. Un test, qui est par exemple réalisé à l'aide d'une sonde, permet de valider la présence de chacune des primitives sur l'ensemble de la forme. Le paramétrage conduit plutôt à une approche heuristique. Dans ce cas, on ne cherche pas nécessairement une représentation complète, mais seulement des indices significatifs. De la même façon que dans le cas du codage, ces indices sont des primitives définis a priori.

L'une des raisons pour lesquelles cette étape pose un problème est que la plupart des techniques d'extraction s'accompagnent d'une perte d'information irrémédiable. De ce fait, il faut effectuer un compromis entre quantité et qualité de l'information, mais peu d'études théoriques sont faites sur ce sujet où l'intuition prédomine.

Pour un problème de classification donné, la principale qualité recherchée pour un ensemble de caractéristiques est sa faculté de rassembler les objets appartenant à une même classe dans une même partition de l'espace de représentation, tout en éloignant autant que possible des autres. Cette qualité est communément appelée pouvoir discriminant de l'ensemble de caractéristiques.

3.5.3 Apprentissage et Classification

La classification est l'élaboration d'une règle de décision qui transforme les attributs caractérisant les formes en appartenance à une classe (passage de l'espace de codage, vers l'espace de décision). Avant qu'un modèle de décision ne soit intégré dans un système de reconnaissance des formes, il faut avoir procédé auparavant à deux étapes : l'étape d'apprentissage et l'étape de test. L'étape d'apprentissage consiste à caractériser les classes de formes de manière à bien distinguer les familles homogènes de formes. C'est une étape clé dans le système de reconnaissance. On distingue deux types d'apprentissages : apprentissage supervisé, apprentissage non supervisé et apprentissage semi-supervisé.

Dans le cas de l'apprentissage supervisé, un échantillon représentatif de l'ensemble des formes à reconnaître est fourni au module d'apprentissage. Chaque forme est étiquetée par un opérateur appelé professeur, cette étiquette permet d'indiquer au module d'apprentissage la classe dans laquelle le professeur souhaite que la forme soit rangée. Cette phase d'apprentissage consiste à analyser les ressemblances entre les éléments d'une même classe et les dissemblances entre les éléments de classes différentes pour en déduire la meilleure partition de l'espace des représentations. Les paramètres décrivant cette partition sont stockés dans une table d'apprentissage à laquelle le module de décision se référera ensuite pour classer les formes qui lui sont présentées.

Dans le cas de l'apprentissage non supervisé, on fournit au système de reconnaissance un grand nombre de formes non étiquetées. L'étape de la classification va se charger d'identifier automatiquement les formes appartenant à une même classe.

Le troisième type concerne les méthodes semi-supervisées [COZ 03] qui utilisent les informations connues, c'est-à-dire les formes et classes connues, pour estimer les caractéristiques des classes et leurs fonctions d'appartenances tout en utilisant également l'apprentissage non supervisé pour détecter les nouvelles classes et apprendre leurs fonctions d'appartenance.

L'étape de test permet d'évaluer la performance du classifieur pour un apprentissage donné. C'est une étape importante, car elle peut mettre en cause le choix des primitives ou le choix de la méthode d'apprentissage. En effet, il est difficile de trouver a priori les primitives pertinentes et la méthode d'apprentissage la plus adaptée au problème posé. D'où l'utilité de procéder par itérations successives. Ces itérations consistent à extraire des primitives jugées utiles au problème de reconnaissance à résoudre et à tester la performance du système avec cet ensemble de primitives. Au fur et à mesure que les performances du système souhaitées ne sont pas atteintes, alors il suffit de trouver à nouveau une nouvelle famille de primitives ou de combiner les primitives extraites avec de nouvelles primitives.

Ces phases d'apprentissage et de test sont réalisées préalablement à l'intégration du module de décision dans le système de reconnaissance. Dans tous les cas, on peut permettre au système de reconnaissance d'itérer les phases d'apprentissage et de test tant qu'on n'a pas atteint les performances désirées. Le calcul de cette performance est le résultat du classifieur utilisé. Pour construire un classifieur, il existe trois approches : structurelle, statistique et hybride.

3.6 Les approches de classification

3.6.1 Les approches statistiques

Cette approche consiste à déterminer des caractéristiques extraites d'une forme pour les caractériser d'une manière statistique. Elle a besoin d'un nombre élevé d'exemples

afin de réaliser un apprentissage correct des lois de probabilité des différentes classes. Autrement dit, cette approche bénéficie des méthodes d'apprentissage automatique s'appuyant sur des bases théoriques connues telles que la théorie de la décision bayésienne, les méthodes de classification non supervisées et l'analyse en composantes principales. Les deux principales familles de méthodes utilisées sont les méthodes paramétriques et les méthodes non paramétriques.

Les méthodes paramétriques opèrent sous l'hypothèse que les classes étudiées suivent une distribution de probabilité d'une certaine forme connue a priori. La prise de décision consiste à déterminer la classe pour laquelle la forme inconnue présente la probabilité d'appartenance maximale. Elles exigent des bases d'apprentissage assez importantes pour une estimation correcte des paramètres de la distribution supposée. L'approche statistique englobe : la règle de Bayes [BEL 92], les méthodes neuronales et les chaînes de Markov [MAH 99].

Les méthodes non paramétriques, les lois de probabilité sont inconnues pour une des classes. Le problème revient à établir des frontières de décision entre les classes. Les techniques les plus utilisées en reconnaissance de formes sont : la méthode du plus proche voisin [BEL 92], et la méthode d'appariement de graphes [BEL 92]. Pour plus amples d'informations, Kunt a décrit un ensemble de méthodes statistiques en reconnaissance de formes [KUN 00].

3.6.2 Les approches structurelles

Ces approches consistent à mettre en relation la structure des formes analysées et la syntaxe d'un langage formel. La description des formes est réalisée par l'intermédiaire de phrases et le problème de classification est ramené à un problème d'analyse de grammaire (parsing). De manière générale, les approches syntaxiques ou structurelles permettent la description de formes complexes à partir de formes élémentaires. Ces dernières, encore appelées caractéristiques, sont extraites directement des données présentes en entrée du système.

La différence principale entre ces méthodes et les méthodes statistiques est que ces caractéristiques sont des formes élémentaires et non pas des mesures. Une autre différence est qu'elles introduisent la notion d'ordre dans la description d'une forme. Les méthodes les plus répandues utilisent le calcul de distance d'édition entre deux chaînes [ALB 12] et la programmation dynamique [MIC 84].

Malgré leur nature différente, les approches statistiques et structurelles peuvent être combinées aux mêmes domaines d'application. Le choix d'une approche peut être lié à des contraintes matérielles telles que la taille de la base d'apprentissage disponible, le temps de calcul requis et la taille mémoire nécessaire. L'utilisation conjointe des deux approches peut être une solution optimale pour le problème de reconnaissance de formes.

3.6.3 Les approches hybrides

Une dernière approche est envisageable : l'approche hybride. La reconnaissance par modèles de Markov cachés on est une. En effet, ils utilisent une approche statistique tout en ayant la possibilité d'utiliser des descriptions structurelles. Il existe également des approches qui couplent plusieurs techniques à savoir : les modèles de Markov et les réseaux de neurones, les modèles de Markov cachés et k- plus proches voisins, etc.

3.7 Nature des formes traitées

Dans la pratique, on peut être confronté à deux types de formes selon l'évolution des caractéristiques avec le temps à savoir : les formes statiques et les formes dynamiques. La différence entre les deux aspects se résume dans la manière de représenter les formes. Si la forme est représentée par un point dans l'espace alors nous sommes en train de parler de la forme statique. Le deuxième aspect est une représentation multidimensionnelle de la trajectoire, ce qui nécessite l'ajout du temps comme une dimension supplémentaire [ANG 01]. Cette représentation peut être aussi appliquée sur les classes des formes. En effet, les classes peuvent aussi être statiques ou dynamiques. Les classes statiques sont représentées par des zones restreintes contenant des formes similaires dans l'espace de représentation. Pour ces classes, l'ordre d'arrivée des formes n'est pas lié à leur valeur d'appartenance et les paramètres du classifieur restent inchangés.

On trouve dans la littérature, plusieurs méthodes qui traitent les formes statiques. Les plus communément utilisées sont : des K-Plus Proches Voisins (KPPV) [DAS 91], les Séparateurs à Vaste Marge (SVM) [VAP 98], les méthodes Bayésiennes [FRE 89], l'Analyse en Composantes Principales (ACP) [JOL 02 ; ELS 09], Fuzzy Pattern Matching (FPM) [CAY 82], Fuzzy C-Means (FCM) [BEZ 13], ainsi que de nombreuses déclinaisons de ces méthodes, etc. Cependant, les systèmes ont tendance à être constamment en évolution entre leurs différents modes de fonctionnement, on parle alors de systèmes évolutifs (Angelov, 2010)[ANG 10] pour lesquels il est nécessaire d'utiliser des méthodes de classification dynamique [KIF 04].

3.8 Méthodes de reconnaissance des formes

Nous allons aborder, dans cette section, les méthodes de reconnaissance de formes les plus populaires. Vu le nombre important des variantes proposées dans la littérature, nous allons détailler uniquement les versions de base en faisant référence à quelques versions améliorées.

3.8.1 Réseaux de neurones

La théorie des réseaux de neurones est une approche s'appliquant principalement au monde de l'intelligence artificielle, venue renforcer l'insuffisance de celle-ci dans certains domaines particuliers tels que l'analyse des images et de la parole, la classification, et la compréhension du langage naturel.

Les techniques neuronales sont par nature des méthodes de classification paramétriques puisque la phase d'apprentissage consiste à estimer les paramètres qui permettront de choisir la fonction de discrimination dans un ensemble de fonctions possibles.

Les réseaux de neurones sont des méthodes connexionnistes dont la structure est organisée en différentes couches qui sont reliées les unes aux autres. Ils sont basés sur une représentation mathématique (statistique) simplifiée des neurones biologiques ; on parle alors de l'utilisation de neurones formels.

Les réseaux de neurones possèdent une couche d'entrées, une ou plusieurs couches cachées et une couche de sortie. La couche d'entrées correspond aux vecteurs formes de l'espace de représentation. Les couches cachées contiennent les prototypes des classes. La couche de sortie définit le nombre de classes connues et permet de connaître le degré d'appartenance d'une observation à une classe.

Pour chaque nouvelle observation présentée à l'entrée d'un réseau de neurones, un neurone formel fait une somme pondérée des potentiels d'actions qui lui parviennent, puis s'active suivant la valeur de cette sommation pondérée. Chacun de ces potentiels est une valeur numérique qui représente l'état du neurone qu'il a émis. Si cette somme dépasse un certain seuil, le neurone est activé et transmet une réponse (sous forme de potentiel d'action) dont la valeur est celle de son activation. Si le neurone n'est pas activé, il ne transmet rien.

Lorsqu'on aborde le domaine des réseaux de neurones, on regroupe un certain nombre de modèles. Pour décrire un modèle, il convient d'indiquer ses caractéristiques :

- Une architecture : soit monocouche ou multicouches.
- Une règle de propagation et d'activation.
- Une règle d'apprentissage.

3.8.1.1 Le Perceptron de ROSENBLATT

Introduit par Frank Rosenblatt en 1958[ROS 58], le Perceptron est organisé en couches, La première couche appelée «*rétilne*» est constituée de cellules d'entrée, recueille les signaux d'entrée. La deuxième couche, appelée «*couche d'association* », permet d'interconnecter la couche d'entrée et la couche de sortie. Les connexions entre cette couche et la couche précédente sont pondérées par des poids qui sont fixés et qui n'évoluent pas pendant l'apprentissage. La troisième couche, appelée «*couche de décision* », contient les neurones généralement logiques et à fonction d'activation à seuil.

3.8.1.2 Adaline de Windrow–Hoff

Inventé en 1960 par Bernard Windrow et Ted Hoff [WID 60], ce modèle possède deux couches, une couche d'entrée et une couche de sortie. Les neurones de la première couche envoient directement leurs entrées vers la couche de sortie sans modifier le signal. Chaque neurone de la couche de sortie effectue la somme pondérée de ses entrées et compare le résultat avec un seuil.

3.8.1.3 Le modèle de KOHONEN

Également connu sous l'appellation de " Carte Auto – organisatrice de KOHONEN "[KOH 82]. Ce type de réseau de neurones consiste en une situation où plusieurs neurones entrent en compétition pour être activés. Dans cette situation, un seul neurone est activé (winner-takes-all neuron) suite à la compétition. Le réseau combine une couche d'entrée avec une couche compétitive de neurones entraînés de façon non supervisée (la sortie désirée n'est pas définie).

3.8.1.4 Le Perceptron multicouche

Un réseau multicouche est une extension du modèle monocouche issu du Perceptron [RUM 86], avec une ou plusieurs couches cachées successives entre la couche d'entrée et la couche de sortie. La connectivité la plus courante consiste à définir les connexions uniquement d'une couche vers la couche suivante. Il possède les caractéristiques suivantes :

- Chaque neurone est connecté en entrée à tous les neurones de la couche précédente.
- La sortie d'un neurone est connectée à tous les neurones de la couche suivante.
- Il n'y a pas de boucle.
- Le nombre de couches cachées ainsi que le nombre de neurones par couche peut être choisi en fonction de l'application.

3.8.1.5 Le modèle de HOPFIELD

Ce modèle est un réseau totalement connecté dont les valeurs d'entrée et de sorties appartiennent à l'intervalle $(-1,1)$ et les poids sont symétriques[HOP 86]. La symétrie des connexions assure la convergence du réseau. L'apprentissage dans un réseau de HOPFIELD est parfois qualifié d'apprentissage par cœur, par opposition avec les algorithmes d'apprentissage par essai et erreur.

Plusieurs améliorations des réseaux de neurones ont été proposées dans la littérature, nous citons : evolving Neural Network [ANG 04], Similarity MAPping [HER 09], General Fuzzy Min-Max Neural Network [GAB 00], fuzzy Adaptive Resonance Theory MAPs [GRO 88], et la liste n'est pas exhaustive. Ces réseaux ont une structure évolutive, c'est-à-dire que leur nombre de neurones et de connexions change en utilisant les nouvelles

informations.

3.8.2 Méthodes à noyau

Les méthodes basées noyau utilisent un apprentissage non paramétrique pour la modélisation des formes complexe. Cette dernière n'utilise aucune connaissance a priori sur la distribution des données d'apprentissage. En effet, la minimisation de l'erreur se fait d'une manière empirique en utilisant des modèles simples et adaptés à différents types de classes.

Les méthodes à noyaux aboutissent à un problème de minimisation convexe, sous forme duale ou le plus souvent quadratique dont la résolution est rapide. On trouve dans la littérature plusieurs noyaux, on peut citer le noyau gaussien [KEE 03], le noyau d'ondelette [ZHA 04], le noyau polynomial [KIM 02], etc.

Actuellement, les Séparateurs à Vaste Marge (SVM) sont l'une des méthodes à noyau les plus utilisées. L'SVM est une méthode adaptée à la classification en ligne. Elle est fortement développée dans la littérature [SUY 99 ; RUP 01 ; DUF 09 ; VAP 13]. De nombreuses versions de SVM existent afin de s'adapter à toutes sortes de problèmes en utilisant différents paramètres et optimisations. On peut citer les SVM linéaires, dont les vecteurs de supports permettent de discriminer les classes en définissant une frontière de décision linéaire et les SVM non linéaires adaptés aux classes dont la séparation ne peut pas être linéaire sans transformation de l'espace de représentation [BOS 92].

Dans leur version préliminaire, les SVM utilisent un apprentissage supervisé et ne peuvent attribuer que des décisions binaires. Donc, le modèle utilise au maximum deux classes pour la classification d'une forme dans un espace de décision. Cependant, des versions multiclassées ont été développées afin de considérer les cas plus complexes comprenant plus de deux classes.

Lorsque plusieurs classes sont présentes, le problème de classification est sous-traité par plusieurs SVM. Pour cela, il y a la stratégie « un contre tous » [GUE 02 ; BOR 03] où l'on construit autant de SVM qu'il y a de classes ; on cherche alors à distinguer les données d'une classe parmi les autres classes. Une autre stratégie appelée « un contre un » [SAL 01] consiste à construire autant de SVM qu'il y a de paires de classes. Pour classifier les formes, les méthodes SVM cherchent l'hyperplan, c.-à-d. la frontière de décision, qui maximise la distance entre les points de chaque classe.

Dans la phase d'apprentissage, la méthode SVM calcule une fonction qui lie chaque point de l'ensemble d'apprentissages à une classe. Cette fonction définit l'hyperplan séparateur. L'hyperplan est basé sur un certain nombre de points caractéristiques de la séparation des classes c.-à-d. les vecteurs de supports. Le nombre et le choix des vecteurs de supports sont estimés en minimisant le critère d'optimisation des paramètres. L'optimisation des paramètres passe par une formulation duale du problème, généralement plus

simple à traiter. À partir de cette formulation, les coefficients d'optimisation (multiplicateurs de Lagrange) sont obtenus par minimisation d'une fonction objective quadratique. Cette phase d'optimisation est réalisée initialement dans la phase d'apprentissage puis elle peut être poursuivie en ligne.

Dans le cas non linéaire, les SVM réalisent une transformation de l'espace de représentation initial en un espace de Hilbert comprenant plus de dimensions. Il est alors nécessaire d'utiliser d'autres types de noyaux (gaussien, polynomial, etc.) pour obtenir une frontière de décision non linéaire par rapport à l'espace initial, mais linéaire par rapport au nouvel espace. Cette mise en œuvre permet aux SVM d'être plus résistants que certaines méthodes de classification aux problèmes de dimensionnalité et de bruit. Il peut être reproché à la méthode SVM de reposer uniquement sur des vecteurs de support (données existantes) pour chercher à séparer les classes. De même, la complexité des méthodes SVM est importante et le choix initial de leurs paramètres (dont le noyau) est difficile et change complètement les résultats.

3.8.3 Les arbres de décision

La construction des arbres de décision à partir de données est une discipline déjà ancienne. Les statisticiens en attribuent la paternité à Morgan et al. [MOR 63] qui sont les premiers à utiliser les arbres de régression dans un processus de prédiction et d'explication. Il s'en est suivi toute une famille de méthodes, étendues jusqu'aux problèmes de discrimination et classement, qui s'appuyaient sur le même paradigme de la représentation par arbres [MOR 73]. On considère généralement que cette approche a connu son apogée avec la méthode CART (Classification and Regression Tree) de Breiman et al. [BRE 84].

Les arbres de décisions sont des algorithmes de classification qui prennent une décision en fonction du chemin parcouru par une nouvelle forme dans un arbre. Un arbre est composé d'une racine, de nœuds, de branches et de feuilles. Les branches sont les connexions reliant la racine, les nœuds et les feuilles. Chaque nœud d'un arbre de décision représente un test réalisé par rapport à un attribut. Chaque branche parcourue à partir d'un nœud correspond à une sortie possible du test réalisé. Chaque feuille contient l'estimation d'une classe. Pour classifier une nouvelle forme, la forme doit parcourir un chemin passant par la racine de l'arbre de décision pour aller jusqu'à une feuille. En fonction de l'arbre de décision développé, les tests contenus dans les nœuds et les feuilles peuvent être des règles de classification plus ou moins complexes. Lors de la phase de création d'un arbre, les feuilles sont progressivement remplacées par des nœuds. L'attribut qui est testé à un nœud est choisi par comparaison de tous les attributs disponibles en fonction de certaines mesures heuristiques.

Les arbres de décision ont connu un très grand succès dans les années 90. Malgré

un grand nombre de publications visant à en améliorer les performances, aucune avancée décisive n'a été produite en matière de taux de reconnaissance par rapport aux algorithmes de référence tels que :ID3, CHAID, CART et C4.5 [RAK 97 ; LIM 00].

L'intérêt apporté par toutes ces contributions a permis de mieux maîtriser les propriétés des arbres. Il est possible de caractériser les variantes et le cadre dans lequel elles fonctionnent le mieux. Si elles se révèlent bien souvent performantes sur les données artificielles construites à partir de fonctions logiques, elles sont peu décisives sur des bases réelles ; elles permettent surtout d'obtenir des classifieurs de taille réduite sans dégrader les performances [BRE 97].

La première catégorie de variantes vise à améliorer l'algorithme de recherche dans l'espace des solutions, soit en améliorant la méthode d'élagage, soit en procédant à une optimisation globale plus puissante, soit en procédant à des recherches en avant lors de la segmentation [RAG 93]. Le résultat de ces techniques est un arbre plus concis au prix d'un temps de calcul plus élevé [MUR 94].

La seconde catégorie de variantes cherche à modifier itérativement l'espace de recherche en produisant au fur et à mesure de nouveaux descripteurs. Connue sous le terme d'induction constructive, l'objectif est de trouver un espace de représentation plus approprié en élaborant des combinaisons de variables [BAG 90].

D'autres variantes ont voulu changer carrément le concept lui-même. Deux types de représentation sont généralement rencontrés : les arbres obliques qui utilisent une combinaison linéaire des variables lors de la segmentation des sommets de l'arbre, cette variante permet de lever la contrainte « parallèle aux axes » lors du partitionnement, dans l'espace de représentation. Généralement, l'arbre produit est plus concis ; en revanche la lecture des règles de décision est un peu plus compliquée [CAN 03]. Les graphes d'induction introduisent un nouvel opérateur « fusion » dans l'algorithme d'apprentissage ; le modèle de prédiction n'est donc plus un arbre, mais un graphe ; l'objectif est de permettre le regroupement d'individus de mêmes caractéristiques et d'assurer ainsi une meilleure résistance à la fragmentation des données [RAK 97]. En réalité, c'est plutôt du côté de la mise en œuvre des arbres de décision dans un cadre plus large qu'il faut trouver les principales innovations de ces dernières années. On peut citer par exemple leur utilisation dans les méthodes d'agrégation de classifieurs [FRE 97]. Autre innovation intéressante de ces dernières années, l'extension des algorithmes d'induction aux données non tabulaires avec les données floues et symboliques [OLA 03].

3.8.4 K plus proches voisins

Un des inconvénients majeurs de la méthode du Plus Proche Voisin est que celle-ci présente une sensibilité élevée aux abords des frontières entre classes. Le plus proche voisin d'un objet peut être d'une classe incorrecte, alors que la majorité de ses voisins ne

le sont pas. Afin de contrer cet effet, la classe assignée à un objet peut être celle qui est la plus représentée parmi les k plus proches prototypes trouvés. La méthode porte dans ce cas le nom de « k Plus Proches Voisins ». La fonction discriminante d'une classe est alors simplement le nombre de prototypes de cette classe qui se situent parmi les k plus proches voisins de l'objet à classer.

3.8.5 K plus proches voisins flous

Le K-ppv est l'une des plus importantes méthodes parmi les algorithmes non paramétriques, cependant cet algorithme présente quelques inconvénients. L'un de ces inconvénients est que tous les voisins ont le même poids pour l'affectation d'une classe à un vecteur de données. L'autre inconvénient est qu'il ne spécifie pas le degré d'appartenance du vecteur à la classe qui lui a été attribué. Keller et al. [KEL 85] ont proposé le classifieur K-ppv flou pour pallier aux limitations du K-ppv. L'algorithme K-ppv flou alloue au vecteur de données un degré d'appartenance à une classe donnée. Le principe est d'allouer le degré d'appartenance à une classe en fonction de la distance du vecteur de ces K-ppv et de l'appartenance de ces voisins à la classe.

3.8.6 Les modèles de Markov cachés

Les chaînes de Markov forment depuis plusieurs dizaines d'années un sujet de choix aussi bien au niveau des investigations mathématiques [MEY 12] que des applications à base de chaînes de Markov cachées en reconnaissance des formes [COR 11]. Au niveau théorique on considère souvent des modèles à temps discret ou continu avec des espaces d'états finis ou infinis aussi bien discrets que continus.

Les modèles de Markov cachés est une approche à deux processus stochastiques dont l'un est caché. La différence fondamentale avec les modèles classiques est que l'on n'observe pas directement les états, mais seulement les symboles produits par ces états. De plus, on code l'évolution temporelle dans la séquence d'états cachés, dont l'espace des valeurs possibles est généralement de cardinal beaucoup plus faible que celui des observations.

3.8.7 AdaBoost

AdaBoost est un méta-algorithme qui peut être utilisé conjointement avec de nombreux autres algorithmes d'apprentissage afin d'améliorer leurs performances. Le boosting est une méthode itérative qui trouve une classification très précise, en combinant un grand nombre de classificateurs de base [FRE 97].

Dans la phase d'entraînement de l'algorithme AdaBoost, une première distribution de poids est construite sur l'ensemble d'entraînement. Le mécanisme de boosting sélectionne

ensuite le classificateur de base qui donne le minimum d'erreurs. Ainsi, l'algorithme favorise le choix d'un autre classificateur qui est plus performant sur les données restantes au cours de la prochaine itération.

3.9 Conclusion

Ce chapitre est une brève introduction à la reconnaissance des formes. Les motivations qui conduisent à faire effectuer des tâches de reconnaissance de l'homme seront toujours un moteur pour le progrès dans ce domaine. Toutefois ceci doit se faire non pas pour remplacer l'homme par une machine, mais pour soulager des tâches ennuyeuses et dangereuses. L'aspect multidisciplinaire de la reconnaissance des formes est indiqué. Malgré l'absence d'une théorie unifiée de la reconnaissance des formes, on peut dégager deux approches presque complémentaires qui sont les approches statistiques et structurelles.

4

Suivi des objets en mouvement

4.1 Introduction

Le suivi des objets est une tâche importante dans le domaine de la vision par ordinateur. L'augmentation du nombre d'ordinateurs de grande puissance, la disponibilité des caméras vidéo de haute qualité, le faible coût et la nécessité croissante d'une analyse automatique des vidéos ont généré un grand intérêt pour les algorithmes de suivi d'objet. Il y a trois étapes clés dans l'analyse vidéo : la détection d'objets en mouvement, le suivi de ces objets d'un frame à un autre, et l'analyse des objets suivis afin de reconnaître leurs comportements. Par conséquent, l'utilisation du suivi d'objet est pertinente dans les tâches suivantes :

- La reconnaissance basée mouvement,
- La surveillance automatique,
- Indexation des vidéos,
- Interaction homme-machine,
- La surveillance du trafic routier,
- Navigation des véhicules.

Dans sa forme la plus simple, le suivi peut être défini comme le problème de l'estimation de la trajectoire d'un objet dans l'image lorsqu'il se déplace dans une scène. En d'autres termes, un traqueur donne des étiquettes conformes aux objets suivis dans les différents frames d'une vidéo. De plus, selon le domaine du suivi, un tracker peut également fournir des informations sur l'objet centré, telles que l'orientation, la région, ou la forme d'un

objet. Le suivi d'objets peut être très complexe en raison de :

- La perte d'informations causée par la projection du monde 3D sur une image 2D,
- Le bruit dans les images,
- Les mouvements complexes de l'objet,
- La nature non rigide ou articulés des objets
- Les occlusions partiales ou complètes des objets,
- La forme complexe des objets,
- Les changements d'illumination de la scène,
- Et les exigences du traitement en temps réel.

On peut simplifier le suivi en imposant des contraintes sur le mouvement et/ou l'apparence des objets. Par exemple, presque tous les algorithmes de suivi supposent que le mouvement de l'objet est lisse, sans changements brusques. On peut en outre limiter le mouvement de l'objet à être de vitesse constante ou une accélération constante sur la base de l'information a priori. La connaissance a priori du nombre et de la taille des objets, ou l'apparence de l'objet et sa forme, peut aussi être utilisée pour simplifier le problème. De nombreuses approches pour le suivi d'objets ont été proposées. Celles-ci diffèrent principalement les uns des autres en fonction de la façon dont ils abordent les questions suivantes :

- Quelle est la représentation de l'objet la plus adaptée pour le suivi ?
- Quelles sont les meilleures caractéristiques d'image qui doivent être utilisées ?
- Comment modéliser le mouvement, l'apparence et la forme de l'objet ?

Les réponses à ces questions dépendent fortement du contexte/environnement dans lequel le suivi est effectué et l'utilisation finale pour laquelle l'information de repérage est demandée. Un grand nombre de méthodes de suivi ont été proposées pour tenter de répondre à ces questions par des scénarios variés.

4.2 Représentation de l'objet

Dans un scénario de suivi, un objet peut être défini comme tout ce qui est d'intérêt pour de plus amples analyses. Les objets peuvent être représentés par leurs formes et par leurs apparences. Dans cette section, nous allons d'abord décrire les formes les plus couramment utilisées pour la représentation de l'objet dans le suivi ensuite, décrire le traitement par représentations de l'apparence.

4.2.1 Les Points

L'objet est représenté par un point comme le centre de gravité ou par un ensemble de points [SER 04]. En général, la représentation par un point est adaptée pour le suivi des objets qui occupent de petites régions dans l'image.

4.2.2 La forme géométrique

La forme de l'objet est représentée par un rectangle, par une ellipse ou par une autre forme géométrique. Le mouvement de l'objet pour ces représentations est généralement modélisé par les transformations affines, projectives ou translations [COM 03]. Bien que des formes géométriques primitives soient plus appropriées pour la représentation des objets rigides simples, ils sont également utilisés pour le suivi d'objets non rigides.

4.2.3 La silhouette et le contour des objets

La représentation par le contour définit la limite d'un objet. La région qui se trouve à l'intérieur du contour est appelée la silhouette de l'objet. La représentation basée sur la silhouette et le contour est appropriée pour le suivi des formes non rigides et complexes [YIL 04].

4.2.4 Les modèles de forme articulés

Les objets articulés sont constitués de parties du corps qui sont tenues par des joints [STR 12]. Par exemple, le corps humain est un objet articulé avec torse, les jambes, les mains, la tête et les pieds reliés par des articulations. La relation entre les parties est régie par des modèles cinématiques de mouvement, par exemple, l'angle de l'articulation, etc. Afin de représenter un objet articulé, on peut modéliser les éléments constitutifs en utilisant des cylindres ou des ellipses.

4.2.5 Les modèles basés sur le squelette

Le squelette d'un objet peut être extrait en appliquant des transformations jusqu'à l'obtention d'un axe médian. Ce modèle est couramment utilisé comme une représentation de la forme pour la reconnaissance d'objets [ALI 01]. La représentation basée squelette peut être utilisée pour modéliser à la fois des objets articulés et rigides.

4.3 Les caractéristiques d'apparence des objets

Il existe un certain nombre de moyens pour représenter les caractéristiques d'apparence des objets. On remarque que les représentations de forme peuvent également être combinées avec les représentations d'apparence pour le suivi.

4.3.1 Densités de probabilité de l'apparence de l'objet

Les estimations de la densité de probabilité de l'apparence de l'objet peuvent être soit paramétrique, comme une gaussienne ou un mélange de gaussiennes [PAR 02], ou

non paramétrique, tel que les histogrammes [COM 03]. Les densités de probabilité des caractéristiques d'apparence de l'objet (couleur, texture) peuvent être calculées dans les régions d'image spécifiées par les modèles de forme (région intérieure d'une ellipse ou un contour).

4.3.2 Les modèles d'apparence

Les modèles sont formés en utilisant des formes ou des silhouettes géométriques simples [FIE 97]. Un avantage d'un modèle est qu'il porte à la fois l'information spatiale et l'apparence. Les modèles ne codent que l'apparition de l'objet généré à partir d'une seule vue. Ainsi, elles ne conviennent que pour le suivi des objets dont la pose ne varie pas sensiblement au cours du suivi.

4.3.3 Les modèles d'apparence actifs

Les modèles d'apparence actifs sont générés par la modélisation simultanée de la forme et de l'apparence de l'objet. En général, la forme de l'objet est définie par un ensemble de points de repère [STE 00]. Similaire à la représentation basée contour, les repères peuvent résider sur la frontière de l'objet ou, par défaut, ils peuvent résider à l'intérieur de la zone de l'objet. Pour chaque point de repère, un vecteur d'aspect est stocké qui est sous la forme de couleur, de texture, ou de l'intensité de gradient. Les modèles d'apparence actifs nécessitent une phase d'apprentissage où la forme et son apparence sont apprises à partir d'un ensemble d'échantillons à l'aide de l'analyse en composante principale.

4.3.4 Les modèles d'apparence à multiples vues

Ces modèles codent différentes vues d'un objet. L'une des approches les plus utilisées pour représenter les multiples vues de l'objet consiste à générer un sous-espace de vues attribuées [LEL 03]. L'apprentissage s'effectue soit par une analyse en composantes principales (ACP) ou une analyse en composantes indépendantes (ACI) qui sont utilisées à la fois pour la représentation par la forme et la représentation par l'apparence. Une autre approche pour l'apprentissage de vues d'un objet consiste à utiliser un ensemble de classificateurs tels que les machines à vecteurs de support ou les réseaux bayésiens. Une des limites de ce modèle est que les apparitions dans tous les points de vue sont nécessaires à l'avance.

En général, il existe une très forte relation entre la représentation de l'objet et l'algorithme de suivi. Les représentations de l'objet sont généralement choisies en fonction de l'application. Pour le suivi d'objets, qui apparaissent très petits dans une image, la représentation par point est généralement plus appropriée. Une représentation basée point

a été utilisée par Veenman et al. [VEE 01] pour le suivi de grains dans une séquence de plat. Une représentation similaire a été proposée par Shafique et al. [SHA 05] pour le suivi d'un troupeau d'oiseaux à distance. Les formes géométriques sont plus appropriées lorsqu'on veut représenter des objets avec des formes bien définies ou avec un rapprochement vers une autre forme géométrique simple telles que les rectangles ou les ellipses. Comaniciu et al. [COM 03] ont utilisé une représentation de forme elliptique et emploient un histogramme de couleur calculée à partir de la région elliptique pour la modélisation de l'apparence. Pour représenter l'apparence de l'objet, Black et al. [BLA 98] ont utilisé les vecteurs propres générés à partir de modèles d'objets rectangulaires. Pour le suivi d'objets de formes plus complexes telles que les humains, un contour ou une représentation basée silhouette est mieux approprié [HAR 00].

4.4 Sélection des caractéristiques pour le suivi

Choisir les bonnes caractéristiques joue un rôle essentiel dans le suivi. En général, la propriété la plus souhaitable d'un élément visuel est son caractère unique, de sorte que les objets peuvent être facilement distingués dans l'espace des caractéristiques. La sélection des fonctionnalités est étroitement liée à la représentation de l'objet. Par exemple, la couleur est utilisée en tant que caractéristique pour la représentation de l'apparence basée sur un histogramme, tandis que pour la représentation basée contour, les bords de l'objet sont généralement utilisés en tant que caractéristiques. En général, de nombreux algorithmes de suivi utilisent une combinaison de ces caractéristiques.

4.4.1 Les couleurs

La couleur apparente d'un objet est influencée principalement par deux facteurs physiques à savoir la distribution d'énergie spectrale de la source et les propriétés de réflectance de la surface de l'objet [PAS 01]. Dans le traitement de l'image, l'espace de couleur RGB est généralement utilisé pour représenter la couleur. Cependant, l'espace RGB n'est pas une perception uniforme de l'espace de couleurs, c'est-à-dire les différences qui existent entre les couleurs dans l'espace RGB ne correspondent pas aux différences de couleurs perçues par l'homme. De plus, les dimensions RGB sont fortement corrélées. En revanche, les espaces de couleur tels que HSV, HSL sont parfaitement uniforme à la perception humaine. Toutefois, ces espaces de couleur sont sensibles au bruit. En résumé, il n'y a pas de dernier mot sur lequel l'espace de couleur est plus efficace, par conséquent, une série d'espaces de couleurs ont été utilisés dans le suivi.

4.4.2 Les contours

Les Limites de l'objet génèrent habituellement de fortes variations d'intensité de l'image. La détection de bord est utilisée pour identifier ces changements. Une propriété importante est que les bords sont moins sensibles aux variations d'éclairage par rapport aux caractéristiques de couleur. Les algorithmes qui permettent le suivi des contours d'un objet utilisent, dans la plupart des cas, les bords des objets comme caractéristiques. En raison de sa simplicité et de sa précision, l'approche la plus populaire de détection de contour est le filtre de Canny. Une évaluation des algorithmes de détection de bord est fournie par Bowyer et al. [BOW 99].

4.4.3 Le flux optique

Le flux optique est un champ dense de vecteurs de déplacement qui définit la transposition de chaque pixel dans une région. Il est calculé en utilisant la contrainte de la luminosité, qui suppose que la luminosité des pixels, correspondant à des trames consécutives, est constante. Le flux optique est couramment utilisé comme un élément de segmentation et de suivi des applications. Les techniques populaires pour calculer la densité du flux optique comprennent les méthodes proposées par : Black et al.[BLA 96], Szeliski et al. [SZE 97],Horn et al. [HOR 81], Lucas et al. [LUC 81].

4.4.4 Les textures

La texture est une mesure de la variation d'intensité d'une surface qui quantifie les propriétés telles que la douceur et la régularité. Par rapport à la couleur, la texture nécessite une étape de traitement pour générer les descripteurs. Il existe différents descripteurs de texture, les plus utilisés sont :

- Les matrices de cooccurrence qui utilisent un histogramme d'intensité dans une direction bien spécifique comme descripteur de texture [HAR 73].
- Les mesures de Laws [LAW 80] qui se base sur la génération de 21 filtres de deuxième dimension sur la base de cinq filtres de première dimension .
- Les transformations en ondelette [MAL 89] et les pyramides orientées [GRE 94].

Les méthodes de modélisation de texture peuvent être classées dans quatre catégories : les modèles statistiques, les modèles structurels et les modèles fondés sur des filtres. Les modèles statistiques mesurent la distribution spatiale des valeurs des pixels [BOU 99] et les matrices de cooccurrence [HAR 73]. Dans les méthodes structurelles, la texture est représentée par une répétition d'éléments structurels et la texture est modélisée comme étant un arrangement spatial de ces éléments [VIL 86]. Quant aux modèles fondés sur les filtres, ils consistent à appliquer un ensemble de filtres à l'image afin d'en étudier la réponse. On retrouve les filtres dans le domaine spatial (filtres de Sobel, de Canny,

Robert, etc.), ceux dans le domaine fréquentiel (filtres de Fourier) ou dans le domaine spatio-temporel (filtres de Gabor, transformée en ondelettes, etc.). Les textures sont moins sensibles au changement d'intensité lumineuse par rapport aux couleurs.

Dans la plupart du temps, les caractéristiques sont choisies manuellement par l'utilisateur en fonction de l'application et le domaine. Cependant, le problème de la sélection automatique des caractéristiques a reçu une attention significative dans la communauté de la reconnaissance des formes. Les méthodes de sélection automatique des caractéristiques peuvent être divisées en méthodes de filtrage et les méthodes d'encapsulation. Les méthodes par filtrage essaient de sélectionner les caractéristiques en se basant sur des critères généraux (par exemple les caractéristiques doivent être non corrélées). Tandis que les méthodes d'encapsulation sélectionnent les caractéristiques sur la base de l'utilité de ces derniers dans un problème particulier (par exemple, les performances de classification en utilisant un sous-ensemble de fonctions).

4.5 Le suivi des objets en mouvement

L'objectif principal du suivi d'un objet en mouvement est de tracer sa trajectoire dans un laps de temps bien défini en repérant à chaque instant sa position dans l'image. Donc, le suivi fournit les régions occupées par un objet à chaque instant. La détection de l'objet consiste à établir une correspondance entre deux instances de l'objet se retrouvant chacun dans des frames séparés. Cette dernière peut être réalisée séparément ou conjointement. Le premier cas consiste à localiser les régions possibles de l'objet dans chaque image par le biais d'un algorithme de détection d'objet, puis d'appliquer un suivi par une correspondance entre les objets dans toutes les images. Il s'agit dans le deuxième cas d'effectuer la détection et la correspondance de l'objet simultanément par une mise à jour itérative de l'emplacement de l'objet dans la région obtenue à partir de trames précédentes.

4.5.1 Le suivi de Point

Le suivi peut être formulé comme la correspondance des objets détectés représentés par des points à travers les images [YIL 06]. Les objets détectés dans les frames successifs sont représentés point par point, et l'association entre les points est basée sur l'état précédent de l'objet incluant la position et le mouvement de l'objet. Cette approche nécessite une méthode de détection des objets dans chaque frame de la vidéo. Il est très difficile et parfois impossible d'effectuer une correspondance entre les points surtout dans la présence des occlusions, les mauvaises détections, la présence d'un nouvel objet ou la sortie de ce dernier de la scène [VEE 01]. Donc le plus difficile dans cette catégorie n'est pas la représentation elle-même de l'objet, mais l'utilisation d'une bonne méthode de correspondance entre les points. On trouve dans la littérature deux principales catégories à

savoir, les méthodes déterministes et les méthodes statistiques. Les méthodes déterministes utilisent des heuristiques de mouvement qualitatives pour résoudre le problème de correspondance. Les méthodes probabilistes prennent explicitement en compte la mesure de l'objet et les incertitudes afin d'établir la correspondance.

4.5.1.1 Les méthodes déterministes

Dans cette catégorie, les méthodes déterministes essayent d'associer chaque objet, dans les précédents frames, à un seul objet du frame courant. Cette association est guidée par un coût calculé en utilisant un ensemble de contraintes de mouvement [YIL 06]. La minimisation de ce coût est formulée en tant que problème d'optimisation combinatoire. Le coût de correspondance est habituellement défini en utilisant une combinaison des contraintes suivantes.

- **La proximité** : cette contrainte suppose que l'emplacement ne change pas d'un frame à un autre.
- **Vitesse maximale** : c'est une contrainte qui limite la vitesse maximale d'un objet et par conséquent limite les correspondances possibles au voisinage circulaire du même objet.
- **Changement de vitesse** : cette contrainte suppose que l'objet ne peut pas changer subitement et de façon radicale sa direction et sa vitesse.
- **Mouvement commun** : la contrainte de mouvement suppose que l'objet garde la même vitesse dans son voisinage et par conséquent la vitesse est également limitée dans le voisinage de l'objet. Cette contrainte est appropriée pour les objets représentés par des points multiples.
- **La rigidité de l'objet** : Suppose que les objets dans le monde 3D soient rigides, par conséquent, la distance entre deux points quelconques de l'objet réel restera inchangée.

Plusieurs travaux ont été proposés dans cette catégorie. Dans les travaux de Sethi et al. [SET 87], les auteurs ont proposé de résoudre la correspondance par une approche gloutonne basée sur la proximité et la rigidité des contraintes. Une version améliorée a été proposée par Salari et al. [SAL 90] en effectuant une correspondance pour les points détectés puis, d'étendre le suivi des objets manquants par l'ajout d'un nombre de points aléatoires. Intille et al. [INT 97] ont proposé une correspondance entre le centre de gravité de l'objet. Les objets sont détectés à l'aide de la soustraction du fond. Les auteurs traitent explicitement la variation du nombre d'objets en examinant les régions spécifiques dans l'image pour détecter d'éventuelles entrées / sorties avant de calculer la correspondance.

Afin de préserver la cohérence temporelle de la vitesse et de la position, Shafique et al. [SHA 05] propose une approche multiframe. Le problème est présenté comme une correspondance multiple entre plusieurs frames et la solution consiste à utiliser la théorie des graphes pour trouver le meilleur chemin pour chaque point.

4.5.1.2 Les méthodes statistiques

Les méthodes statistiques permettent de résoudre les problèmes de suivi liés aux bruits générés par les capteurs vidéo ou par les perturbations aléatoires qui affectent les mouvements de l'objet [YIL 06]. Cela est effectué en prenant la mesure et les incertitudes des modèles lors de l'estimation de l'état de l'objet. Les méthodes statistiques utilisent une approche spatio-étatique pour modéliser les propriétés de l'objet comme la position, la vitesse et l'accélération [VAS 03]. Les mesures se calculent par rapport à la position de l'objet dans l'image obtenue par un mécanisme de détection.

Le filtre de Kalman C'est un filtre qui est utilisé pour estimer l'état d'un système linéaire dans lequel l'état est supposé suivre une distribution gaussienne. Le filtre est composé de deux étapes à savoir la prédiction et la correction [ROS 99]. La prédiction utilise le modèle de l'état de l'objet pour prédire le nouvel état selon les équations suivantes :

$$\bar{X}_t = DX_{t-1} + W \quad (4.1)$$

$$\bar{\Sigma}_t = D\Sigma_{t-1}D^T + Q_t \quad (4.2)$$

Où \bar{X}_t et $\bar{\Sigma}_t$ exprime l'état et la covariance de la prédiction dans le temps t . D est une matrice de transition d'état qui définit la relation entre la variable d'état à l'instant t et $t - 1$. W est le bruit et Q est la covariance du bruit W . De façon analogue, l'étape de correction utilise l'observation actuelle pour la mise à jour de l'état de l'objet selon les équations 4.3, 4.4 et 4.5

$$K_t = \bar{\Sigma}_t M^T [M^T \bar{\Sigma}_t M^T + R_t]^{-1} \quad (4.3)$$

$$X_t = \bar{X}_t + K_t \underbrace{[Z_t - M\bar{X}_t]}_V \quad (4.4)$$

$$\Sigma_t = \bar{\Sigma}_t - K_t M \bar{\Sigma}_t \quad (4.5)$$

Où Z_t est l'observation actuelle, M est la matrice de mesure, K est le gain de Kalman et V exprime l'innovation. Lorsque le système n'est pas linéaire, la prédiction est effectuée avec une extension du filtre de Kalman. Ce dernier utilise un développement en séries de Taylor pour obtenir un système linéaire. Malgré son efficacité et sa large utilisation dans le domaine de la vision, le filtre de Kalman présente un inconvénient majeur lié aux conditions d'utilisation. En effet, le filtre de Kalman suppose que la variable d'état suit une loi normale ce qui n'est pas toujours le cas dans le suivi des objets. Cette limitation peut être contournée en utilisant le filtre à particules.

Le Filtre à particules Dans le filtre à particules, la densité conditionnelle de l'état de l'objet $P(X_t/Z_t)$ à l'instant t est représentée par un ensemble d'échantillons de N par-

ticules $\{s_i^{(n)} : n = 1, \dots, N\}$ pondérée chacun par un poids $\pi_i^{(n)}$. Le poids définit l'importance de l'échantillon exprimée par la fréquence d'apparition de ce dernier. Afin de minimiser le temps de traitement, un poids cumulatif $c_{(n)}$ est également sauvegardé pour chaque couple $(s^{(n)}, \pi^{(n)})$ [MAC 98]. Les nouveaux échantillons à l'instant t sont tirés de l'étape précédente à l'instant $t - 1$ sur la base des différents schémas d'échantillonnage $S_{t-1} = \{s_{t-1}^{(n)}, \pi_{t-1}^{(n)}, c_{t-1}^{(n)} : n = 1, \dots, N\}$. Le filtre à particules peut être initialisé par les premières mesures $s_0^{(n)}$, avec un poids $\pi_0^{(n)} = \frac{1}{N}$ ou par un apprentissage sur l'ensemble d'échantillons. Une étape supplémentaire est effectuée afin de garder les meilleures particules de suivi. En effet, les échantillons représentés par des poids très faibles seront supprimés de l'ensemble. Dans cette dernière décennie, le filtre à particules est devenu incontournable dans le domaine de la vision par machine [YIL 06].

Filtre de probabilité de données conjointe Lors du suivi de plusieurs objets à l'aide du filtre de Kalman ou du filtre à particules, il faut associer de façon déterministe la mesure la plus probable pour un objet particulier à l'état de l'objet. Ce problème de correspondance doit être résolu avant que ces filtres puissent être appliqués. La méthode la plus simple pour effectuer la correspondance est d'utiliser l'algorithme du plus proche voisin (KNN) [RAS 01]. Toutefois, si les objets sont proches les uns des autres, alors il y a toujours un risque que la correspondance soit incorrecte induisant ainsi à un échec de convergence du filtre.

Il existe plusieurs mécanismes d'association de données statistiques pour lutter contre ce problème. Le filtre de probabilité de données conjointe est une alternative pour associer les données [RAS 01]. Soit une plage de valeurs définie comme une séquence de mesures qui sont supposées provenir du même objet. On suppose que nous avons N plages de valeurs à l'instant t avec leurs mesures $Z(t) = z_1(t), \dots, z_m(t)$ et nous avons besoin de relier ces mesures avec les plages existantes. Soit η un ensemble d'assignation de mesures. On suppose également que le nombre de plages de valeurs est constant dans le temps. Soit $v_{i,l}$ l'innovation, au sens de Kalman, associé à la plage l de la mesure Z_i . Le filtre de probabilité conjointe associe toutes les mesures avec chaque plage de valeurs selon l'équation 4.6

$$v^l = \sum_{i=1}^{m_k} \beta_i^l v_{i,l} \quad (4.6)$$

Où β_i^l est la probabilité a posteriori que la mesure i soit originaire de l'objet associé à la plage l

$$\beta_i^l = \sum_{\eta} P[\eta_l(k) | Z^l] \tau_{i,l}(\eta)$$

$\tau_{i,l}$ est une variable binaire qui permet d'indiquer l'association de la mesure $z_i(k)$ avec la plage de valeurs l .

Suivi par des hypothèses multiples Dans les méthodes précédentes, l'association s'effectue entre deux frames qui se suivent. Il est fort possible, dans un tel cas, d'effectuer une correspondance incorrecte entre les occurrences du même objet. Le suivi par des hypothèses multiples permet de retarder la décision de la correspondance pour un meilleur rendement [HUE 02]. Cet algorithme permet de sauvegarder les correspondances pour chaque objet dans chaque frame sous forme d'hypothèses de correspondances, et cela dans une durée bien déterminée en nombre de frames. Au final, l'algorithme retient uniquement les correspondances les plus probables pour une plage de valeurs. L'algorithme commence par la création d'un ensemble d'hypothèses contenant des prédictions de la position de chaque objet. Les prédictions seront comparées avec les mesures actuelles et un ensemble de correspondances est établi pour chaque hypothèse. Chaque nouvelle hypothèse va représenter une nouvelle plage de valeurs en se basant sur les mesures actuelles. Cette méthode dresse la liste de toutes les correspondances possibles ce qui augmente considérablement le temps de calcul. Plusieurs propositions ont essayé de réduire ce temps soit par considérer que les associations sont indépendantes et de cette façon éviter de calculer toutes les possibilités [STR 94], soit par la sélection des meilleures hypothèses [COX 96] ou par l'utilisation d'une plateforme d'hypothèses de suivi multiple [CHA 99].

4.5.2 Le suivi à base de noyau

Dans cette catégorie, le suivi consiste à calculer le mouvement d'une forme ou d'une apparence de l'objet dans les frames successifs. Dans ce cas, le suivi est un algorithme qui permet d'estimer le mouvement par des transformations géométriques telles que la translation, la rotation et les transformations affines. Le suivi basé noyau est effectué en calculant les mouvements de l'objet d'un frame à un autre. Les mouvements de l'objet sont généralement de type paramétrique ou des champs de flux dense calculés dans les images ultérieures. Ces algorithmes se diffèrent souvent dans l'apparence utilisée pour la représentation de l'objet, le nombre d'objets suivis et la méthode utilisée pour estimer le mouvement de l'objet. En général, on peut soit comparer directement les noyaux du modèle (m) et de l'objet cible (o), soit comparer leur distribution [COM 02a]. Pour ce faire, on peut définir des mesures directes selon les équations 4.7, 4.8 et 4.9

$$D_{L1}(m, O) = \sum_{i \in \text{noyau}} |m(i) - O(i)| \quad (4.7)$$

$$D_{LSQ}(m, O) = \sum_{i \in \text{noyau}} (m(i) - O(i))^2 \quad (4.8)$$

$$D_{corr}(m, O) = \sum_{i \in \text{noyau}} m(i) \cdot O(i) \quad (4.9)$$

Ou bien utiliser des mesures basées sur les distributions de probabilités selon les équations suivantes :

$$D_{EMD}(P, Q) = \text{EarthMoverDistance}(P, Q) \quad (4.10)$$

$$D_{inter}(P, Q) = \text{Intersection}(P, Q) \quad (4.11)$$

$$D_{Hellinger}(P, Q) = \sqrt{1 - \text{Bhattacharyya}(P, Q)} \quad (4.12)$$

Ou P et Q représentent respectivement les distributions de probabilité du modèle m et de l'objet O . L'algorithme de suivi de noyau le plus utilisé est le mean shift [COM 02b], qui utilise une fonction de densité des histogrammes de couleur pour appairer le modèle et la cible. Plutôt que de prendre les données directement pour former l'histogramme de l'objet P et les cibles $P(y)$, on effectue une pondération à l'aide d'une fonction de densité. Donc, le mean-shift consiste à trouver, de façon itérative, le déplacement d'un objet en mettant en correspondance leurs histogrammes pondérés par une fonction de densité [COM 02b].

4.5.2.1 Appariement de modèle

L'appariement de modèle est parmi les approches les plus communément utilisées pour le suivi [SCH 02b]. C'est une méthode qui permet d'effectuer la recherche d'une image I_t par le biais d'une région similaire à un modèle O_t défini dans les frames précédents. La position du modèle dans l'image est calculée en utilisant une mesure de similarité. Le modèle est bâti en utilisant les caractéristiques chromatiques de l'image telles que la couleur, les intensités des pixels et les gradients [SCH 02b]. Ces derniers sont très efficaces à cause de leurs faibles variations par rapport au changement de luminosité.

L'inconvénient majeur de cette méthode est le coût de calcul très élevé à cause de la vérification de la similarité entre un modèle et toutes les régions possibles dans l'image en court de traitement. En revanche les performances peuvent être améliorées en limitant la recherche dans les zones voisines susceptibles de contenir l'objet.

4.5.2.2 Les modèles d'apparence basés densités

En plus des modèles, l'apparence de l'objet peut être également utilisée pour le suivi. En effet, les histogrammes de couleur et les mixtures de modèle peuvent être facilement calculés en utilisant l'apparence de l'objet à l'intérieur d'un rectangle. Plusieurs méthodes ont été proposées dans ce type de suivi. L'utilisation de la moyenne des couleurs des pixels à l'intérieur d'un rectangle a été proposée par Fieguth et al. [FIE 97]. Un histogramme pondéré calculé dans une région circulaire a été proposé par Comaniciu et al. [COM 02a]. Jepson et al. [JEP 03] proposent un suivi par un modèle à trois mixtures contenant des caractéristiques d'apparence stable, des caractéristiques transitoires et un processus de bruit. Shi et al. [SHI 94] ont proposé une méthode de suivi qui permet de calculer de manière itérative les translations de région centrée autour d'un point d'intérêt. Tao et al.

[TAO 02] proposent une méthode basée sur la totalité de l'image comme un ensemble de couches. La représentation inclut une seule couche pour l'arrière-plan et une couche pour chaque objet. Chaque couche a comme forme a priori une ellipse, des mouvements (translation, rotation) et une apparence. Isard et al. [ISA 01] proposent une modélisation conjointe de l'arrière-plan et les régions de premier plan pour le suivi. L'apparence de l'arrière-plan et de tous les objets sont représenté par un mélange de gaussiennes et les formes des objets sont considérées comme des cylindres.

4.5.2.3 Les modèles d'apparence à multiples vues

Les modèles d'apparence cités précédemment sont générés en court de traitement. Cette approche modélise les objets en se basant sur les informations collectées dans les précédents frames. Cette démarche n'est pas pratique pour les objets qui ont des changements brusques et rapides. Effectivement, l'objet peut changer d'apparence d'une position à une autre et s'il y a un changement de vue radicale dans le suivi, l'objet peut ne pas être reconnu à cause du changement du modèle. C'est pour cette raison que plusieurs travaux ont proposé des modèles à vue multiples pour permettre de remédier à ces problèmes. Black et al. [BLA 98] proposent d'utiliser des sous-espaces pour calculer les transformations affines. Pour ce faire, des agents vont reconstruire l'image résultante en se basant uniquement sur les transformations affines calculées précédemment. Avidan [AVI 04] propose un système similaire basé sur les SVM. Le système utilise les scores générés par les SVM pour indiquer le degré d'appartenance à la classe positive qui contient les images des objets en mouvements. Tandis que la classe négative contient l'arrière-plan des autres régions.

4.5.2.4 Le suivi de silhouette

Les méthodes de suivi basées sur la silhouette peuvent être considérées comme une segmentation d'objet appliquée dans le domaine temporel en utilisant les informations générées à partir des images précédentes [KAN 03]. Ces méthodes utilisent l'information codée à l'intérieur de la région de l'objet. Cette information peut être sous la forme d'une densité de l'apparence ou d'un modèle de la forme de l'objet représenté sous la forme d'une carte de bord [COM 02a]. Compte tenu des modèles d'objet, la silhouette est suivie en effectuant une correspondance de forme ou en utilisant une évolution du contour de l'objet.

Les méthodes basées sur la silhouette de l'objet sont très efficaces lorsqu'il s'agit d'un objet avec une forme rigide [YIL 04]. Effectivement, les objets tels que le corps d'un humain ou d'un animal ne peuvent pas être bien définis par une forme ou un ensemble de formes géométriques. Pour cela, la silhouette constitue un bon moyen pour la description de ces objets. Dans ce type d'approche, il est question de retrouver les objets en

mouvement via un modèle généré en utilisant des informations collectées a priori dans les frames précédents. Parmi les caractéristiques les plus communément utilisées dans ce type d'approche on trouve : les histogrammes de couleur [KAN 04 ; COM 02a], les bords [HAR 00] et les contours des objets [SAT 04].

Appariement de forme Cette approche permet de chercher la silhouette de l'objet dans le frame courant en se basant sur des modèles générés. L'appariement est effectué en calculant le degré de similarité entre l'objet et le modèle généré. Le modèle est bâti sur un ensemble d'hypothèses, sur les silhouettes probables de l'objet, extraites à partir des frames précédents.

Dans cette approche, les mouvements non rigides des objets ne sont pas explicitement traités. En effet, la silhouette est supposée bougée d'un frame à un autre et c'est pour cette raison que les mouvements qui exercent une transformation au niveau de la silhouette ne peuvent être traités. Dans chaque nouveau frame et après la localisation de l'objet, une mise à jour est appliquée sur le modèle de l'objet afin qu'il puisse gérer les changements et les problèmes de suivi liés aux changements de conditions d'éclairage ainsi que le mouvement de l'objet non rigide.

Plusieurs travaux basés sur la silhouette ont été publiés pour traiter le problème du suivi. Huttenlocher et al. [HUT 93] ont proposé une approche pour l'appariement des formes en utilisant une représentation basée sur les bords de l'objet. Les auteurs ont utilisé la distance de Hausdorff pour construire une surface de corrélation à partir de laquelle le minimum est sélectionné en tant que nouvelle position de l'objet. Une approche similaire a été proposée par Li et al. [LI 01] en utilisant la distance de Hausdorff pour la vérification des trajectoires. Le suivi se résume dans l'évaluation du flot optique des vecteurs calculés par rapport au modèle de la silhouette généré de sorte que la moyenne du flot détermine la nouvelle position de l'objet. Une autre manière de calculer l'appariement des formes est de détecter les silhouettes correspondantes dans deux frames successifs. Cette approche est identique à l'appariement de point dans le processus, mais elle est totalement différente dans la représentation de l'objet.

Suivi basé contour Cette méthode utilise un processus itératif qui permet d'appliquer un ensemble de transformations au contour initial, extrait des frames précédents, pour arriver à la position actuelle de l'objet. Il est nécessaire qu'une partie de l'objet dans le frame courant chevauche avec une partie de l'objet dans le frame précédent. Dans la littérature, plusieurs techniques ont été proposées afin de faire évoluer les contours. Deux processus ont trouvé un très grand succès[YIL 06], le premier utilise un modèle spatiotemporel pour modéliser le contour et son mouvement tandis que le second essaye de minimiser l'énergie des contours par la descente de gradient.

Modèle spatiotemporel Chaque état de l'objet est caractérisé par la forme et le mouvement du contour. La mise à jour de l'état de l'objet est effectuée en maximisant la probabilité a posteriori qui dépend fortement de l'état précédent et du voisinage. Ce dernier est défini par la distance entre le contour et les bords de l'objet. L'état de l'objet peut être également défini comme une suite dynamique des points de contrôle [TER 93]. D'autres travaux ont tenté de définir l'état comme des paramètres sous forme de baguette et des paramètres de mouvement affine [ISA 98]. Chen et al. [CHE 01] ont utilisé des contours sous forme d'ellipse associée à un modèle de Markov caché.

Minimisation de l'énergie des contours L'énergie du contour est définie comme une information temporelle sous forme de gradient [BER 00 ; MAN 02 ; CRE 03], ou une apparence statistique générée à partir d'un objet [YIL 04] ou d'un arrière-plan [RON 94]. L'utilisation du gradient de l'image est largement utilisée dans le calcul du flot optique. Bertalmio et al. [BER 00] ont utilisé des contraintes pour appliquer des transformations sur le contour et cela de manière consécutive dans chaque frame en utilisant une représentation multiniveau sur la direction du flux. Les contraintes du flot optique ont été également utilisées par Mansouri et al. [MAN 02] pour le suivi en calculant uniquement les limites de l'objet. Cremers et al. [CRE 03] utilisent le flot optique pour calculer l'évolution du contour en posant des contraintes sur l'homogénéité des régions. Ronfard [RON 94] a proposé une autre alternative du flot. Il a exploité les statistiques calculées pour chaque objet pour définir l'évolution du contour de l'objet.

4.6 Conclusion

Nous avons présenté, dans ce chapitre, une étude approfondie des méthodes et techniques couramment utilisées pour le suivi des objets en mouvement. Nous avons également donné un bref aperçu des sujets connexes et les contraintes relatives aux domaines d'application. Le processus de suivi consiste à conserver l'identité des objets détectés et de suivre leurs évolutions au cours du temps. Le suivi d'objets est par conséquent dépendant de la représentation des objets et de l'incertitude sur les observations. Nous avons divisé les méthodes de suivi en trois principales catégories basées sur l'utilisation et la représentation d'objets, à savoir, les méthodes basées sur l'établissement d'une correspondance entre les points de la forme, les méthodes utilisant des modèles géométriques primitifs, et les méthodes basées contour. Toutes les méthodes présentées nécessitent une détection préalable de l'objet avant d'effectuer le suivi et cela que ce soit à chaque nouveau frame (méthodes basées sur les points) ou uniquement lors de la première apparition dans la scène (les approches géométriques et les approches basées sur le contour).

Deuxième partie

**Proposition d'un Système de
Vidéosurveillance**

5

Détection des formes mobiles dans une vidéosurveillance

5.1 Introduction

La reconnaissance des formes, le traitement des images ainsi que la vision par machine sont des domaines de recherche très vaste qui ont pour rôle principal d'automatiser l'analyse des images et de la vidéo. L'application de ces derniers vise une liste non exhaustive de domaines tels que l'imagerie médicale, l'interaction homme-machine, la compression vidéo, etc. La vidéosurveillance est un domaine en émergence qui a su profiter des avancements technologiques des autres domaines et qui est devenu un axe de recherche à part entière[FAR 12]. Le besoin de sécuriser les endroits publics tels que les aéroports, les banques, les stations de métro, etc. à également attirer l'attention des chercheurs dans la dernière décennie[FAR 12].

Le rôle des systèmes de vidéosurveillance est d'offrir une plateforme qui permet d'assister les opérateurs humains dans leur tâche de surveillance par le biais d'une alerte visuelle ou sonore. Cette alerte est généralement liée à un comportement ou une situation qualifiée d'anormale dans le champ de vision d'une caméra. Un tel système va permettre aux opérateurs de surveillance de visualiser simultanément plusieurs environnements en focalisant son attention sur les points critiques annoncés par le système et cela en temps réel[FAR 13b]. Le système peut être également utilisé ultérieurement comme un support de recherche grâce à la méthode d'indexation et de recherche vidéo pour d'éventuels pro-

blèmes tels que les crimes et les vols.

Les tentatives terroristes, dans les différents capitaux du monde tel que New York, Madrid et Londres, ont donné un grand coup de pouce dans le développement rapide de ces systèmes qui sont devenus une partie de notre vie quotidienne. Cependant les systèmes de vidéosurveillance ne sont pas une restriction sur les applications de sécurité, ils ont également contribué dans l'amélioration des qualités de service[FAR 12].

5.2 La vidéosurveillance

5.2.1 Définition

La vidéosurveillance est la surveillance à distance d'un lieu, public ou privé, de machines ou d'individus. Elle est employée dans de nombreuses situations, généralement pour des raisons de sécurité.

5.2.2 Présentation

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

5.2.3 Historique

La première génération des systèmes de vidéosurveillance est le CCTV (closed Circuit Television Network). Dans ce système les caméras vidéo analogiques sont connectées par des câbles coaxiaux à des moniteurs de surveillance pour être utilisées par les opérateurs humains ou la caméra sera connectée à un enregistreur de vidéo pour l'archivage[FAR 11]. Les premiers systèmes CCTV ont été développés dans la fin des années 1940 ou elles étaient utilisées par les militaires allemands pour observer les missiles. Les systèmes CCTV ont été également utilisés par les militaires américains pour tester les bombes atomiques[FAR 11].

La deuxième génération des vidéosurveillances a remplacé les enregistreurs vidéo avec les DVR avec enregistrement sur disque dur. Dans la troisième génération, les systèmes sont à adresse IP, ou les données sont périodiquement transférées via le réseau. Aujourd'hui les caméras de vidéosurveillance sont omniprésentes.

5.2.4 Objectifs majeurs

les systèmes de vidéo-surveillance ont plusieurs objectifs[FAR 11], nous citons :

- L'analyse de scène pour l'interprétation d'événements.
- Le déclenchement d'alertes.
- La maîtrise de l'espace public (propreté urbaine, la lutte contre les dépôts sauvages, éclairage public, etc.).
- La lutte contre la délinquance (sécurité des personnes et des biens, vols à la portière, agressions, flagrants délits, etc.).
- Le signalement des stationnements anarchiques engendrant des perturbations du flux routier.

5.2.5 Domaines d'application

La vidéosurveillance couvre plusieurs domaines d'applications. La lutte contre le terrorisme est l'un des plus connus. De ce point de vue, les attentats de Londres au cours de l'été 2005 ont prouvé l'efficacité de la vidéosurveillance, non pas tant sur le plan préventif, mais pour l'enquête criminelle, en permettant d'identifier des visages et de reconstituer le trajet des terroristes (grâce à l'analyse des images enregistrées sur leur parcours).

L'installation de caméras dans les transports en commun (tramway, trains, bus et métros) poursuit trois objectifs : la sécurité des voyageurs, la lutte contre la fraude et la protection contre les dégradations (non seulement dans les rames, mais aussi dans les gares, sur les quais et, depuis peu, le long de certaines voies). La surveillance des sites sensibles, publics et privés, liés à des intérêts vitaux et stratégiques sur un plan national (énergie, défense, transport), représente également l'un des domaines d'application les plus courants. Elle permet de contrôler l'accès à des zones réservées en associant identification biométrique (iris, empreintes digitales), reconnaissance faciale et vidéosurveillance. Plusieurs aéroports internationaux sont équipés de systèmes de ce type, destinés à fluidifier le passage aux frontières ou encore à automatiser l'accès du personnel aux différentes zones réservées (bagages, fret, pistes).

Enfin, le contrôle et la gestion de la circulation automobile, tant en ville que le long des axes routiers, constituent des domaines d'application prometteurs. Ce contrôle va de la lecture à distance de plaques d'immatriculation (avec envoi automatique d'amendes aux propriétaires des véhicules en situation irrégulière) à la surveillance de voies routières et de carrefours dangereux ou encombrés. Dans ce cas, la vidéosurveillance est couplée à un centre de gestion du trafic qui permet, par exemple, de dérouter la circulation ou encore de fluidifier le trafic en faisant évoluer la signalisation. Sans oublier, les installations privées importantes concernent les hôtels, les restaurants, les casinos, les salles de jeux qui font régulièrement appel à des systèmes d'identification faciale pour reconnaître les fraudeurs.

Cette masse importante de données complexes rend les infrastructures actuelles insuffisantes pour garantir une surveillance vigilante par un opérateur humain en raison de la fatigue dans des périodes de surveillance longue[FAR 11]. Pour permettre d'assister

les opérateurs humains dans leurs tâches de surveillance des événements, les systèmes de vidéosurveillance intelligente peuvent être utilisés. De tels systèmes nécessitent une rapidité et une robustesse pour la détection, le suivi, la reconnaissance et l'analyse des événements des objets en mouvement. Nous proposons dans cette thèse des méthodes pour la détection des objets en mettant la robustesse et la rapidité d'exécution comme un objectif principal dans ce travail.

5.3 Le document vidéo

La structure d'un document vidéo est définie comme une combinaison de deux ou trois flux d'information. L'image et le son sont les principales sources d'informations qui composent ce document. La sauvegarde et la lecture de ce dernier de manière synchrones permettent de générer une histoire. Le flux visuel comporte une séquence d'images fixes qui selon l'axe temporel apparaissent animées à une fréquence de 24 à 30 images par seconde. Le flux sonore est composé d'un ou plusieurs canaux (mono, stéréo). Le signal sonore est typiquement échantillonné entre 16000 et 48000 Hertz. Un troisième flux d'information généralement associée aux documents vidéo est le texte. Il provient soit d'un flux séparé, soit il est dérivé des sources audio et visuelle. Pour le stockage, la manipulation et la recherche de documents vidéo, il est nécessaire de se doter d'un moyen d'organiser l'information. On peut considérer qu'il existe plusieurs niveaux de structure liés à la donnée vidéo, il s'agit d'une organisation hiérarchique issue du monde de la production vidéo ou cinématographique. Cette hiérarchie met en évidence des séquences de granularités différentes. Un premier niveau de structure, appelé « plan », correspond au premier niveau de montage. Les niveaux supérieurs visent à monter en abstraction de façon hiérarchique, et tendent vers un niveau sémantique. En principe, ces différents niveaux de hiérarchie sont nommés selon le type de contenu du document vidéo (scène, histoire, documentaire, etc.). Une séquence vidéo brute est une suite d'images fixes, qui peut être caractérisée par trois principaux paramètres : la luminance, la résolution spatiale et la résolution temporelle.

5.4 Caractéristiques d'une vidéo

5.4.1 Nombre de frames par seconde

Le nombre d'images immobiles par unité de période de vidéo s'étend de six ou huit frames par seconde (frame/s) pour de vieux appareils-photo mécaniques à 120 frames ou plus par seconde pour de nouveaux appareils-photo professionnels. Pal (L'Europe, l'Asie, l'Australie, etc.) et SECAM (La France, la Russie, régions de l'Afrique, etc.) utilisent 25 frames/s, alors que NTSC (les États-Unis, le Canada, le Japon, etc.) utilise 29.97 frames/s.

Le taux minimum de frames pour réaliser l'illusion d'une image mobile est environ quinze frames par seconde. Le Débit binaire est une mesure du taux de contenu de l'information dans un flux vidéo. Il est quantifié en utilisant l'unité bit par seconde (bit/s ou bps) ou Méga de bits par seconde (Mbit/s). Un débit binaire plus élevé permet une meilleure qualité vidéo. Par exemple : un CD-Vidéo a un débit binaire d'environ 1 Mbit/s et il est de qualité inférieure à un DVD qui a un débit binaire d'environ 5 Mbit/s. Le Débit binaire variable (VBR) est une stratégie pour maximiser la qualité vidéo visuelle et pour réduire au minimum le débit binaire. Sur des scènes de mouvement rapide, un débit binaire variable utilise plus de bits qu'il fait sur des scènes de mouvement lent de la durée semblable, pourtant réalise une qualité visuelle cohérente.

5.4.2 Entrelacement

La vidéo peut être entrelacée ou progressive. L'entrelacement a été inventé comme une manière de réaliser la bonne qualité visuelle dans les limitations d'une largeur de bande étroite. Les lignes de balayage horizontales de chaque frame entrelacé sont numérotées consécutivement et divisées dans deux champs : champ impair composé des lignes impaires et champ pair contenant les lignes paires. NTSC, PAL et SECAM sont des formats entrelacés. Les caractéristiques visuelles abrégées de résolution incluent souvent i pour indiquer l'entrelacement. Par exemple, le format visuel de PAL est souvent indiqué par $576i50$, où 576 indique la résolution horizontale, i indique l'entrelacement, et 50 indique 50 frames par seconde. Dans le balayage progressif des systèmes, chaque frame inclut toutes les lignes de balayage. Le résultat est une résolution perçue beaucoup plus élevée. Un procédé connu sous le nom de « interlacing » peut être employé pour convertir un signal entrelacé analogique d'un DVD ou d'un satellite vers un balayage progressif sur un écran LCD, TFT ou un écran à plasma. Ce processus diminue considérablement la qualité visuelle

5.4.3 Résolution d'affichage

La taille d'une image visuelle est mesurée en Pixel pour la vidéo numérique et par les lignes de balayage verticales et horizontales de la résolution pour la vidéo analogique. Dans le domaine numérique, la résolution est indiquée comme $720/704/640 \times 480i60$ pour NTSC et $768/720 \times 576i50$ pour la résolution de PAL ou de SECAM. Cependant dans le domaine analogique, le nombre de lignes de balayage visibles reste constant ($486NTSC/576PAL$), tandis que la mesure horizontale change avec la qualité du signal.

5.4.4 Proportion

La proportion décrit les dimensions des écrans vidéo et les éléments images de la vidéo. Tous les formats vidéo populaires sont rectangulaires, et ainsi peuvent être décrits par un rapport entre la largeur et la hauteur. La proportion de l'écran d'une télévision traditionnel est 4 :3, soit environ 1.33 :1. Les télévisions à haute définition utilisent une proportion de 16 :9, soit environ 1.78 :1. Les pixels sur des moniteurs d'ordinateur sont habituellement carrés. Cependant les pixels utilisés dans une vidéo numérique ont souvent des proportions différentes, comme ceux utilisés dans les variantes de PAL et de NTSC du standard CCIR 601. Par conséquent, une image de NTSC DV qui est de 720 / 480 pixels est affichée avec une proportion de 4 :3 si les pixels sont minces et affichée avec une proportion de 16 : 9 si les pixels sont gros.

5.5 Analyse du contenu de la vidéo

Les documents vidéo peuvent avoir des contenus extrêmement varié (télévisés, documentaires, films, publicité, vidéosurveillance, etc.). La plupart du temps, ces documents ont une (ou plusieurs) structure interne. Comme les documents eux-mêmes, ces structures peuvent être très variées. Le type de structure est relatif au contenu sémantique du document. Il s'agit de structures qui ont un sens pour l'utilisateur. À ce titre, elles peuvent parfois apparaître mal définies, ambiguës ou subjectives. En ce qui concerne l'analyse du contenu, on peut distinguer deux niveaux de descriptions qui sont liés aux données vidéo :

- **Le niveau signal (bas niveau) :** proche de la représentation numérique des documents, il s'attache à décrire les caractéristiques « physiques » des segments d'une vidéo comme la couleur, la texture et la forme. Les informations correspondant à ce niveau sont en général de type numérique (tableaux de nombres, histogrammes de couleur par exemple).
- **Le niveau sémantique (haut niveau) :** proche de la façon dont les humains représentent le contenu des documents, il vise la description des concepts présents et des relations entre eux. Les informations correspondantes à ce niveau sont en général de type symbolique (concepts, relations, graphes).

5.6 Modélisation du Contenu Vidéo

La structure la plus classique consiste à décomposer la vidéo en des unités (segments) dont chacune représente un niveau de description. Cette décomposition est similaire à la modélisation hiérarchique du contenu vidéo. Dans un document vidéo, on distinguera des séquences qui correspondent à des unités sémantiques de la vidéo (thème, sujet, etc.). Dans des vidéos de taille importante, plus le contenu est varié, plus le nombre de sé-

quences est important. Les séquences peuvent elles-mêmes être décomposées en scènes. Ces scènes sont composées d'un ou plusieurs plans. Chaque plan correspond à une prise de vue, avec ou sans mouvement de caméra (plan fixe). Les plans sont séparés par des transitions.

Dans une vidéo, on peut définir des entités élémentaires (éléments d'information) significatives comme des personnages, des éléments décors, des objets, etc. On peut également définir des objets de granularité plus fine : des parties d'éléments comme le visage du personnage. Ces entités sont regroupées au niveau description du contenu vidéo et sont appelées des entités « Classe ». On peut également définir une entité événement (event) : un événement est « quelque chose qui arrive ». Il peut être décrit par les classes ou les occurrences de classe qui interviennent dans « ce qui arrive » et par des relations qui définissent « ce qui arrive ». Un événement se déroule en général pendant un certain intervalle de temps. La figure suivante illustre la structure d'un document vidéo.

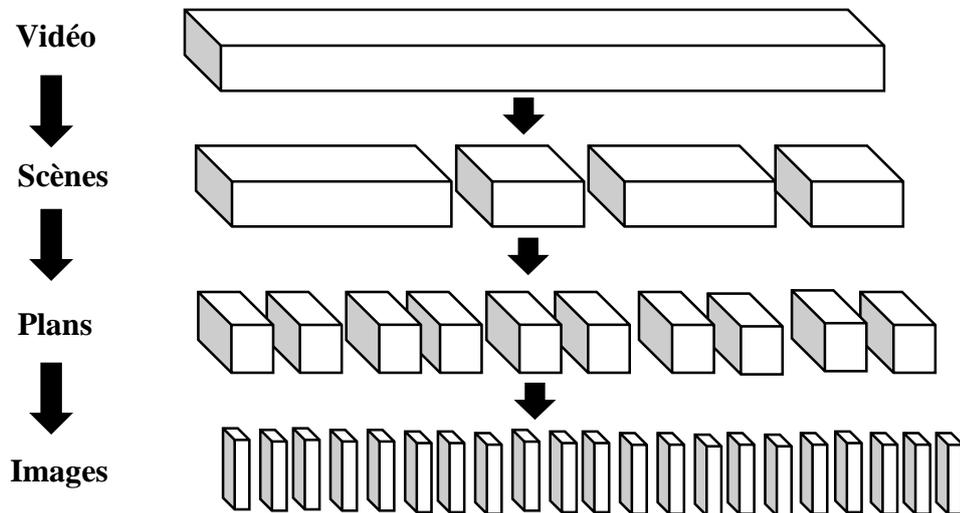


FIGURE 5.1 – Structure d'une vidéo

Pour la détection des transitions franches, chaque image est divisée en plusieurs blocs sans chevauchement, puis est extrait un histogramme pour chacun. Il suffit de comparer la similarité entre les images de la fenêtre coulissante avec l'image courante à travers le calcul d'une distance euclidienne par exemple, en donnant un poids à chaque bloc. Ensuite, on trie de façon décroissante les distances obtenues et on analyse ce classement. L'algorithme arrive à détecter une coupure lorsque le nombre d'images en aval classées dans la première partie du classement est supérieur à un seuil préfixé. Il a été observé qu'un mouvement rapide dans un plan peut fausser le résultat de la segmentation. Pour cela, on donne plus d'importance aux régions qui représentent le fond de l'image que les régions centrales. Ceci a permis l'obtention d'un gain de performance sur les données.

5.7 Les formats de fichiers vidéo numériques

Le format est une compression des informations vidéo dans une structure spécifique. La compression vidéo est très importante, car sans elle une seconde de vidéo occuperait une taille considérable causant ainsi des problèmes de manipulation et de transport. En effet, la vidéo non compressée nécessite énormément de place sur un disque dur et prend beaucoup de temps dans la transmission. Pour remédier à ces problèmes, le contenu de la vidéo doit être compressé pour réduire le débit des données. Cependant, la compression engendre toujours une perte d'information et donc une dégradation de la qualité des images de la vidéo.

L'algorithme de compression essaye de supprimer les redondances accumulées dans la vidéo. Ces derniers sont de nature spatiale ou temporelle. La redondance spatiale se base sur la supposition que, dans une image, deux points voisins sont souvent similaires. Tandis que la redondance temporelle se base sur l'hypothèse que deux images successives sont souvent fort similaires. Donc, la tâche de compression consiste à déterminer ces redondances et les éliminer. En se basant sur les notions citées précédemment, deux méthodes de compression sont envisagées : le codage par différence entre les images successives et le codage par prédiction de mouvement.

Plusieurs formes de compressions existent, plus ou moins standardisées, qui réduisent la taille des fichiers pour une manipulation plus aisée sur ordinateur. La vidéo est compressée par ce que l'on appelle des CODECS (Codeurs/Décodeurs). Il existe deux types de CODECS, les CODECS logiciels et les CODECS matériels. Le codec est un algorithme qui a pour rôles de compresser la vidéo pour la sauvegarde et l'enregistrement et de la décompresser ensuite pour une éventuelle lecture. Le choix du CODEC utilisé doit dépendre du type de séquence vidéo, du système d'exploitation utilisé, mais aussi de plusieurs facteurs à savoir :

- La qualité.
- Le taux de transfert des données.
- Le volume de stockage.
- Le temps nécessaire à la compression et à la décompression des informations.

L'encodage des séquences vidéo donne naissance à différents formats. Les formats les plus courants sont décrits dans le tableau ci-dessus :

5.8 Les images de références

L'image de référence est une image qui va être utilisée pour prédire une image et estimer le mouvement à compenser. Dans le cas échéant, l'image n'est rien d'autre qu'une image précédemment codée dans la séquence. Elle est souvent utilisée sous la forme d'objet mosaïque construit à partir d'une séquence de vues de l'objet considéré. Le principe de

Formats	Utilisation	Logiciel/Support
AVI	Prévoit une définition d'image maximale de 320 X 240 pixels et une cadence maximale de 30 fps. C'est un "conteneur" très pratique, le plus utilisé	Tous les logiciels de lecture vidéo
WMV	Un format conteneur de Microsoft et sa famille de codecs vidéo. On le trouve souvent chez des vidéos en streaming ou en téléchargement	Windows Media Player
MOV	Format conteneur d'Apple, pouvant contenir un très large choix de codecs, y compris haute définition	QuickTime
DivX Xvid	Les codecs vidéo les plus utilisés. Ils permettent de diviser par 6 la taille d'un film en DVD sans trop de perte de qualité. Xvid est le concurrent libre de DivX	Nécessite l'installation des codecs DivX / Xvid à jour, qui sont inclus dans de nombreux lecteurs vidéo
MKV	Format conteneur russe, pouvant contenir de très nombreux codecs	Se lit avec VLC, Media Player Classique et quelques autres après l'installation des codecs
H264 MPEG	C'est le codec utilisé pour la haute définition	Se retrouve dans les Blu-ray. Se lit avec la majorité des lecteurs vidéo à jour
FLV	Le format Flash Vidéo, utilisé sur Internet	s'ouvre avec d'Adobe Flash-Player, et VLC
Real Vidéo	Codec vidéo propriétaire de Real Networks	RealPlayer, RealAlternative
MPEG1	Appelé aussi format VCD 360 X 240 (pixels)	La plupart des logiciels de lecture vidéo
MPEG2	Appelé aussi format SVCD 480 X 576 (pixels)	La plupart des logiciels de lecture vidéo
MPEG2	Appelé aussi format DVD 720 X 576	La plupart des logiciels de lecture vidéo

TABLE 5.1 – Description des principaux formats vidéo

construction consiste à recalculer les images traitées dans le référentiel de l'objet mosaïque. Ces images peuvent être utilisées pour représenter sur une seule image la totalité d'une scène. Cette mosaïque est généralement plus grande qu'une image de la séquence.

Il est important de souligner qu'il est inutile d'utiliser toutes les images d'un plan pour obtenir une bonne extraction des caractéristiques visuelles. En effet, il serait par la suite impossible de conserver et d'utiliser cette information qui est par ailleurs redondante. Ainsi, une sélection d'une ou plusieurs images représentatives des plans serait nécessaire. Or, on peut déjà noter que si un plan présente de faibles mutations, il pourra être considéré comme statique. Il suffit alors de choisir l'image qui est la plus identique aux autres. En

pratique, cette recherche exhaustive est difficilement réalisable. Les approches empiriques sélectionnent simplement la première, la dernière ou l'image médiane du plan. Les avantages de la détection dans des images fixes et l'aspect temporel d'une vidéo permettent d'améliorer la détection et d'effectuer le suivi des objets en mouvement.

5.9 Architecture du système proposée

Cette partie décrit l'architecture du système proposé pour la reconnaissance des formes dans une séquence d'images extraites à partir d'une caméra fixe. Nous nous sommes intéressés à la détection et la reconnaissance des formes en mouvement dans un contexte ouvert et sans aucune condition a priori sur la nature de l'environnement ou le type d'objet à prendre en considération.

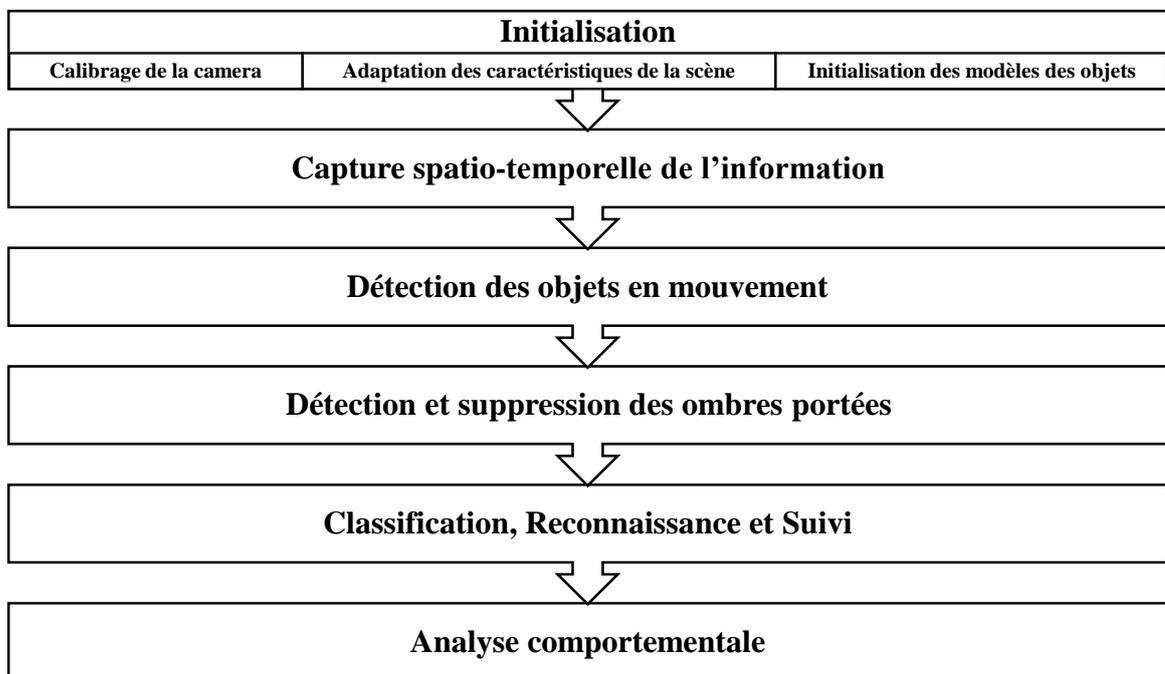


FIGURE 5.2 – Architecture du système de vidéosurveillance proposée

5.9.1 Initialisation

L'initialisation couvre les actions nécessaires pour assurer que le système commence les traitements par une interprétation correcte de la scène. Parfois, ce terme est également utilisé pour la représentation des données[FAR 11]. La plupart des méthodes d'initialisation sont accomplies hors ligne avant même le commencement des traitements, tandis que d'autres systèmes incluent cette phase comme un prétraitement. L'initialisation concerne

dans la plupart des temps le calibrage de la caméra, l'adaptation du système pour un nouvel environnement et l'initialisation des modèles des objets à suivre[FAR 12].

5.9.1.1 Calibrage de la caméra

Tous les systèmes de vision ont besoin d'une étape primordiale à savoir le calibrage de la caméra. Ceci peut être effectué d'une manière hors ligne ou en ligne selon les besoins. L'utilisation d'une caméra fixe nécessite dans la plupart des cas un seul recalibrage de la caméra lors de la première utilisation. Cependant, lorsque la caméra bouge, des procédés en ligne sont préférables pour éviter tout dysfonctionnement du système[FAR 15b].

5.9.1.2 Adaptation des caractéristiques de la scène

L'adaptation des caractéristiques de la scène est étroitement liée aux suppositions imposées sur l'apparence des formes dans l'étape du suivi des objets en mouvement. Dans de tels systèmes, une initialisation hors ligne est effectuée pour fixer un seuil et pour capturer une image de référence qui sera utilisée par le système dans les étapes ultérieures.

5.9.1.3 Initialisation des modèles des objets

L'initialisation de modèles est concentrée sur deux principaux procédés : la position initiale de l'objet à traiter et le modèle qui permet de représenter ce dernier[FAR 11 ; FAR 13b]. Les deux sont reliées à l'estimation de la position et l'attitude de l'objet. Le problème dans les approches basées modèles est d'arriver à trouver une posture initiale et de l'incrémenter d'un frame à un autre.

Dans la majorité des systèmes, l'ensemble des problèmes est réduit par la supposition d'avoir une posture initiale connue à l'avance ou par des opérateurs humains. Quelques systèmes ont exploité cette idée soit par un opérateur qui permet de spécifier la posture dans chaque frame de la vidéo ou par une spécification manuelle de certains frames clés en utilisant une interpolation entre les frames. Dans la littérature, on trouve qu'une fine partie des systèmes qui utilise une initialisation automatique des postures de départ.

La qualité du résultat de l'initialisation du modèle dépend fortement de la manière dont l'objet d'intérêt correspond au modèle du système. Quelques systèmes utilisent un modèle général calculé à partir d'un ensemble d'apprentissages. D'autres systèmes créent des modèles sur mesure pour chaque objet de la base. Dans ce dernier l'analyse de la posture initiale de l'objet est effectuée dans un premier temps par un modèle initial qui sera raffiné au fur et à mesure que le système traite des informations.

5.9.2 Capture spatio-temporelle de l'information

L'entrée du système est un ensemble de frames extraits à partir d'une caméra fixe. La nature analogique ou numérique de la caméra n'est pas importante vu que le système utilise des informations purement numériques sous format RGB et sans compression.

L'acquisition des frames s'effectue selon le processus suivant : la caméra capture le frame de la scène, ensuite l'envoie vers le système ; un dispositif matériel d'acquisition va enregistrer les images envoyées par la caméra dans une mémoire tampon ; le système va ensuite prendre les images de la mémoire vidéo pour les traiter par la suite. Dans les applications en temps réel, la caméra ne capture aucune image jusqu'à ce que le traitement du frame précédent sera totalement achevé. Nous avons utilisé, dans notre système, une seule caméra. Cependant, de multiples caméras peuvent être également utilisées surtout dans le contexte de la stéréovision ou dans le contexte d'un suivi multicible. Pour des contraintes techniques, nous avons séparé le module de la capture de la scène et les autres modules du système. En effet, tous les traitements proposés sont effectués sur des vidéos déjà capturées et sauvegardées.

5.9.3 Détection des objets en mouvement

La détection de mouvement peut être considérée comme un problème de segmentation ou il s'agit de classer dans chaque nouveau frame, les pixels en deux classes distinctes, l'arrière-plan (background) et l'avant-plan (foreground). Les éléments de l'arrière-plan représentent la structure de la scène et correspondent à l'ensemble des objets figés tous le long de la vie du système, tandis que l'avant-plan fait référence à tout objet ou structure de la scène en mouvement. Cette définition est purement théorique et ne peut être vérifiée dans la pratique étant donné que certains éléments de l'arrière-plan peuvent changer d'apparence à cause des conditions climatiques ou les changements de luminosité. D'autres objets peuvent avoir des mouvements partiels dans des périodes bien spécifiques telles que les mouvements des branches d'un arbre dans la présence du vent.

L'estimation de l'arrière-plan est une étape importante dans de nombreuses applications de vidéosurveillance. Toutefois, la détection de mouvement n'est pas considérée comme une fin en soi. Effectivement, il s'agit d'analyser temporellement les valeurs des pixels afin d'en extraire des informations de bas niveau qui seront utilisées dans des traitements de plus haut niveau à savoir : l'extraction, la reconnaissance, le suivi d'objets et l'analyse de comportement.

5.9.4 Détection et suppression des ombres portées

La détection et la suppression des ombres s'imposent comme une étape indispensable pour aboutir à un système de détection des objets en mouvement cohérent. Il y a plusieurs

types d'ombres générées par une ou plusieurs sources de lumière. Dans ce type d'application, les bruits sont principalement générés par les ombres portées. En effet, les ombres portées partagent les mêmes habitudes de déplacement que les objets en mouvement et il est pratiquement impossible aux méthodes de détection de mouvement de les discriminer. Ces bruits indésirables provoquent régulièrement de fausses détections dans les algorithmes de segmentation d'objets engendrant des fusions d'objets et des distorsions dans leurs formes. L'utilisation d'un traitement particulier pour la détection et la suppression d'ombre permet d'augmenter considérablement la qualité de la segmentation. Dans notre système, nous ne nous intéressons qu'aux ombres projetées par les objets en mouvements. Celles incorporées dans l'arrière-plan ne sont pas analysées et considérées quasi constantes dans le temps. L'algorithme de suppression d'ombre ne traite que les pixels appartenant aux éléments de l'arrière-plan.

5.9.5 Classification, Reconnaissance et suivi

Le principe de base du processus de reconnaissance est d'effectuer une classification pour retrouver un objet spécifique dans une vidéo. Cette étape consiste à extraire les connaissances à partir des données en utilisant les méthodes de classification. La difficulté dans cette étape est de maximiser la reconnaissance sachant qu'il n'existe aucune règle générale pour choisir le meilleur classificateur. Dans notre système, les étapes de reconnaissance et de suivi sont regroupées dans un module commun. En général, les deux processus sont séparés et les informations de reconnaissance sont utilisées pour le suivi des objets en mouvement. Cependant, l'emplacement des objets en mouvement peut également être utilisé pour minimiser les défauts de reconnaissance en limitant les espaces de recherche. Il permet également de corriger et de mettre à jour les classifieurs lorsque les objets en mouvement changent de formes dans le temps.

5.9.6 Analyse comportementale

L'analyse comportementale est un processus de haut niveau qui se trouve généralement comme un module au sommet de la hiérarchie dans des systèmes de vision. Ce processus cognitif permet d'extraire le contenu sémantique en détectant et analysant les comportements spécifiques dans la vidéo. Il s'agit de fournir une information sémantique sur ce qui se déroule dans la scène et de, par exemple, déclencher une alerte lorsqu'un événement particulier se produit. Les types d'événements à détecter peuvent contenir des informations sémantiques plus ou moins complexes et se diffèrent d'une application à une autre.

5.10 Conclusion

L'objectif de ce chapitre était de mettre les projecteurs sur l'un des domaines les plus actifs en ce moment à savoir la vidéosurveillance. Nous avons commencé par définir la vidéosurveillance, expliquer ces objectifs et présenter les domaines d'application les plus pertinents.

Les anciennes plateformes existantes ont montré leurs limites. En effet la fiabilité et la robustesse des systèmes de vidéosurveillance sont fortement liées aux opérateurs humains qui, à cause de leur nature, ne peuvent accomplir la tâche de surveillance en tout moment et avec une totale concentration sur une multitude d'écrans de surveillance. Pour cela, un traitement automatique de quelques tâches est indispensable afin d'augmenter la rentabilité de ces systèmes. Nous avons ensuite présenté les caractéristiques liées à la vidéo d'une manière générale pour une meilleure compréhension des processus et traitements qui seront appliqués ultérieurement à cette dernière. Le reste de ce chapitre a été dédié à l'architecture de notre système et aux différents modules qui le composent.

Nous allons essayer, dans les chapitres suivants, de présenter chaque module à part et d'essayer de donner le maximum de détail pour une meilleure visibilité des contributions apportées à ce domaine.

6

Analyse Spatio-Temporelle Détection des objets en mouvement

6.1 Panoramas

La détection automatique des objets en mouvement est une tâche très importante dans les systèmes de vidéo de surveillance. Les anciens systèmes basés sur les compétences des agents de sécurité ont montré leurs limites. Effectivement, beaucoup de séquences visuelles importantes peuvent leur échapper. Pour cela, un système de vidéosurveillance s'impose permettant de seconder les agents dans leurs travaux [DEN 10 ; GAO 10].

Les infrastructures, existantes, des systèmes de surveillance actuels offrent seulement la possibilité de capturer, sauvegarder et de distribuer les vidéos [LAI 10] afin de les traiter manuellement par des policiers ou par des agents spécialistes dans la détection des anomalies et des comportements suspects des humains.

Ce domaine a connu une expansion en termes de méthodes utilisées pour résoudre le problème de la vidéosurveillance dans ses différents niveaux [BIS 10 ; WOO 10], à savoir la segmentation, la détection, la reconnaissance et le suivi des objets en mouvement. Ce processus commence d'abord par l'extraction des objets qui n'appartiennent pas à l'arrière-plan [LI 10]. Pour cela la méthode la plus couramment utilisée est la modélisation de l'environnement [HED 10].

Dans la littérature ces approches se divisent en trois parties : les méthodes d'extraction de l'arrière-plan qui sont basées sur la comparaison entre l'image actuelle et l'image

de référence [KIM 10], la différence temporelle qui permet de découper la vidéo en un ensemble de fragments séparés dans un laps de temps constant et l'utilisation du flot optique qui permet d'estimer les mouvements grâce à un ensemble de points similaires d'un objet à travers plusieurs fragments [KIM 10 ; SAL 10]. La deuxième étape consiste à classifier les objets extraits de la première étape. Cette phase utilise des techniques telles que la détection des contours et les descripteurs spécifiques d'un objet [KIM 10 ; SAL 10]. La troisième étape est le suivi des objets à travers la scène.

6.2 Travaux similaires

Généralement, on peut distinguer deux types de contribution. La première est basée sur l'influence des caractéristiques dans le rendement des systèmes grâce à leur pouvoir discriminatif tandis que la deuxième se focalise sur le choix d'une bonne méthode de séparation entre l'avant-plan et l'arrière-plan avec une meilleure prise en charge des variations possibles. La méthode choisie doit également offrir un bon compromis entre temps d'exécution et espace mémoire pour une utilisation en temps réel.

Parmi les caractéristiques utilisées, on trouve les textures [YUK 11 ; DOU 10] qui ont pris un grand usage grâce à leurs pouvoirs d'être invariants par rapport aux changements de luminosité. Mais ces dernières ont la réputation d'être trop gourmandes en temps d'exécution et en espace mémoire limitant ainsi leur utilisation dans les environnements complexes et les applications en temps réel. Les espaces de couleur [LI 11 ; CHE 11 ; CHA 10 ; EL 09 ; LEI 09 ; HOR 99] restent le moyen de discrimination le plus utilisée dans les systèmes de vidéosurveillance. La puissance de ces derniers est mise en cause par les changements d'intensité lumineuse qui influence énormément sur leur rendement. L'hybridation entre les textures et les couleurs [LIN 10 ; AZA 10] est un moyen très efficace pour garder un bon compromis entre le temps d'exécution et l'invariance par rapport aux changements d'intensité lumineuse. D'autres caractéristiques telles que les épars [UZA 09], les transformations en odelettes [GAO 08 ; MEN 11], les transformations en cosinus [WAN 08b], les caractéristiques spectrales [LI 04b] et spatio-temporelle [LI 04b ; VEM 09] ainsi que les champs de tenseurs [CAS 10] ont pris un grand intérêt comme un moyen de contourner le paradoxe qualité et temps d'exécution. L'intérêt apporté aux caractéristiques n'a pas suffi pour mettre en œuvre des systèmes de qualité et de fiabilité. En effet, les méthodes utilisées pour la séparation avant-plan/arrière-plan jouent un rôle très important pour réaliser un système performant. Dans la littérature, des dizaines de méthodes pour la modélisation de l'arrière-plan. Cette diversité est une conséquence directe de vouloir réaliser un système standard capable de donner de bons résultats, quelles que soient les conditions d'utilisation. Parmi les méthodes utilisées, on trouve : les échelles invariantes [YUK 11], les approches basées régions [LEI 09 ; UZA 09], les champs de Markov [LUO 06], la logique floue [AZA 10], les approches par bloc [LIN 10 ; VEM 09],

l'approximation des densités séquentielles [WAN 08b], le filtre médian sur plusieurs niveaux [MA 06] et l'utilisation d'un Codebook [KIM 04]. La règle de décision bayésienne a aussi pris une part dans la modélisation de l'arrière-plan [XU 09 ; LI 04b].

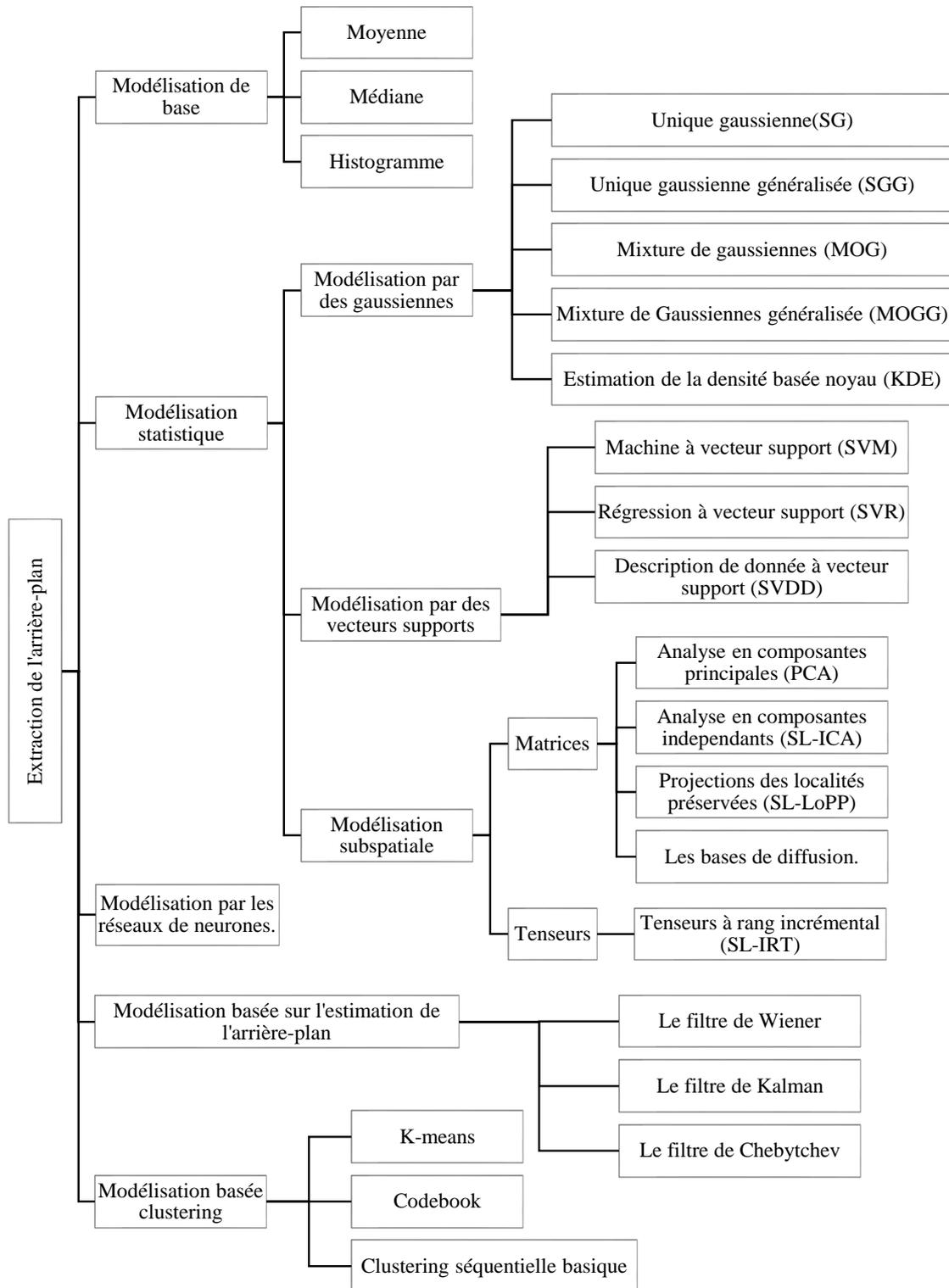


FIGURE 6.1 – Taxonomie des systèmes d'extraction de l'arrière-plan

6.3 Taxonomie proposée

Le suivi des objets en mouvement est un problème important dans les systèmes automatiques de vidéosurveillance. Un tel système nécessite une opération primordiale à savoir : la séparation entre l'arrière-plan et l'avant-plan. Plusieurs travaux de recherche ont été proposés afin d'améliorer le rendement de cette opération. Effectivement, une bonne méthode d'extraction engendre une simplification des traitements ultérieurs permettant ainsi un gain en temps d'exécution et en espace mémoire. Nous avons essayé de regrouper les méthodes proposées dans une taxonomie afin de mieux visualiser les apports de recherche effectués dans ce domaine (Figure 6.1).

6.4 Motivations

Les GMM sont parmi les méthodes les plus utilisées pour la modélisation de l'environnement [YU 10a]. Les travaux de Hedayati et al. [HED 10] ont montré que les GMM offrent un bon compromis entre qualité d'extraction et temps d'exécution par rapport aux autres méthodes d'extraction d'arrière-plans. Les GMM sont performants dans le cas où la vidéo ne contient pas de variation locale où un changement de luminosité [ZHA 10]. Cependant l'apprentissage d'un arrière-plan encombré et contenant plusieurs objets en mouvement reste le problème majeur de cette méthode [ZHA 10 ; HED 10].

Plusieurs travaux ont essayé d'améliorer le rendement de cette méthode dans les environnements à multiple variation de luminosité et à forte condensation de l'arrière-plan. Les premières idées se sont focalisées sur la proposition d'un remplaçant des caractéristiques couleur utilisées [CAS 10]. Des modèles hybrides tels que les GMM et la méthode K-means [CHA 10], les GMM et la logique floue [EL 09] ainsi que les GMM et les modèles d'arrière-plan adaptatif [DOU 10] ont été proposés pour pallier les inconvénients des GMM. D'autres travaux se sont focalisés sur l'amélioration de la vitesse d'apprentissage [KAN 10 ; SUO 08] à travers un taux d'apprentissage adaptatif ; et sur le temps d'exécution [LI 11] par des algorithmes utilisant un parallélisme réel sur des machines multiprocesseurs. D'autres systèmes utilisent deux arrières plans [CHE 11] pour résoudre le problème de changement de luminosité entre le jour et la nuit.

6.5 Proposition d'un système d'extraction d'arrière-plan

Nous allons nous intéresser à la détection des objets en mouvement dans une vidéosurveillance par une caméra fixe où la modélisation de l'arrière-plan est basé sur le modèle de mixture de gaussiennes (GMM). Pour surmonter les problèmes de variations locales et de la luminosité, l'image initiale de la vidéo est segmentée en plusieurs régions, ensuite chaque région est assignée à un superviseur qui peut détecter des changements dans cette

région en comparant les histogrammes et seuls les segments qui ont subi un changement sont mis à jour par le GMM. Pour une meilleure visibilité et selon la nature du traitement effectué, le système est organisé autour de trois parties (figure 6.2) comme suit : acquisition ; traitement et détection.

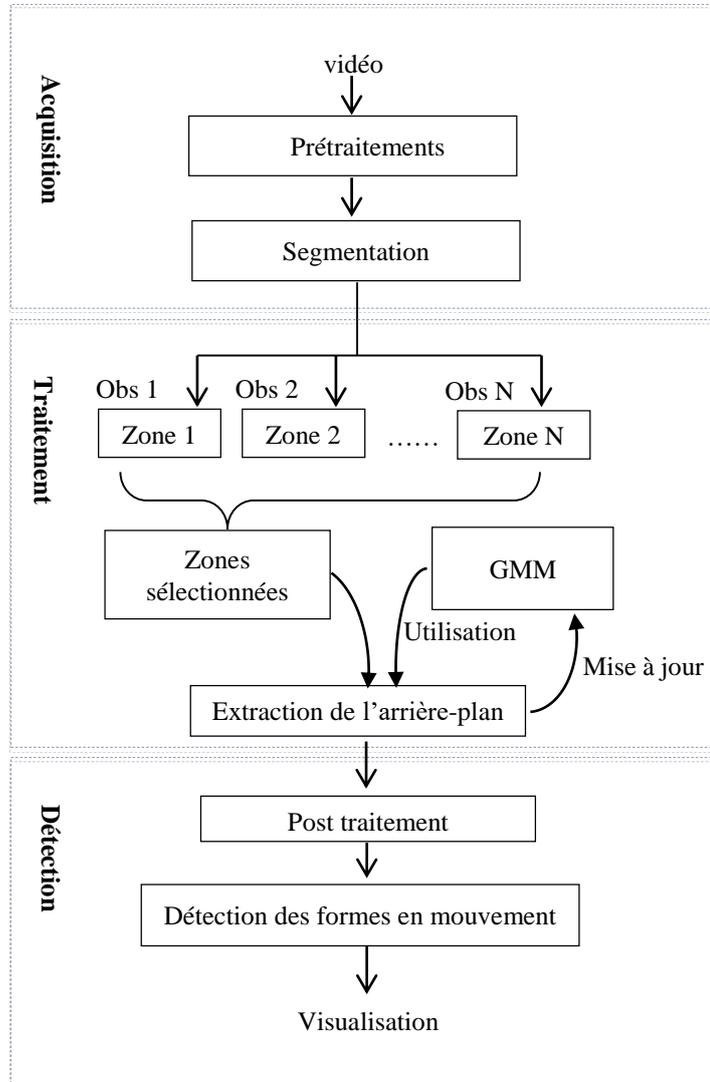


FIGURE 6.2 – Architecture du module de détection et d'extraction des formes en mouvement

6.6 Acquisition

Nous avons regroupé dans cette partie tous les traitements liés à l'acquisition et les prétraitements permettant ainsi de donner une structure plus simplifiée pour les blocs suivants. Généralement, le terme acquisition désigne la capture de la scène ou de l'environnement via une caméra. Dans notre système, la capture est effectuée a priori à cause de plusieurs contraintes techniques. L'aspect sécurité est un autre problème que nous avons

rencontré lors de l'installation des caméras pour l'expérimentation.

6.6.1 Prétraitement

Le prétraitement permet d'appliquer un ensemble de techniques pour améliorer les données qui seront traitées par la suite. Nous avons d'abord appliqué une fragmentation qui permet de transformer la vidéo capturée en un ensemble de frames ou chaque frame représente une séquence de la vidéo. Ensuite, un filtre médian est appliqué afin de corriger les erreurs de capture. Le filtre médian est réputé d'être très simple à implémenter, rapide et efficace pour rendre l'image uniforme. Les images extraites de la vidéo sont en mode RGB. Cependant, cette représentation n'est pas adéquate à cause de l'influence de la luminosité sur la description des objets [HAQ 10]. Pour cela nous avons effectué un transfert vers le modèle HSV reconnu d'être le modèle le plus proche de la perception humaine et dans lequel la luminosité n'affecte qu'un seul composant [HAQ 10 ; FAR 13b]. La conversion est effectuée selon l'équation suivante :

$$h = \begin{cases} 0 & \text{si max} = \text{min} \\ \left(60^\circ \times \frac{g-b}{\text{max}-\text{min}} + 360^\circ\right) \bmod 360^\circ & \text{si max} = r \\ 60^\circ \times \frac{b-r}{\text{max}-\text{min}} + 120^\circ & \text{si max} = g \\ 60^\circ \times \frac{r-g}{\text{max}-\text{min}} + 240^\circ & \text{si max} = b \end{cases} \quad (6.1)$$

$$s = \begin{cases} 0 & \text{si max} = 0 \\ 1 - \frac{\text{min}}{\text{max}} & \text{sinon} \end{cases} \quad v = \text{max}$$

Ou $h \in [0, 360]$ et $s, v, r, g, b \in [0, 1]$

r, g, b désignent respectivement les coordonnées RGB,

h, s, v désignent respectivement les coordonnées HSV.

max désigne la plus grande valeur entre r, g et b ; et min la plus petite.

6.6.2 Segmentation

Cette opération est utilisée uniquement dans la phase d'initialisation du système. La première image est divisée en N segments de taille égale afin de minimiser les variations locales et de simplifier la tâche des observateurs d'arrière-plan [FAR 13b]. Le choix du nombre de segments varie avec la taille des images originales, l'environnement et le domaine d'utilisation. Nous avons constaté que le nombre de zones a une grande influence sur la qualité du système. La sursegmentation conduit à une baisse de performance et la sous-segmentation conduit à une baisse dans la qualité ce qui implique que le choix du nombre de segments doit être compris entre la performance et la qualité. En effet, un grand

nombre de régions nous amène au point de départ (approche basée sur le pixel). Dans le cas où le nombre de zones est petit (la taille de la zone est grande), les variations locales accumulées dans la même zone forceront le système à considérer ces derniers comme une variation intense ce qui conduit le système à appliquer une mise à jour sur tous les pixels appartenant à la région. Cette fausse mise à jour va dégrader considérablement les performances du système. Pour éviter les problèmes de chevauchement, nous supposons que toutes les régions sont complètement distinctes équation 6.2 . Soit A une région de l'image I et N le nombre total de régions dans la même image.

$$\forall i, j \in \{1, \dots, N\} \text{ si } A_i \in I \text{ et } A_j \in I \text{ alors } i \neq j \implies A_i \cap A_j = \emptyset \quad (6.2)$$

6.7 Traitement

Cette étape représente le cœur du système de détection et d'extraction de l'arrière-plan. Le travail collaboratif entre les observateurs d'arrière-plan et le GMM ont permis d'assurer la stabilité de l'arrière-plan. Nous allons essayer dans ce qui suit de détailler chaque partie à part en mettant l'accent sur les algorithmes de détection et de mise à jour[FAR 13a].

6.7.1 Le modèle de mixture de gaussiennes

Stauffer et Grimson [STA 99b] ont présenté une nouvelle modélisation de l'arrière-plan basée sur une mixture de plusieurs gaussiennes. Les MoG ont prouvé leurs efficacités face aux changements de luminosité, les mouvements répétitifs, l'encombrement, l'introduction et le retrait des objets de la scène et les mouvements lents des objets. La motivation sur laquelle est basée la proposition de ce modèle est qu'un modèle unimodal de l'arrière-plan ne pouvait gérer le bruit d'acquisition d'image, le changement de lumière et l'existence de multitude de surfaces pour un pixel particulier dans le même temps[FAR 12 ; FAR 13b].

Une MoG est un modèle statistique qui suppose que les données proviennent d'une somme pondérée de plusieurs distributions de gaussiennes. Stauffer et Grimson [STA 99b] ont présenté une méthode GMM adaptative pour modéliser un fond dynamique dans des séquences d'images. Si les K distributions de gaussiennes sont utilisés pour décrire l'historique d'un pixel, l'observation d'un pixel donné sera forcément dans l'un des K états à la fois [CHA 10]. K détermine la multimodalité de l'arrière-plan et la sélection de K est généralement basé sur la mémoire disponible et de la puissance de calcul. Stauffer et Grimson [STA 99b] ont proposé de fixer K entre 3 et 5.

6.7.1.1 Initialisation

Tout d'abord, chaque pixel est caractérisé par son intensité dans l'espace de couleurs HSV. Ensuite, la probabilité d'observer la valeur du pixel courant X_t est donnée par l'équation suivante dans le cas multidimensionnel.

$$P(X_t) = \sum_{i=1}^k \omega_{i,t} \eta(X_t, \mu_{i,t}, \Sigma_{i,t}) \quad (6.3)$$

Où : k est le nombre de gaussiennes associées à chaque pixel, $\omega_{i,t}$ est le poids calculé, $\mu_{i,t}$ est la moyenne et $\Sigma_{i,t}$ est la matrice de covariance qui sont respectivement évalués pour la $i^{\text{ème}}$ gaussienne à l'instant t . η est la fonction de densité de probabilité gaussienne :

$$\eta(X_t, \mu, \Sigma) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(X_t - \mu)^T \Sigma^{-1} (X_t - \mu)} \quad (6.4)$$

Où D exprime la dimension du vecteur X .

Au début du processus d'extraction de l'arrière-plan, pour chaque pixel, les distributions des K gaussiennes sont initialisées avec une moyenne prédéfinie, une variance élevée et un faible poids. Pour obtenir un arrière-plan stable, nous avons utilisé les 50 premiers frames pour l'apprentissage des mixtures de gaussiennes.

6.7.1.2 Mise à jour des paramètres

Il existe plusieurs techniques disponibles pour estimer les paramètres d'une GMM. La méthode la plus populaire et la plus utilisée pour l'estimation est le maximum de vraisemblance (ML). Le but de l'estimation ML est de trouver les paramètres du modèle qui maximisent la vraisemblance du GMM sur un ensemble de données d'apprentissage. Pour une séquence de T vecteurs d'apprentissage de $X = x_1, \dots, x_T$, si on suppose qu'il existe une indépendance entre les vecteurs, alors la vraisemblance du GMM peut être écrite selon l'équation 6.5 suivante.

$$P(X|\lambda) = \prod_{t=1}^T P(X_t|\lambda) \quad (6.5)$$

Malheureusement, cette expression est une fonction non linéaire des paramètres λ et la maximisation directe est impossible. Cependant, les estimations des paramètres ML peuvent être obtenues en utilisant de manière itérative un cas particulier de l'algorithme espérance-maximisation (EM).

L'idée de base de l'algorithme EM est de commencer par un modèle initial λ pour estimer un nouveau modèle $\hat{\lambda}$, tel que $P(X|\hat{\lambda}) \geq P(X|\lambda)$. Le nouveau modèle devient alors le modèle initial pour la prochaine itération et le processus est répété jusqu'à ce qu'un seuil de convergence soit atteint. Le modèle initial est généralement obtenu en

utilisant une forme binaire de l'histogramme. Pour chaque itération d'EM, la réestimation des équations 6.6, 6.7 et 6.8 est utilisée pour garantir une augmentation monotone de la valeur de vraisemblance du modèle.

$$\bar{\omega}_i = \frac{1}{T} \sum_{t=1}^T \dot{P}(i|X_t, \lambda) \quad (6.6)$$

$$\bar{\mu}_i = \frac{\sum_{t=1}^T \dot{P}(i|X_t, \lambda) X_t}{\sum_{t=1}^T \dot{P}(i|X_t, \lambda)} \quad (6.7)$$

$$\bar{\sigma}_i^2 = \frac{\sum_{t=1}^T \dot{P}(i|X_t, \lambda) X_t^2}{\sum_{t=1}^T \dot{P}(i|X_t, \lambda)} - \bar{\mu}_i^2 \quad (6.8)$$

Ou $\bar{\omega}_i$, $\bar{\mu}_i$ et $\bar{\sigma}_i^2$ représente respectivement les nouvelles valeurs du poids, de la moyenne et de la variance. La probabilité a posteriori pour le composant i est donnée par l'équation suivante :

$$\dot{P}(i|X_t, \lambda) = \frac{\omega_i \eta(X_t, \mu_{i,t}, \Sigma_{i,t})}{\sum_{k=1}^M \omega_k \eta(X_t, \mu_{k,t}, \Sigma_{k,t})} \quad (6.9)$$

Malgré la réputation de l'algorithme ML ce dernier occupe énormément d'espace mémoire et nécessite une puissance de calcul non négligeable ce qui rend son utilisation inappropriée dans un environnement en temps réel ou d'un système embarqué. En plus, EM converge vers un maximum local ; mais nous n'avons aucune garantie que ce maximum soit « bon », c-a-d soit proche en valeur du maximum global. Pour se rapprocher de ce but, on doit encore alourdir le calcul. Enfin, EM présente le risque de trouver des solutions « dégénérées » présentant au moins une composante ajustée sur un seul individu ; or la vraisemblance d'une telle Gaussienne peut être infinie (i.e. on se rapproche d'un pic de Dirac), mettant ainsi à mal notre algorithme.

Pour une éventuelle application en temps réel, nous avons utilisé l'algorithme d'approximation K-Means [CHA 10 ; XU 10 ; FAR 13a] pour la mise à jour du modèle. Chaque valeur d'un nouveau pixel est vérifiée, à travers les k distributions de gaussiennes, pour déterminer si la distance entre le pixel et chacune des gaussiennes est inférieure à l'écart type de distribution selon l'équation suivante :

$$\frac{|P_t - \mu_i|}{\sigma_i} < 2.5 \quad (6.10)$$

Si aucune distribution ne vérifie l'équation 6.10, alors le pixel est associé à l'avant-plan. Les paramètres, de la gaussienne la moins probable, sont remplacés par la moyenne, la variance et le poids du pixel courant suivant les équations 6.11, 6.12 et 6.13

$$\mu_k = X_t \quad (6.11)$$

$$\sigma_k^2 = \text{unegrandevarianceinitiale} \quad (6.12)$$

$$\omega_k = \text{faiblepoid} \quad (6.13)$$

Les poids de toutes les gaussiennes seront mis à jour selon l'équation suivante :

$$\omega_{k,t} = (1 - \alpha) \cdot \omega_{k,t-1} + \alpha M_{k,t} \quad (6.14)$$

Où α est le coefficient d'apprentissage qui détermine la vitesse d'adaptation de modèle. $M_{k,t}$ est égal à 1 pour la gaussienne qui correspond à l'arrière-plan et égal à 0 pour les autres. Après la mise à jour des poids, une étape de normalisation est effectuée pour garantir que la somme des poids est toujours égale à 1.

Les autres paramètres des distributions ayant vérifié l'équation 6.10 sont mis à jour selon les équations 6.15, 6.16 et 6.17.

$$\mu_{k,t} = (1 - \phi_k) \cdot \mu_{k,t-1} + \phi_k \cdot P_t \quad (6.15)$$

$$\sigma_{k,t}^2 = (1 - \phi_k) \cdot \sigma_{k,t-1}^2 + \phi_k (P_t - \mu_{k,t})^T (P_t - \mu_{k,t}) \quad (6.16)$$

avec

$$\phi_t = \alpha \cdot \eta (P_t | \mu_k, \sigma_k) \quad (6.17)$$

6.7.1.3 Extraction de l'arrière-plan

Pour décider si X_{t_j} est inclus dans l'arrière-plan, les distributions sont ordonnées selon la valeur $w_{k,t} / \sigma_{k,t}$. Cet ordre utilise l'hypothèse selon laquelle un pixel d'arrière-plan correspond à un poids élevé avec une variance faible pour la raison que l'arrière-plan est pratiquement constant et il est plus présent que les objets en mouvement [FAR 13b]. Les premières B distributions qui vérifient l'équation 6.18 sont sélectionnées pour représenter l'arrière-plan.

$$B = \arg \min_b \sum_{k=1}^b \omega_{k,t} > T \quad (6.18)$$

Le seuil T représente la portion minimum du poids total attribué au modèle d'arrière-plan. Si une petite valeur T est choisie, l'arrière-plan devient unimodal. Si T est plus élevée, une distribution multimodale causée par un mouvement répétitif de l'arrière-plan peut résulter de plusieurs objets de l'arrière-plan ce qui permet aux pixels de l'arrière-plan d'accepter plus d'une distribution gaussienne. L'utilisation d'un seuil T unique pour le modèle engendre une mauvaise classification surtout lorsque la scène contient à la fois des zones dynamique et statique. Un seuil plus élevé permet d'avoir une meilleure classification dans un environnement dynamique, mais génère des détections incorrectes pour les objets en mouvement dans un arrière-plan statique.

La figure 6.3 montre un simple processus de classification où les pixels accumulés définissent l'arrière-plan représenté par les distributions de gaussiennes tandis que les

pixels dispersés sont classés comme premier plan

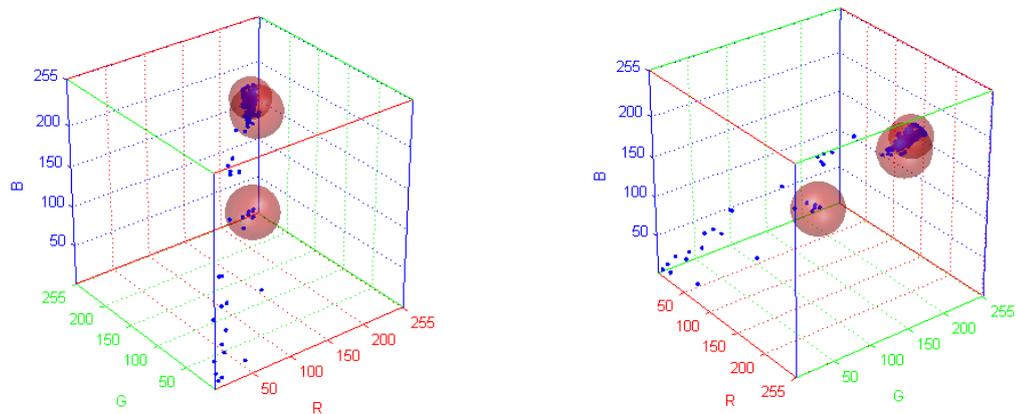


FIGURE 6.3 – Exemple de modélisation par une mixture de gaussiennes



FIGURE 6.4 – Résultat de l'extraction de l'arrière-plan

6.7.2 Les observateurs de l'arrière-plan

Les méthodes basées sur GMM utilisent la valeur du pixel pour détecter un changement probable dans l'arrière-plan en se basant sur l'hypothèse selon laquelle un objet en mouvement est un ensemble de pixels en mouvement.

Cette vision est très utile, car elle ne nécessite aucune connaissance a priori sur la nature des objets et leurs trajectoires. Cependant, l'environnement naturel est bien loin d'être parfait. La présence de poussière, la variation de la luminosité, la pluie, le vent, etc. influencent sur la valeur des pixels conduisant à une mauvaise détection de mouvement.

Une mauvaise détection conduit le système à faire des erreurs dans les étapes suivantes, soit en déformant les objets en mouvement ou en signalant un faux mouvement. Pour

surmonter ces problèmes, nous avons proposé l'utilisant d'un ensemble de threads que nous avons appelé « les observateurs de l'arrière-plan ». Nous avons défini et attribué pour chaque observateur, une zone dans une image. Le rôle de l'observateur est de surveiller et de signaler la présence de l'activité dans sa zone.

Après attribution d'un observateur à une zone, il crée et enregistre l'histogramme de cette dernière dans sa mémoire. Ainsi, à chaque nouveau traitement, l'observateur peut détecter des changements dans la zone en comparant l'état courant de l'histogramme avec l'état stocké [FAR 15a]. La présence d'un objet se déplaçant dans une ou plusieurs zones déclenche simultanément et indépendamment de tous les observateurs de l'arrière-plan. Ce processus permettra d'éliminer les variations locales, parce que seules les régions qui ont subi des changements importants et significatifs seront traitées par le système.

Dans la littérature, il existe plusieurs méthodes pour mesurer la similarité (dis-similarité) entre deux distributions de probabilités P et Q et les plus utilisés sont les suivants : la distance de Bhattacharyya [DJO 90] ; la distance de Kullback-Leibler et la divergence Jensen-Shannon [LIN 91]. Ces mesures sont définies respectivement comme suit.

6.7.2.1 Bhattacharyya distance

$$D_B(p, q) = -\ln(BC(p, q)) \quad (6.19)$$

avec

$$BC(P, Q) = \sum_{x \in X} \sqrt{P(x)Q(x)} \text{ et } 0 \leq BC \leq 1 \quad (6.20)$$

6.7.2.2 Divergence de Kullback-Leibler

$$D_{KL}(P||Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)} \quad (6.21)$$

6.7.2.3 Divergence de Jensen-Shannon

$$JSD(P||Q) = \frac{1}{2} \sum P(\log P - \log M) + \frac{1}{2} \sum Q(\log Q - \log M) \quad (6.22)$$

avec

$$M = \frac{P+Q}{2}$$

Notez que dans les équations 6.21 et 6.22

$$P = Q \text{ si et seulement si } D_{KL}(P||Q) = JSD(P||Q) = 0$$

Dans notre cas la possibilité d'avoir l'égalité des deux distributions de probabilité est presque impossible. Pour cette raison, l'utilisation du coefficient de Bhattacharyya BC

semble appropriée, car on peut choisir un seuil T qui se trouve en 0 et 1. Par conséquent, si la différence entre P et Q est supérieure au seuil T , l'observateur de l'arrière-plan déclenche un signal indiquant la détection d'un changement dans la région[FAR 13a].

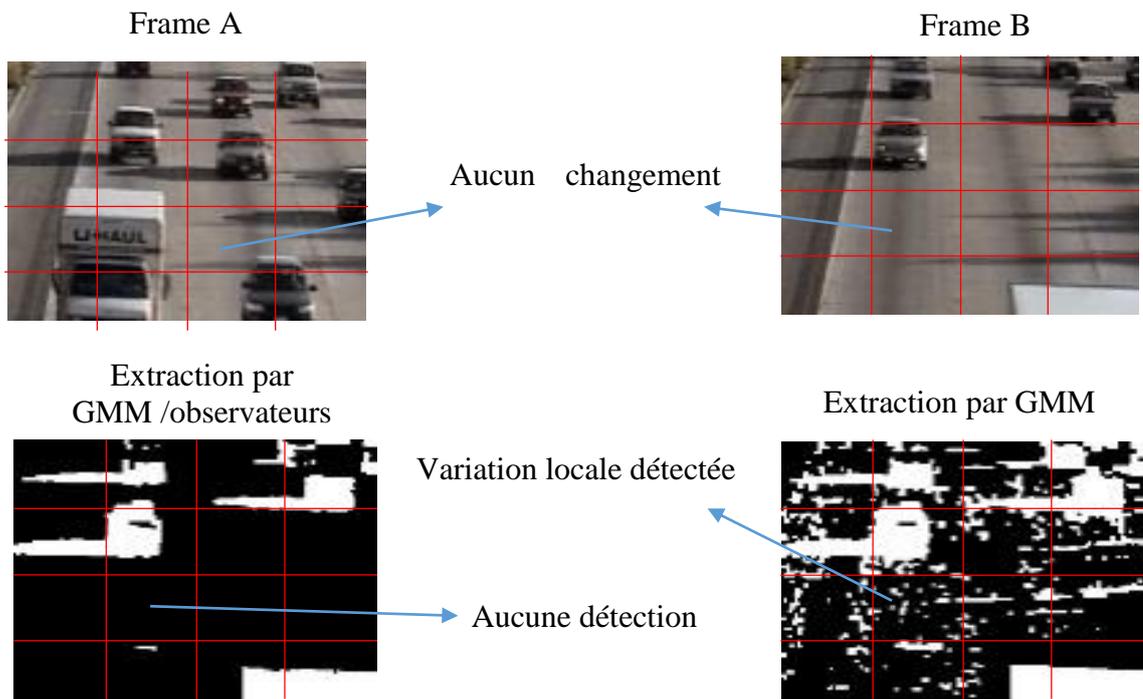


FIGURE 6.5 – L'effet de la variation locale sur les performances des GMM

Dans notre système, le coefficient de Bhattacharyya calculée pour les frames A et B de

la figure 6.5 donne presque la même valeur. Cette égalité indique à l'observateur que la région n'a eu aucun changement significatif. Par conséquent, les pixels perturbateurs ne seront pas mis à jour et ne seront pas considérés comme faisant partie d'un objet en mouvement. Ce mécanisme n'affecte pas la capacité du GMM à s'adapter par des changements lents de luminosité. En effet, tout changement important aura une incidence directe sur l'ensemble des valeurs de pixels dans la zone, ce qui influe sur la forme de l'histogramme[FAR 15a]. Voici un algorithme qui permet de résumer le travail coopératif des observateurs de l'arrière-plan et de la GMM.

Algorithme 1 : Extraction de l'arrière-plan

Initialisation

- 1) Diviser la première image en plusieurs zones de taille égale
- 2) Attribuer un observateur pour chaque zone
- 3) Convertissez tous les pixels de l'image de RGB à HSV
- 4) Calculer et stocker l'histogramme de couleur de chaque zone
- 5) Initialiser les paramètres de la GMM

Itération

pour chaque nouveau frame **faire**

 Convertissez tous les pixels de l'image de RGB à HSV

pour chaque zone **faire**

 Calculer l'histogramme de couleur

 Mesurer le degré de similitude entre l'histogramme calculé et celui stocké

si la différence est supérieure à un seuil T **alors**

 Mémorisez le nouvel histogramme

 Mise à jour des paramètres de la GMM

fin

fin

fin

6.8 Détection

Ce bloc a pour objectif de raffiner les étapes précédentes, de supprimer les bruits indésirables et de détecter définitivement des objets en mouvement.

6.8.1 Post-traitement

Le post-traitement permet de corriger les erreurs commises dans l'étape précédente. En effet, les sorties des algorithmes d'extraction de l'arrière-plan que nous avons présenté dans les sections précédentes contiennent généralement des bruits et donc les données ne sont pas appropriées pour un traitement ultérieur. Il y a différents facteurs qui causent le bruit dans l'extraction de l'arrière-plan, nous citons :

Bruit de l'appareil d'acquisition. c'est un bruit causé par les composants d'acquisition d'image. L'intensité d'un pixel qui correspond à une arête entre deux objets de couleurs différentes dans la scène peut être capturée avec la couleur d'un objet dans un frame et par la couleur de l'autre objet dans le frame suivant.

Bruit de l’algorithme d’extraction. la nature statistique du modèle utilisé génère des bruits au niveau de la modélisation (initialisation et mise à jour) et au niveau de la décision (seuil) affectant ainsi l’opération de l’extraction de l’arrière-plan qui laisse un nombre important de bruit et de trous, ce qui influence par la suite sur les algorithmes de segmentations et sur le comportement du système.

Bruit de réflexion. Si une source de lumière, par exemple le soleil, se déplace, elle incite certaines parties de l’arrière-plan de la scène à réfléchir la lumière. Ce phénomène affecte le rendement des algorithmes de détection des objets en mouvement par la détection des régions de réflectance de premier plan.

Bruit de couleurs. Certaines parties des objets peuvent avoir la même couleur que l’arrière-plan de référence qui se trouve derrière eux. Cette ressemblance provoque une erreur au niveau de l’algorithme d’extraction de l’arrière-plan qui va confondre entre l’objet et l’arrière-plan et détecte l’objet en mouvement comme faisant partie de l’arrière-plan.

Changement soudain d’illumination. le changement soudain d’illumination (par exemple allumer une lampe dans une pièce surveillée) affecte les algorithmes de détection qui ne parviennent plus à détecter les objets en mouvement avec précision.

Présence d’ombres. Les ombres projetées sur les objets sont détectées comme avant-plan par la plupart des algorithmes de détection. Ce problème sera traité séparément dans les prochaines sections.

6.8.1.1 Application des opérations morphomathématiques

Pour remédier aux problèmes relatifs aux couleurs, aux réflexions et à la nature de l’algorithme, nous avons proposé comme post-traitement des opérations morphomathématiques. La dilatation permet de boucher les trous laissés à l’intérieur des objets, tandis que l’érosion permet d’éliminer les points indésirables générés par la présence de poussière ou par un changement infime de luminosité non détecté par les agents. L’ordre de ces opérations affecte la qualité des résultats et augmente la complexité de calcul dans la suppression du bruit. Par exemple, si nous appliquons la dilatation suivie par l’érosion, les grandes régions de bruit isolées ne peuvent être supprimées, car l’opération de dilatation permet d’élargir leurs frontières avec un pixel et l’érosion va supprimer les pixels ajoutés par la dilatation ce qui nous amène au point de départ. Cependant, cet ordre permet d’éliminer les bruits qui se trouvent à l’intérieur des objets détectés comme avant plan.

Dans le cas où nous appliquons ces opérations dans l’ordre inverse, qui est l’érosion suivie d’une dilatation, les régions de bruits isolés seront supprimées, mais cette fois, nous ne serions pas en mesure de fermer les trous qui se trouvent à l’intérieur des objets. Après

avoir expérimenté différentes combinaisons de ces opérations, nous nous sommes fixé à la séquence suivante : deux niveaux de dilatation suivie par trois niveaux d'érosion et enfin un niveau de dilatation.

La première opération de dilatation élimine les trous qui se trouvent à l'intérieur des objets du premier plan qui ont été détectés comme arrière-plan. Elle permet également d'élargir les frontières des régions et de corriger les coupures qui se trouvent au niveau des contours d'objet.

Dans l'étape suivante, les trois niveaux d'érosion suppriment les pixels isolés et les contours supplémentaires des régions ajoutées par l'étape précédente. La dernière étape, un niveau de dilatation, est utilisée pour compenser les effets supplémentaires de l'érosion.

La figure 6.6 montre que ces opérations sont nécessaires pour améliorer les images après l'extraction de l'arrière-plan permettant de mieux détecter les contours des objets et simplifiant les autres tâches du système.

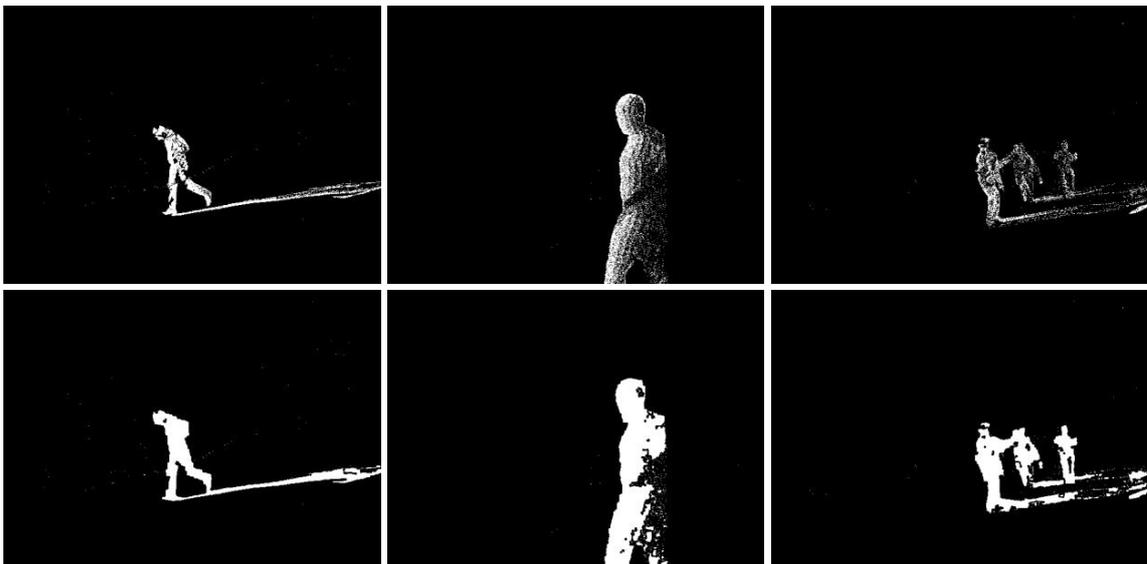


FIGURE 6.6 – Résultat des opérations morphomathématiques sur les objets

6.8.1.2 Élimination des changements soudains de luminosité

La détection soudaine de changement d'éclairage est aussi un post-traitement très important qui permet à un système de surveillance intelligent de continuer à détecter et à analyser le comportement des objets correctement. Un changement global peut par exemple se produire à cause d'un soleil couvert ou découvert par des nuages dans les environnements extérieurs ou à cause d'une lumière éteinte dans un environnement intérieur. Ces deux changements génèrent un changement soudain de luminosité dans la scène où même les modèles d'arrière-plan les plus sophistiqués ne peuvent les gérer.



FIGURE 6.7 – Changement soudain de la lumière. (a) La scène avant la variation soudaine de lumière (b). La même scène après un changement brusque de lumière.

La figure 6.7 montre des échantillons d'images avant et après un changement soudain de lumière. Notre méthode de détection de changement de lumière utilise le fait qu'un changement soudain global de la lumière provoque une fausse classification de près de 50 % des pixels de la scène où ils seront classés comme faisant partie des objets en mouvement. Cependant, dans certaines situations, où les objets déplacent très près de la caméra, cette hypothèse devient trop simpliste et échoue.

Ainsi, dans le but de distinguer un changement de lumière globale pour les mouvements étendus de l'objet, nous avons exploité le fait que dans le changement soudain global de lumière, les modifications apportées sur la topologie des contours des objets dans la scène sont minimales et que les frontières des régions détectées ne sont pas réelles alors que dans le cas d'un mouvement étendu, des objets, les frontières des régions détectées sont réelles dans l'image. Afin de vérifier si les frontières des régions détectées correspondent aux frontières réelles dans l'image actuelle, nous avons utilisé les gradients de l'image actuelle et de l'image de l'arrière-plan.

L'extraction des gradients s'effectue en prenant la différence de la luminosité entre des pixels consécutifs dans les deux directions horizontale et verticale. L'extraction des gradients est appliquée à la fois sur l'image de l'arrière-plan et sur l'image en cours de traitement. Après l'extraction des gradients, un seuil est appliqué et le résultat est converti en image binaire.

Nous allons ensuite calculer la différence entre l'image de l'arrière-plan et l'image en cours de traitement pour extraire les contours des objets en mouvement. Une érosion est appliquée sur les objets en mouvements de l'extérieur vers l'intérieur. Si les régions de l'avant-plan sont plus petites par rapport à l'image originale, alors la scène a forcément subi un changement global de lumière. Dans ce cas, le modèle de l'arrière-plan est réinitialisé avec l'image courante et les 10 frames précédents. Les ondelettes sont une autre

alternative qui peut également être utilisée à la place des gradients pour la détection d'un changement global soudain de la lumière.

6.8.2 Localisation des objets en mouvement

Après avoir détecté les objets en mouvement et appliqué les opérations de post-traitement pour éliminer les bruits, les pixels classés comme avant plan sont filtrés, regroupés en composantes connexes et étiquetés.

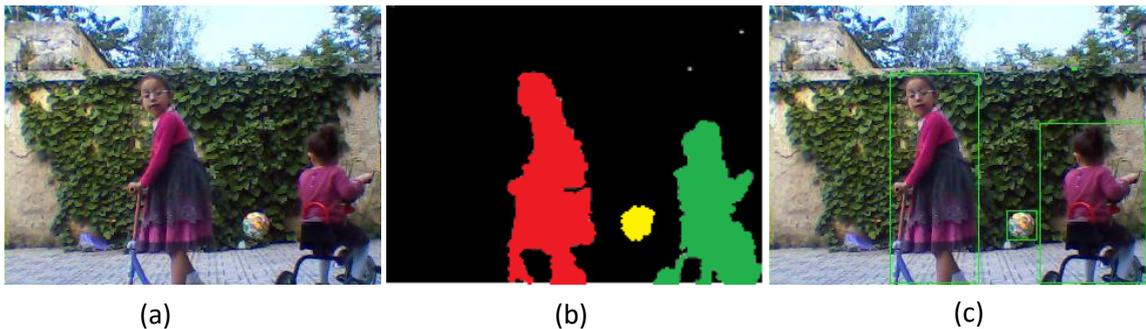


FIGURE 6.8 – Opération d'étiquetage et de cadrage des composantes connexes. (a) l'image originale, (b) détection et étiquetage des composantes connexes, (c) cadrage des objets en mouvement

En analysant les histogrammes de projection horizontale et verticale, les objets sont localisés et cadrés selon un processus itératif. À chaque étape, l'image est segmentée en partie suivant le nombre de pics entre deux minimums d'un histogramme. Ce processus est appliqué successivement à chaque partie jusqu'à ce que cette dernière ne puisse plus être segmentée. Chaque pic entre deux minimums représentera un objet en mouvement.

6.9 Conclusion

Nous avons proposé dans ce chapitre, la construction d'un modèle d'extraction de l'arrière-plan pour la détection des objets mobiles dans des séquences d'images extraites à partir d'une caméra fixe. Ce modèle est basé sur un travail collaboratif entre les mixtures de gaussiennes et les observateurs de l'arrière-plan. Nous avons d'abord fait un passage de l'espace de couleur RGB vers l'espace HSV. Cette conversion a permis de canaliser la luminosité dans le seul composant (V).

Pour surmonter le problème de la variation local, nous avons divisé l'image en N régions et attribué à chacune d'elle un observateur d'arrière-plan. Ce dernier permet de sélectionner uniquement les régions qui ont subi un très grand changement en se basant sur l'analyse des histogrammes de couleur. La conversion vers l'espace de couleur HSV a permis de diminuer considérablement l'effet de la lumière sur le comportement du système. Tandis que la segmentation de l'image en régions et l'utilisation des observateurs

d'arrière-plan ont éliminé les variations locales causées principalement par la présence de poussière.

La GMM est une méthode basée sur la variation indépendante de chaque pixel. Cela signifie que toute variation dans la valeur du pixel va influencer directement sur le comportement des GMM. La présence de quelques pixels perturbateurs séparés ne génère aucun problème, car ils seront facilement supprimés dans le post-traitement. Toutefois, le nuage de pixels généré par les variations locales va induire le système en erreur. Parce qu'il peut être considéré, même après les étapes de post-traitement, comme un objet en mouvement.

Afin de raffiner davantage les résultats de l'extraction de l'arrière-plan, nous avons appliqué plusieurs post-traitements. Les post-traitements proposés respectent les contraintes de temps réel et de portabilité. Nous avons proposé des solutions qui minimisent au maximum la complexité algorithmique et l'espace mémoire nécessaire pour l'application de ces derniers. Une suite d'opérations morpho-mathématiques ont permis de réduire les bruits causés par : la présence de poussière, quelques effets indésirables de luminosité et l'existence de couleurs similaire entre l'objet et l'arrière-plan. Les bruits générés principalement par la poussière ont laissé plusieurs pixels isolés lors de l'extraction de l'arrière-plan, tandis que la ressemblance de couleur a généré des creux au niveau des objets détectés.

Pour prendre en compte les variations soudaines de lumière, nous avons proposé une détection basée sur les gradients. La détection et la suppression de ces variations s'effectuent par une différence entre l'image du modèle et l'image en cours de traitement.

La détection des objets en mouvement n'est rien d'autre que l'étape initiale d'un long processus dans les systèmes de vidéosurveillance. Le chapitre suivant sera consacré à un post traitement spécifique à avoir la suppression des ombres, qui nécessite une intension à part entière.

7

Analyse Spectrale Détection et Suppression des Ombres

7.1 Introduction

Les systèmes de vision par ordinateur dédiés au traitement de la vidéo nécessitent une procédure primordiale à savoir la détection et le suivi des objets en mouvement. Lorsque les objets d'intérêt ont une forme bien définie, plusieurs classificateurs sophistiqués peuvent être utilisés pour segmenter les objets directement à partir de l'image. Ces techniques fonctionnent bien pour des objets avec des contours bien définis, mais sont difficiles à mettre en œuvre pour les objets à contour flexible.

7.2 Motivations

Les mixtures de gaussiennes (GMM) sont parmi les approches les plus communément utilisées pour la détection des objets en mouvement dans une séquence vidéo[FAR 15a]. Cependant, les GMM ont généralement un inconvénient majeur, les ombres ont tendance à être classées comme faisant partie du premier plan conduisant ainsi à une confusion entre l'objet et son ombre. Effectivement, les ombres partagent les mêmes habitudes de déplacement et ont une ampleur similaire au changement d'intensité généré des objets mobiles, ce qui influence sur le rendement des systèmes de vidéosurveillance.

7.3 Taxonomie

Dans les dernières années, de nombreux travaux ont été publiés pour résoudre le problème de la détection et de l'élimination des ombres. Les travaux réalisés peuvent être catégorisés selon l'algorithme utilisé [PRA 03b]; la relation entre les algorithmes, les objets et les environnements traités [ALN 12a]; ou selon le choix des caractéristiques utilisées [SAN 12].

Prati et al. [PRA 03b] ont essayé de regrouper les travaux réalisés dans une taxonomie basée sur deux classes principales : les algorithmes statistiques avec les deux versions paramétriques [WAN 11] et non paramétriques [IRI 08], et les algorithmes déterministes avec les deux versions basées sur un modèle [MCF 11] et non basés sur un modèle [AHM 11]. Il a réalisé aussi plusieurs études comparatives en implémentant un algorithme de chaque méthode.

Une vision différente a été proposée par Al-Najdawi et al. [ALN 12a] pour catégoriser les contributions selon une nouvelle taxonomie. Cette dernière est basée sur la dépendance des méthodes par rapport à l'objet étudié [ZHU 10; CHE 10a; HAF 10; AS 07; SUN 10], par rapport à l'environnement utilisé [SOF 08; GAL 10; AMI 13] ou par rapport au deux [Bian 2012, Ahmad 2011][JIA 12; AHM 11].

Andres Sanin et al. [SAN 12] ont regroupé les systèmes de détection et d'élimination d'ombre sous un autre angle. Ils ont observé que le choix des caractéristiques a une très grande influence sur les résultats par rapport au choix des algorithmes. En se basant sur cette observation, ils ont classé les travaux selon des caractéristiques spectrales tels que : l'intensité [GAL 10; YU 10a], la chromaticité [WAN 11; XUE 08; AHM 11; YAN 12] ainsi que les propriétés physiques [MAK 11]; et selon des caractéristiques spatiales telles que : la géométrie [PAN 13; XU 12; FAN 08] et les textures [LEO 07; LEO 06].

7.4 Évaluation générale et commentaires

L'importance apportée à ce domaine a incité les chercheurs à fournir une grande quantité d'approches afin de résoudre le problème de l'ombre dans les vidéos. Cependant, les approches proposées donnent des résultats uniquement dans des environnements très particuliers et bien définis. En plus, les conditions posées par les auteurs pour le bon fonctionnement de leurs systèmes paralysent totalement ces derniers pour un usage public.

Les études comparatives effectuées dans la littérature ont montré que les résultats obtenus par les caractéristiques spatiales ont une qualité largement supérieure aux résultats obtenus par des caractéristiques spectrales. Toutefois, les caractéristiques spatiales sont trop gourmandes en temps de calcul et en espace mémoire limitant leur utilisation en temps réel et sur des machines à faible puissance. Inversement, les caractéristiques spectrales offrent une très grande rapidité d'exécution, mais présente l'inconvénient d'être

sensible au changement d'intensité lumineuse et donnent de mauvais résultats lorsque les objets ont une intensité ou une couleur similaire à celle d'une ombre.

Afin de contourner les problèmes présentés précédemment, nous allons proposer dans ce chapitre, une méthode pour la détection et l'élimination des ombres à partir des séquences de vidéosurveillance en se basant sur les propriétés chromatiques dans les trois espaces de couleur RGB, HSV et YUV.

7.5 Architecture proposée

La capacité d'extraire des objets en mouvement à partir d'une séquence vidéo est un problème crucial dans de nombreux systèmes de vidéosurveillance. Le rôle principal des opérations de traitement d'image dans un tel système n'est pas la détection précise des détails de l'objet, mais la détection robuste des formes d'objets en mouvement. Malheureusement, ces formes sont généralement déformées par leurs propres ombres. C'est pour ça que nous allons essayer, dans ce chapitre, de détecter et de supprimer les ombres générées par une ou plusieurs sources d'éclairage[FAR 15b].

Dans les scènes dynamiques, tous les points mobiles des objets ou des ombres sont détectés en même temps, les ombres et les objets partagent les mêmes caractéristiques visuelles de mouvement et de détectabilité. Pour ces raisons, nous allons proposer une méthode simple et fiable pour identifier les ombres et les différencier par rapport aux objets. La figure suivante illustre l'architecture globale de notre système.

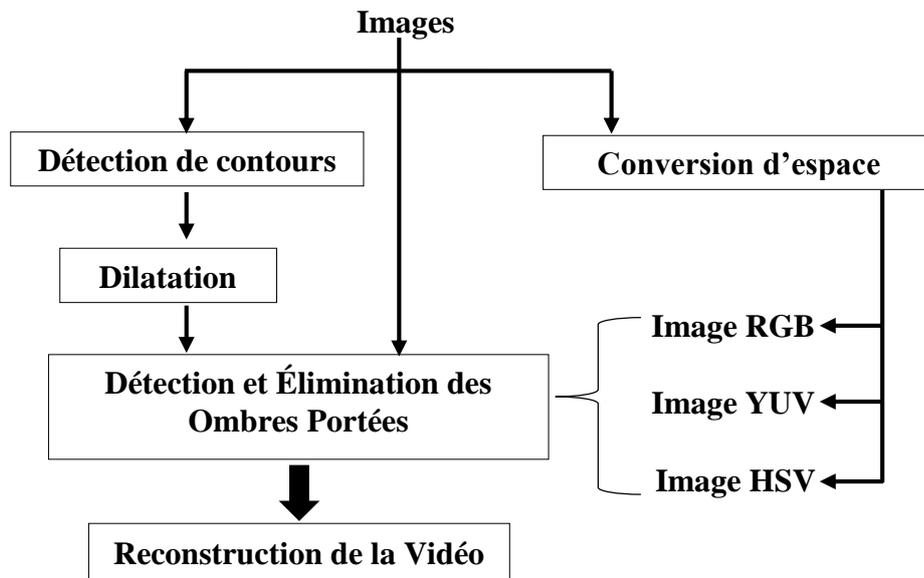


FIGURE 7.1 – Architecture du système proposé

7.6 Détection de contours : Filtre de Canny

Afin de simplifier l'élimination des ombres, nous avons ajouté une étape de détection des contours. Cette dernière va segmenter l'image en un ensemble de composantes connexes permettant ainsi de séparer entre l'ombre propre et l'ombre portée. Après plusieurs tests effectués pour sélectionner le meilleur filtre pour la détection de contour, nous avons choisi le filtre de Canny pour sa simplicité d'implémentation et pour son efficacité par rapport aux autres filtres. En effet, le filtre de Canny présente l'avantage d'être un bon détecteur, car le résultat ne contient qu'un faible taux d'erreur dans la signalisation des contours[FAR 15a]. En plus, la distance entre les contours détectés et les contours réels est très minime ce qui fait de lui un bon estimateur de localisation[FAR 15a]. En remarque également que l'absence des faux positifs rend le filtre très clair à cause de la faible multiplicité des réponses.

7.6.1 Principe général

La méthode de canny implémente une estimation du gradient de l'image à l'aide du filtre de Sobel, suivi d'un seuillage par hystérésis du module de gradient. Un seuil haut et un seuil bas sont à définir. Tous les pixels où le module du gradient est supérieur au premier seuil sont classifiés comme appartenant aux contours de l'image, des contours de l'image sont ainsi formés. Les pixels ayant un module supérieur au seuil bas qui sont reliés aux segments précédents sont définis comme points de contour dans l'image binaire résultante.

7.6.2 Les étapes de l'algorithme de Canny

Le filtre de Canny est un filtre impair à réponse impulsionnelle finie qui suit le processus suivant :

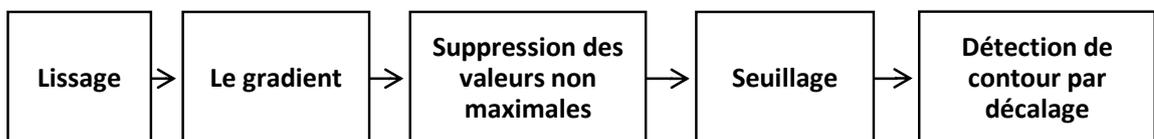


FIGURE 7.2 – Processus de Canny

7.6.2.1 Le lissage

Le lissage est un filtre linéaire passe-bas numérique qui permet d'atténuer les bruits. En effet, le bruit sur une image est inévitable et a plusieurs sources. Ce bruit peut fausser

le résultat d'une détection de contour et causer la détection des lignes ne faisant pas partie du contour. L'image est donc lissée grâce à un filtre gaussien.

$$g(m, n) = G_{\sigma}(m, n) \cdot f(m, n) \quad (7.1)$$

avec

$$G_{\sigma} = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{m^2 + n^2}{2\sigma^2}\right) \quad (7.2)$$

7.6.2.2 Le gradient

Les contours sont exprimés par les plus grandes valeurs d'un gradient. Canny utilise les gradients pour rechercher les zones où la variation de l'intensité lumineuse est la plus importante. Dans un premier temps l'estimation du gradient horizontal et vertical est appliquée. La norme du gradient est donc la distance Euclidienne obtenue comme suit :

$$M(m, n) = \sqrt{g_m^2(m, n) + g_n^2(m, n)} \quad (7.3)$$

avec

$$g_m(m, n) = \frac{\partial g_f(m, n)}{\partial m} \quad \text{et} \quad g_n(m, n) = \frac{\partial g_f(m, n)}{\partial n} \quad (7.4)$$

Ensuite la direction du vecteur doit être définie grâce à l'équation 7.5.

$$\theta(m, n) = \tan^{-1} \left[\frac{g_n(m, n)}{g_m(m, n)} \right] \quad (7.5)$$

7.6.2.3 Suppression des valeurs non maximales

Seuls les extremums locaux sont considérés comme des contours. Les huit pixels voisins sont utilisés pour définir et choisir les extremums locaux.

$$M_T(m, n) = \begin{cases} M(m, n) & \text{si } M(m, n) > T \\ 0 & \text{sinon} \end{cases} \quad (7.6)$$

Où T est un seuil choisi de façon à ce que tous les bruits sont supprimés en gardant bien sûr les contours des objets

7.6.2.4 Seuillage

Les contours potentiels sont définis grâce à un premier seuillage M_T . L'image actuelle ne contient que des contours potentiels après la suppression des valeurs non-maximales. Cependant, ces contours peuvent être dus au bruit restant ou encore aux variations de

couleur. Le moyen le plus rapide et le plus efficace pour supprimer ces faux contours est d'utiliser un double seuillage.

- Tous les pixels supérieurs au seuil maximal défini sont des contours forts.
- Tous les pixels inférieurs au seuil minimal défini sont supprimés.
- Tous les pixels situés entre ces deux valeurs sont considérés comme des contours faibles.

La prochaine étape consistera à définir dans cette 3^{ème} catégorie de contours faibles, ceux que nous devons supprimer.

7.6.2.5 Détection de contour par décalage

Les contours sont détectés par la suppression de tous les contours qui ne sont pas liés au moins à un contour très visible (contour certain). Durant cette étape, tous les contours forts définis lors du double seuillage sont ajustés à l'image finale. Ensuite, une étape de vérification est appliquée en parcourant un à un les contours faibles. Si un contour faible est relié à un contour fort, il est ajouté à l'image finale, sinon le contour faible est supprimé.

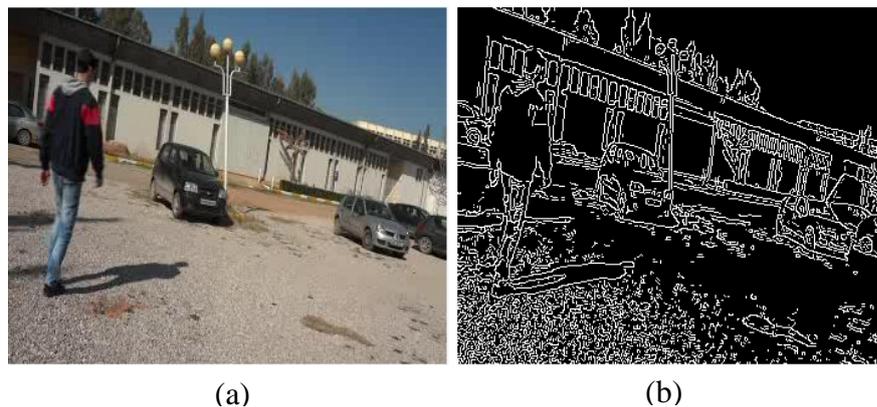


FIGURE 7.3 – Résultats de la détection de contours. (a) l'image originale (b) Masque des contours obtenus avec le filtre de Canny

7.7 Dilatation

La dilatation est une transformation morphologique qui modifie la valeur d'un pixel de l'image en fonction de la valeur de ses voisins. Pour cela, on utilise un ensemble de masques binaires appelés les éléments structurants pour prendre en compte le voisinage du pixel. Ces masques sont spécifiques à l'opération de dilatation ou elle permet de connecter les régions proches et bouche les trous à l'intérieur de celles-ci. Nous avons appliqué la dilatation pour augmenter la taille des contours et supprimer d'éventuelles discontinuités.

7.8 Conversion de l'espace de couleur

On essaye toujours, dans les systèmes utilisant la chromaticité comme un critère de détection et d'élimination des ombres, de choisir un espace de couleur qui permet une séparation naturelle de l'intensité et de la chromaticité. On trouve dans la littérature plusieurs espaces colorimétriques qui permettent de modéliser séparément la luminosité et la chromaticité. Les espaces de couleur HSV, HSL et YUV sont révélés d'être les plus robustes pour la détection d'ombre. Dans ce papier, nous avons choisi une approche dans laquelle nous utilisons les trois espaces de couleurs RGB, HSV et YUV. Chaque espace de couleur est présenté afin de mieux comprendre son effet sur l'ombre.

L'espace RGB RGB est l'espace vectoriel engendré par les trois composantes primaires Rouge, Vert et Bleu. L'ensemble des couleurs produites se représente comme l'intérieur d'un cube. Donc, c'est l'espace de couleur de base avec lequel on peut passer facilement d'un espace à un autre, mais ce dernier ne permet pas d'expliquer l'influence de la luminosité et de la saturation vu que leurs changements affectent les trois composants de base de manière aléatoire.

L'espace HSV Le principe de l'espace HSV est de caractériser les couleurs de façon plus intuitive, conformément à la perception naturelle des couleurs. La teinte (H) est le nom utilisé pour désigner la couleur, qui est idéalement associé à une longueur d'onde. La saturation (S) est le taux de pureté de la couleur, qui doit varier entre la pureté maximale (couleur éclatante) et l'achromatisme (niveau de gris). La valeur (V) est la mesure de l'intensité lumineuse de la couleur, qui doit varier entre le noir absolu et le blanc.

Soit :

$$\begin{aligned} \hat{R} &= R/255, \hat{G} = G/255, \hat{B} = B/255, \\ C_{max} &= \max(\hat{R}, \hat{G}, \hat{B}), C_{min} = \min(\hat{R}, \hat{G}, \hat{B}) \text{ et } \Delta = C_{max} - C_{min} \end{aligned}$$

$$h = \begin{cases} 0 & \text{si } \Delta = 0 \\ 60^\circ \times \left(\frac{\hat{G} - \hat{B}}{\Delta} \bmod 6 \right) & \text{si } C_{max} = \hat{R} \\ 60^\circ \times \left(\frac{\hat{B} - \hat{R}}{\Delta} + 2 \right) & \text{si } C_{max} = \hat{G} \\ 60^\circ \times \left(\frac{\hat{R} - \hat{G}}{\Delta} + 4 \right) & \text{si } C_{max} = \hat{B} \end{cases} \quad (7.7)$$

$$s = \begin{cases} 0 & \text{si } C_{max} = 0 \\ \frac{\Delta}{C_{max}} & \text{sinon} \end{cases} \quad v = C_{max}$$

l'espace YUV Le principe de l'espace YUV est de représenter les couleurs en utilisant une composante de luminance Y, et deux composantes de chrominance U et V, ces composantes de chrominance (u et v) représentent respectivement le contraste Bleu/Jaune et le contraste Rouge/Cyan, la composante de luminance est une moyenne pondérée par la sensibilité humaine relative des primaires. Donc, l'espace de couleur YUV offre une séparation naturelle entre la chromaticité et la luminosité.

$$Y = 0.299 \times R + 0.587 \times G + 0.114 \times B$$

$$U = 0.492 \times (B - Y) \tag{7.8}$$

$$V = 0.877 \times (R - Y)$$

7.9 Détection et élimination des ombres

Nous avons effectué plusieurs recherches et expérimentations pour essayer de comprendre l'impact de l'ombre sur les couleurs[FAR 15b]. Les résultats obtenus nous ont ramené aux remarques suivantes :

- La densité des ombres est parmi les caractéristiques les plus pertinentes et en même temps les plus difficiles à modéliser
- L'ombre dépend fortement de la quantité de lumière réfléchiée par la surface sur laquelle l'ombre est projetée.
- Dans l'espace de couleur HSV, l'ombre ne change pas la teinte d'une couleur, mais affecte seulement la saturation et la luminosité par une augmentation des valeurs. Si nous prenons par exemple un objet de couleur rouge, l'effet de l'ombre sur l'objet ne permet pas de changer sa couleur, l'objet reste toujours de couleur rouge.
- Le passage obligatoire de l'espace RGB vers l'espace HSV ou vers l'espace YUV influence sur l'effet de l'ombre sur les composantes H, S et U, V supposées invariantes par rapport au changement d'intensité lumineuse, ce qui rend la tâche encore plus difficile pour modéliser l'effet de l'ombre sur les couleurs.
- L'utilisation d'un seul espace de couleur n'est pas suffisante pour exprimer tous les effets de l'ombre sur la couleur.
- La valeur Y de l'espace YUV est la meilleure mesure directe de l'intensité lumineuse pour la modélisation de l'ombre.
- Les effets de l'ombre sur les objets provoquent un changement aléatoire, mais avec un effet identique sur les trois composant de l'espace de couleur RGB. Cette condition est nécessaire, mais non suffisante pour déduire la présence d'une ombre sur la couleur.

Après plusieurs expérimentations, nous avons réussi de manière empirique à exprimer l'effet de l'ombre sur la couleur. Après avoir appliqué tous les post-traitements cités pré-

cédemment, nous allons calculer pour chaque nouveau frame de la vidéo, le degré de changement entre l'ancienne et la nouvelle valeur de chaque pixel de l'image dans les trois espaces de couleur RGB, HSV et YUV. Ce processus est garanti par le calcul du taux de variation entre les valeurs des pixels dans le modèle de l'arrière-plan (BG) et les valeurs des pixels du frame courant (F) à l'aide des équations empiriques suivantes. Le changement de rapport des composantes R, G et B dans l'espace de couleur RGB est décrit par les équations suivantes :

$$\begin{aligned} RC_R &= (R_{BG} - R_F) / R_{BG} \\ RC_G &= (G_{BG} - G_F) / G_{BG} \\ RC_B &= (B_{BG} - B_F) / B_{BG} \end{aligned} \quad (7.9)$$

Où $(R_{BG}, G_{BG}, B_{BG}), (R_F, G_F, B_F)$ exprime les valeurs des pixels pour les couleurs Rouge, Vert et Bleu dans l'espace de couleur RGB. Le changement de rapport de la teinte et de la saturation dans l'espace de couleur HSV est décrit par les équations suivantes :

$$\begin{aligned} RC_H &= H_F - H_{BG} \\ RC_S &= S_F - S_{BG} \end{aligned} \quad (7.10)$$

Où $(H_{BG}, S_{BG}), (H_F, S_F)$ et représentent les valeurs des pixels pour les composantes Teinte et Saturation dans l'espace de couleur HSV. Le changement de rapport de la luminosité dans l'espace de couleur YUV est décrit par l'équation suivante :

$$RC_Y = (Y_{BG} - Y_F) / Y_{BG} \quad (7.11)$$

Où Y_{BG}, Y_F représentent les valeurs des pixels pour la composante Y dans l'espace de couleur YUV. La valeur Y de l'espace YUV est la mesure directe de l'intensité de la lumière. L'application d'une ombre sur les pixels diminue leurs valeurs. D'après les propriétés chromatiques, lorsqu'une ombre est appliquée sur le fond d'un pixel elle n'affecte pas sa teinte [FAR 15b]. En plus, les ombres ont tendance à baisser la saturation des pixels. Par conséquent, un pixel est considéré comme faisant partie de l'ombre si les conditions exprimées par les équations 7.12, 7.13 et 7.14 suivantes sont satisfaites :

$$|RC_R - RC_G| < 3 \text{ et } |RC_B - RC_G| < 3 \text{ et } |RC_R - RC_B| < 3 \quad (7.12)$$

$$RC_H < P_1 \text{ et } RC_S < P_2 \quad (7.13)$$

$$0 < RC_Y < |RC_H - RC_S| \quad (7.14)$$

Où P_1, P_2 représentent des seuils optimisés de façon empirique.

7.10 Conclusion

Nous avons proposé, dans ce chapitre, une méthode efficace basée sur les espaces de couleurs RGB, HSV et YUV pour détecter et supprimer les ombres portées générées par les objets mobiles. Le système est capable d'opérer à la fois dans les environnements intérieurs et extérieurs et sans aucune hypothèse préalable sur les conditions d'éclairage. Dans notre méthode, nous avons appliqué le filtre de Canny pour détecter les limites entre les objets et les ombres. Ces informations sont utilisées comme critère pour distinguer entre les ombres portées et les ombres attachées. Nous avons utilisé trois critères calculés dans les espaces RGB, HSV et YUV pour décider si les pixels font partie des ombres ou des objets et. Les pixels qui ont satisfait les trois critères sont classés en tant qu'ombre et supprimés de l'image par une mise à zéro.

8

Analyse comportementale Reconnaissance, suivi et interprétation

8.1 Introduction

Le suivi d'objets est une étape fondamentale dans les systèmes de vidéosurveillance puisqu'il est la base de l'analyse de trajectoires, de comportements et de reconnaissances d'activités. Dans cette étape on essaye de résoudre les problèmes liés au suivi par une association correcte des objets détectés. Historiquement, il s'agit d'un problème d'association de données, dont les premières applications étaient essentiellement l'analyse de données fournies par les radars ou les sonars. Les cibles représentent les objets d'intérêts, tandis que les observations représentent les régions en mouvement obtenues en sortie du module de détection de mouvement. La classification et la reconnaissance viennent ensuite pour valider les estimations des trajectoires proposées par le suivi.

Nous avons décrit dans les chapitres précédents la procédure de soustraction d'arrière-plan appliquée à une vidéo pour extraire les régions en mouvement. Le résultat est un masque binaire dans lequel les pixels de valeurs nulles sont représentatifs de l'arrière-plan, tandis que les pixels de valeurs non nulles sont représentatifs d'un ou plusieurs objets en mouvement. Ce chapitre présente l'étape suivante du traitement qui consiste à extraire, classer et suivre les objets de la scène. Il s'agit de construire et maintenir un ensemble de caractéristiques pour chaque objet de la scène surveillée. La construction de cet ensemble s'effectue par le biais de trois grandes étapes à savoir le suivi, la classification

et l'apprentissage.

8.2 Suivi des objets en mouvement

Le suivi est une tâche qui permet d'estimer les mouvements d'un objet. Les Traqueurs supposent généralement que l'objet est visible tout au long de la séquence. Plusieurs représentations des objets ont été utilisées dans l'état de l'art. Parmi ces représentations, nous citons : les approches basées sur un point [LUC 81 ; SHI 94 ; SAN 08], les approches qui utilisent les modèles articulés [WAN 03 ; RAM 07 ; BUE 08], les méthodes basées contour [BIR 98 ; ISA 98 ; BIB 10] et finalement les méthodes qui utilisent le flot optique pour estimer les mouvements des objets.

Nous allons nous intéresser dans ce chapitre aux méthodes qui représentent l'objet par leurs formes géométriques. L'estimation de leur mouvement est basée sur les informations contenues dans les frames successifs appelés suivis frame par frame. Le suivi basé modèle est parmi les approches les plus efficaces dans ce type de scénario. L'objet étant décrit par un modèle cible qui représente généralement une partie de l'image ou l'utilisation d'un histogramme de couleur. Tandis que le mouvement est défini en tant qu'une transformation qui permet de minimiser le degré de dissemblance entre le modèle cible et le bloc candidat.

Le suivi basé sur un modèle peut être effectué soit d'une manière statique [COM 03] ou adaptative [LUC 81]. Dans le premier cas, le modèle cible ne change pas de caractéristiques à travers le temps. Le modèle adaptatif est utilisé dans le cas où le modèle cible est reconstruit à partir des frames précédents. On trouve également des approches hybrides qui utilisent les modèles statiques et les modèles adaptatifs pour le suivi des objets en mouvement [MAT 04 ; DOW 05 ; RAH 08]. D'autres types d'approches utilisent un processus qui permet une reconnaissance partielle pour les parties, de l'objet, dites fiables du modèle tel que le visage [JEP 03 ; ADA 06]. Les approches basées sur un modèle ont limité les capacités de représentation des objets, car ils permettent de modéliser qu'un seul aspect de l'objet. Pour modéliser plus de variations d'apparence, les modèles génératifs ont été proposés. Les modèles génératifs sont construits soit en mode hors ligne [BLA 96] ou lors de l'exécution [ROS 08 ; KWO 10] du suivi.

Les modèles de suivi génératif modélisent uniquement l'apparence de l'objet. Cette restriction échoue, dans la plupart des cas, dans le suivi dans les environnements encombrés. Afin de remédier à ce problème, les méthodes de suivi actuelles offrent également la possibilité de modéliser les environnements dans lesquels les objets se déplacent. On trouve dans la littérature deux variantes pour la modélisation de l'environnement. La première méthode essaye de chercher l'environnement qui est en corrélation avec l'objet d'intérêt et qui permet de supporter le mouvement de l'objet [YAN 09 ; GRA 10]. Ces supports sont utilisés pour aider le système de suivi à retrouver les objets disparus de la

caméra ou lorsque les objets suivis subissent des transformations difficiles à modéliser.

Dans la seconde variante, l'environnement est considéré comme une classe négative contre laquelle le système doit effectuer une discrimination. L'idée la plus intuitive est de construire un classificateur binaire qui représente la limite de décision entre l'objet et son arrière-plan. Les modèles discriminatifs statiques [AVI 04] effectuent l'apprentissage des classifieurs avant l'exécution du suivi. Cette démarche peut être utilisée uniquement dans les applications qui ont une connaissance a priori parfaite des objets en mouvement. Les modèles dits adaptatifs [COL 05 ; AVI 07 ; GRA 06 ; BAB 09] effectuent une construction et un apprentissage en ligne des classifieurs.

La phase principale dans ce type de méthode est la mise à jour des paramètres. En effet, le voisinage de l'emplacement actuel est utilisé comme un échantillon positif pour l'apprentissage du classifieur tandis que les régions qui se trouvent autour du voisinage sont utilisées comme échantillon d'apprentissage négatif. Ce processus est utilisé pour la mise à jour des paramètres du classifieur et cela pour chaque nouveau frame. Cette stratégie de mise à jour a démontré son efficacité face au grand changement d'apparence, aux occlusions à court terme et par rapport aux environnements relativement encombrés. Cependant, ces méthodes souffrent également de quelques défauts relatifs au mouvement et à la présence de l'objet. En effet, lorsque les mouvements d'un objet se caractérisent par une vitesse lente en continu d'un endroit à l'autre, le classifieur converge vers l'apparence actuelle et efface totalement tous les modèles appris précédemment. Ce cas de figure est similaire à un apprentissage avec un seul exemple pendant plusieurs itérations (sur apprentissage). Le deuxième défaut se présente lorsque l'objet quitte la scène pour une période plus longue que prévu. Pour résoudre ces problèmes, la mise à jour du classifieur a été conditionnée par un autre classifieur auxiliaire dans lequel l'apprentissage a été effectué dans le premier frame [GRA 06] ou par l'apprentissage d'une paire de classifieurs totalement indépendant [TAN 07b ; YU 08].

8.3 Classification et reconnaissance des formes

8.3.1 Détection des formes

La détection des objets est une tâche qui permet de localiser des objets parmi d'autres dans une image. La définition d'un objet varie selon la nature et le nombre d'instances de ce dernier. En effet, un objet peut être une seule instance ou une classe entière d'objets.

Dans la littérature, les méthodes de détection des objets se basent généralement sur les caractéristiques locales de l'image ou sur un ensemble de caractéristiques récupérées à partir d'une fenêtre temporelle glissante [VIO 01]. Les approches basées sur les caractéristiques utilisent généralement un processus de trois étapes à savoir la détection, la reconnaissance et enfin l'appariement de modèle. Les caractéristiques de la surface [LOW 04 ;

LEP 05] et la modélisation 3D [VAC 04] ont également été largement exploitées.

Ces algorithmes ont atteint un très bon niveau de maturité et même fonctionné en temps réel sur des appareils à faible puissance de calcul [TAY 09]. Cette complexité très réduite n'a pas diminué les capacités de ces systèmes pour la détection d'un grand nombre d'objets [PIL 10 ; OBD 05]. La détection des caractéristiques de l'image et la contrainte d'avoir une connaissance géométrique a priori des objets représente à la fois la puissance et la faiblesse de ces approches. Les approches basées sur les fenêtres glissantes [VIO 01] utilisent un balayage de l'image d'entrée par une fenêtre de différentes tailles et vérifie pour chaque fenêtre glissante l'existence ou non d'un objet d'intérêt sous-jacent de la zone sélectionnée par la fenêtre. Dans ce type de méthodes, des centaines voire des milliers de zones sont évalués pour chaque frame. Pour obtenir une efficacité en temps réel, les méthodes de détections basées sur les fenêtres glissantes adoptent une architecture en cascade.

Cette architecture en cascade se base sur le fait que l'apparition de l'arrière-plan dans l'image est plus fréquente que l'objet. Pour cela, un classifieur est devisé en plusieurs étages dont chacun permet un rejet précoce d'un segment de l'arrière-plan, réduisant ainsi le nombre d'étapes qui doivent être évaluées en moyenne. Cependant, une telle architecture nécessite un nombre très important des données d'apprentissage et une puissance de calcul non négligeable dans la phase d'apprentissage pour pouvoir représenter avec précision la limite de décision entre l'objet et l'arrière-plan. Une approche alternative utilise plusieurs modèles pour représenter chaque objet et dans ce cas, il suffit d'ajouter des modèles pour chaque objet et pour chaque nouveau frame.

8.3.2 Apprentissage automatique

La plupart des méthodes de classification utilisent un apprentissage supervisé qui consiste à étiqueter a priori toutes les données d'apprentissage. Cette supposition paralyse totalement les systèmes de vision et en particulier la reconnaissance des formes dans une vidéo. En effet, les objets qui peuvent se manifester dans une vidéo peuvent avoir des postures aléatoires et même si la posture principale est connue lors de la première détection, ce dernier peut se mettre dans un aspect totalement inconnu du classifieur vu qu'il a été entraîné de façon hors-ligne. Pour remédier à ce type de problème, un apprentissage semi-supervisé est envisageable en utilisant, en même temps, des échantillons étiquetés et d'autres non étiquetés. Ces méthodes supposent généralement que les données sont indépendantes et distribuées de manière identiquement en utilisant certaines propriétés, tel que les exemples non étiquetés se regroupent naturellement en clusters dans l'espace des caractéristiques [ZHU 09]. On trouve, dans la littérature, plusieurs algorithmes qui reposent sur des hypothèses similaires. Les plus connus sont : Esperance-Maximisation (EM), Auto-apprentissage et Co-apprentissage.

8.3.2.1 Esperance-Maximisation (EM)

C'est une méthode générique pour trouver des estimations des paramètres du modèle fournis par des données non étiquetées. EM est un processus itératif qui, dans le cas de la classification binaire, alterne la surestimation des étiquettes souples des données non étiquetées et l'apprentissage d'un classificateur. L'algorithme EM a été appliqué avec succès dans plusieurs domaines tels que la classification des documents [NIG 00] et l'apprentissage des catégories d'objets [FER 03] et il est parfois considéré comme une sous variante de l'autoapprentissage [ZHU 09]. L'algorithme EM repose sur hypothèse de la séparation des faibles densités, ce qui signifie que les classes sont bien séparées [CHA 06].

8.3.2.2 L'autoapprentissage

C'est un algorithme qui commence par la formation d'un classificateur initial à partir d'un ensemble d'apprentissages étiqueté ; le classificateur est ensuite évalué sur les données non étiquetées. Les exemples avec les réponses les plus confiantes des classifieurs sont ajoutés à l'ensemble d'apprentissages et le classificateur est mise à jour par un nouvel apprentissage. Nous avons remarqué que la classification peut être améliorée davantage si les données non étiquetées ont été sélectionnées par une mesure indépendante plutôt que la confiance attribuée par classifieur.

On trouve, dans l'état de l'art, plusieurs critiques concernant la fiabilité de l'hypothèse basée sur la séparation par une faible densité et que cette dernière ne permet pas d'avoir une reconnaissance performante des formes par rapport à d'autres approches [ROS 05].

8.3.2.3 Le co-apprentissage

C'est une méthode bâtie sur l'idée que des classifieurs indépendants peuvent être entraînés mutuellement. Pour réaliser des classifieurs indépendants, le co-apprentissage suppose l'existence de deux vecteurs de caractéristiques totalement indépendantes. L'apprentissage est effectué par deux classifieurs séparés en utilisant des données d'apprentissage étiqueté. Les deux classifieurs seront ensuite évalués par des données non étiquetées.

Les échantillons étiquetés du premier classificateur sont utilisés pour augmenter l'ensemble d'apprentissages du deuxième classifieur. La même procédure est effectuée pour l'autre classifieur par un processus itératif.

Le co-apprentissage a montré son efficacité dans les problèmes contenant des modalités indépendantes telles que la classification de texte [BLU 98] et les systèmes de reconnaissance biométrique. Cependant, leur utilisation dans les systèmes de vision par machine n'est pas optimale vu l'absence d'une indépendance multimodale. En effet, les données d'apprentissage sont généralement extraites à partir d'une même image (unique modalité) et les caractéristiques extraites à partir de la même modalité sont fortement

dépendantes. Cette dépendance est totalement contradictoire avec les suppositions du co-apprentissage. Malgré ces problèmes, le co-apprentissage a été utilisé pour la détection des voitures [LEV 03] et la reconnaissance des objets dans les systèmes de vidéosurveillances [JAV 05].

8.4 Travaux similaires

On trouve dans l'état de l'art, plusieurs méthodes qui essaient de faire une combinaison entre l'apprentissage, la classification et le suivi. Williams et al. [WIL 05] ont proposé un apprentissage hors ligne pour valider la trajectoire du suivi. Si la validation échoue, un mécanisme de recherche exhaustive est effectué pour trouver la cible. D'autres approches intègrent des classifieurs avec un filtre à particule. Ces techniques ont été appliquées pour le suivi des visages dans des vidéos à faible taux de rafraîchissement [LI 08], le suivi multiple des joueurs de hockey [OKU 04], ou le suivi des piétons [LEI 07 ; BRE 09].

Les méthodes citées précédemment utilisent un apprentissage hors ligne ce qui est totalement contradictoire avec la nature dynamique des objets en mouvement et celle de la vidéo. En contrepartie, les approches adaptatives ont également utilisé des combinaisons entre la classification, l'apprentissage et le suivi [BAB 09 ; GRA 08 ; TAN 07b]. En effet, ces systèmes utilisent un suivi en se basant sur un apprentissage en ligne des classifieurs pour séparer entre les objets en mouvement et l'arrière-plan. Malheureusement, ces méthodes n'ont pas donné des résultats satisfaisants malgré la nature dynamique des modèles proposée. Cette limitation est due à l'utilisation d'un seul processus qui permet de réaliser en même temps toutes les combinaisons.

8.5 Analyse intrinsèque du comportement

Les régions détectées dans l'image courante sont extraites et leurs caractéristiques sont estimées. L'ensemble des régions forme la liste des candidats à l'instant t . Cette liste est comparée à la liste des objets suivis à l'aide du processus de classification décrit dans la section précédente.

Le suivi est un processus cyclique qui se décompose en plusieurs étapes. Cependant, il faut faire la différence entre le processus de suivi comme une étape de la vision qui peut contenir plusieurs aspects et techniques, et le suivi lui-même qui permet d'estimer l'emplacement prochain de l'objet et qui se résume dans l'application d'un algorithme de suivi bien particulier. Le processus de suivi est composé des étapes suivantes :

Détection et extraction des caractéristiques : Création d'une liste des objets candidats de l'image courante en effectuant une extraction des objets en mouvement ainsi que leurs caractéristiques.

Estimation des mouvements : La prédiction détermine la position la plus probable de l'objet suivi dans l'image courante. La prédiction nécessite la connaissance des états de l'objet dans les images précédentes, ou d'un état initial fourni par le processus de détection d'objet. La prédiction est réalisée à l'aide d'un modèle de mouvement prédéfini, dans lequel est incorporé un modèle d'incertitude.

Classification : c'est une mise en correspondance qui consiste à comparer la position prédite d'un objet avec les observations afin d'identifier les correspondances. Seules les associations d'objets dont les caractéristiques sont proches sont conservées.

Apprentissage et mise à jour : la mise à jour et l'estimation de l'état de l'objet terminent le cycle en fournissant une estimation de l'état de l'objet. Cette mise à jour tient compte du modèle de mouvement choisi (prédiction) et des observations réalisées et des décisions prises par le classifieur. Cette étape doit également prendre en compte les ambiguïtés présentes.

8.6 Propositions et Architectures

La combinaison de la classification, de l'apprentissage et du suivi dans un contexte dynamique est une solution qui semble parfaite pour un système de suivi des objets en mouvement. Cependant, malgré les combinaisons réalisées dans l'état de l'art, les approches existantes n'ont pas réussi à tirer le maximum d'avantage de la puissance de cette combinaison. Pour cela, nous avons proposé un système qui permet à la fois de réaliser la classification, l'apprentissage et le suivi des objets en mouvement dans une vidéo de surveillance.

Notre système se distingue par rapport aux autres méthodes proposées par une architecture parallèle qui permettent de réaliser les trois tâches séparément et efficacement pour tirer le meilleur de cette combinaison. En effet, le classifieur est dans un apprentissage permanent ce qui le rend performant par rapport aux changements de posture des objets en mouvement dans la scène. Le classifieur envoie les nouveaux changements de posture ayant un grand degré de confiance comme de nouvelles données d'apprentissage. Cet aspect cyclique oblige le système à s'adapter avec tous les changements de posture possibles. Le suivi étant la principale tâche, tire les avantages acquis et les propositions des classifieurs pour trouver et tracer les chemins utilisés par un objet en mouvement détecté. En cas d'occlusion ou de collusion de deux ou plusieurs objets, le système utilise les informations estimées des trajectoires des objets en mouvement par le module de suivi pour corriger et parfois annuler le processus d'apprentissage. Les résultats attribués par le module d'extraction de l'arrière-plan permet au classifieur de ne pas tomber dans

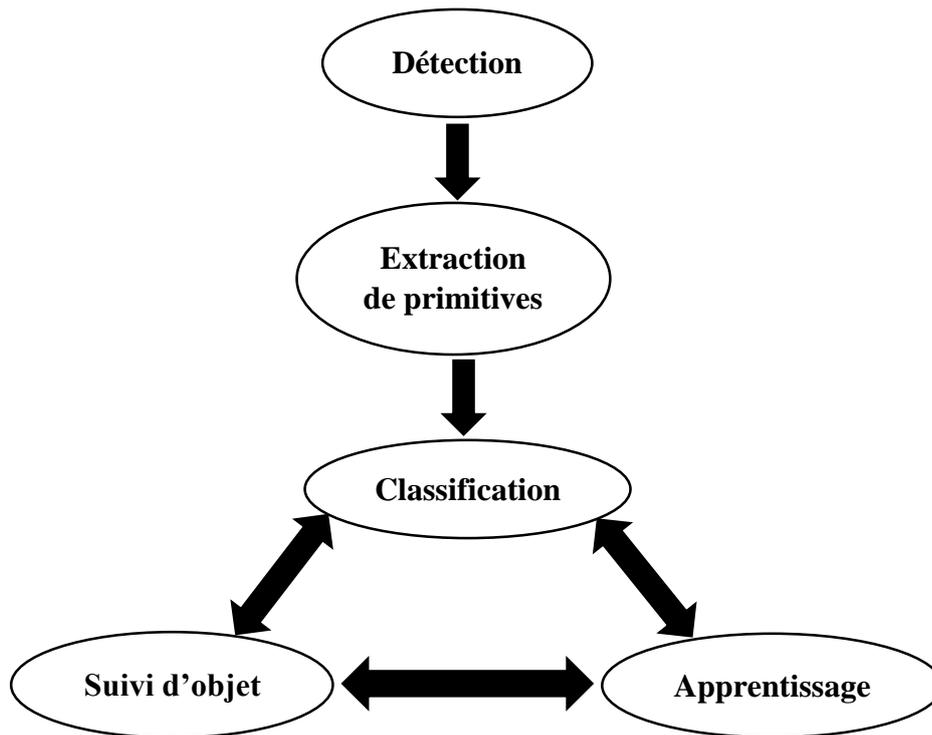


FIGURE 8.1 – Processus dynamique et adaptatif de suivi

une fausse classification ce qui augmente considérablement les capacités du système à s'adapter avec l'environnement.

Dans notre système, nous avons utilisé un algorithme de suivi au niveau de l'objet. Autrement dit, le système ne suit pas les parties d'un objet, telles que les membres d'un être humain, mais suit l'objet dans son ensemble d'un frame à un autre. Les informations extraites par le module de suivi sont adéquates pour la plupart des applications de vidéosurveillance.

La méthode de suivi que nous avons développée est inspirée par les différentes études présentées dans l'état de l'art. Notre approche utilise les caractéristiques de l'objet telles que la taille, le centre de masse, la boîte englobante et la couleur de l'histogramme pour établir une correspondance entre les objets dans les frames consécutifs. En outre, notre algorithme de suivi détecte les occlusions des objets et distingue dans la plupart des cas l'identité des objets après une opération de division des objets occlus. En analysant les informations de trajectoire des objets, notre système de suivi est capable de détecter également les objets qui ont quitté la scène.

8.6.1 Extraction des caractéristiques

Après avoir segmenté l'image en un ensemble de composantes connexes dans l'étape de détection, nous allons maintenant extraire les caractéristiques des objets correspondants à partir de l'image actuelle. Ces caractéristiques sont la taille (S), le centre de gravité

(C_g), l'histogramme de couleur (H_c) et la silhouette du contour de l'objet. Le calcul de la taille de l'objet est une opération triviale effectuée par un simple comptage du nombre de pixels de chaque objet potentiel contenu dans la boîte englobante.

8.6.1.1 Calcul du centre de gravité

Soit $C_g = (X_{C_g}, Y_{C_g})$ le centre de gravité d'un objet O . Le calcul de ce centre s'effectue selon l'équation suivante :

$$X_{C_g} = \frac{\sum_i^n x_i}{n}, \quad Y_{C_g} = \frac{\sum_i^n y_i}{n} \quad (8.1)$$

Ou n représente le nombre de pixels dans l'objet O . Le centre de gravité est une information très importante lorsqu'elle est couplée à d'autres caractéristiques telles que le contour. En effet, la distance entre le centre de gravité et les différents points du contour permet non seulement de reconstruire une géométrie détaillée de l'objet, mais permet également de compléter les informations sur la posture de ce dernier.

8.6.1.2 Calcul de l'histogramme de couleur

L'histogramme de couleur, H_c est calculé sur des valeurs d'intensité monochrome des pixels de l'objet dans l'image actuelle. Nous avons effectué une opération de quantification afin de réduire la complexité des opérations de calcul sur l'histogramme. Soit N , le nombre d'intervalles dans l'histogramme, chaque intervalle va contenir $255/N$ couleur. Le calcul de l'histogramme est effectué de façon itérative sur l'objet O en incrémentant les valeurs de chaque couleur mémorisée correspondant à chaque intervalle de l'histogramme de couleur. En général, les histogrammes ne permettent pas une bonne discrimination vu que plusieurs objets peuvent avoir des couleurs similaires. Cependant, l'architecture du système proposé qui utilise les informations de suivi pour la classification permet de corriger cette inefficacité, car la détection et l'extraction des objets en mouvement ne se basent pas sur les caractéristiques, mais elles sont effectuées dans les étapes précédentes par des mécanismes beaucoup plus sophistiqués.

8.6.1.3 Extraction de la silhouette de l'objet

L'utilisation de la silhouette de l'objet est parmi les caractéristiques les plus utilisées pour modéliser les objets. Cette caractéristique est porteuse à elle seule de nombreuses informations (dimension, allure, etc.) en utilisant uniquement une fine partie des points de l'objet ce qui rend une telle caractéristique très sollicitée pour un usage en temps réel vu que le système traite uniquement les contours de l'objet au lieu de traiter tous les pixels de l'objet. Nous avons utilisé les informations de contour obtenu dans l'étape d'élimination des ombres portées par le biais de l'algorithme de Canny présentée dans le

même chapitre. La réutilisation de cette information a allégé considérablement le temps d'exécution de l'étape de l'extraction des caractéristiques qui prend dans la plupart des systèmes proposés dans la littérature un temps non négligeable par rapport aux autres modules du système.

8.6.2 Classification

La Classification ou reconnaissance de classes est une branche de l'intelligence artificielle qui fait largement appel aux techniques d'apprentissage automatique et aux statistiques. C'est un ensemble de techniques et méthodes visant à identifier des formes à partir de données brutes afin de prendre une décision dépendant de la catégorie attribuée à cette forme.

Les contraintes de vitesse d'exécution liée au système de vidéosurveillance nous ont obligés à utiliser un très simple algorithme de classification pour la reconnaissance et l'identification de l'objet. Nous avons implémenté l'algorithme des K Plus Proches Voisins (cf. chapitre 3). Le nombre de classes varie selon la présence de chaque nouvelle posture détectée. Une suppression est également appliquée lorsqu'une classe n'est plus sollicitée pendant un laps de temps déterminer empiriquement par 100 frames. Effectivement, un objet en mouvement peut carrément sortir de la scène et dans ce cas-là, les informations de ce dernier ne seront plus utilisables et si le même objet réapparaît dans la scène, le système va le considéré comme une nouvelle entité. Cette procédure de suppression permet de réduire l'espace mémoire occupée par le système.

8.6.3 Apprentissage

Nous allons aborder dans cette section l'algorithme d'apprentissage. L'objectif est d'améliorer les performances du système de classification et de reconnaissance en utilisant un apprentissage en ligne. Pour chaque nouveau frame de la vidéo, nous allons évaluer la reconnaissance, le taux d'erreur commis par le classifieur ainsi que le processus de mise à jour pour éviter ces erreurs dans les frames suivants. Pour cela, nous avons utilisé deux types d'erreurs à savoir : les faux positives et les faux négatives. Nous avons également utilisé deux ensembles de données d'apprentissage. L'un est étiqueté et l'autre non étiqueté.

La tâche de l'algorithme proposé est de réaliser l'apprentissage du classifieur par les données étiquetées, ensuite, d'améliorer ses performances avec des données non étiquetées. Les données étiquetées représentent les postures ayant déjà été reconnues par le classifieur tandis que les données non étiquetées représentent les données acquises par le module de suivi. Lorsqu'un objet entre dans la scène pour la première fois, le classifieur n'a aucune classe a priori pour lui. Ce cas est considéré comme une initialisation de classe de l'objet en cours de traitement. Dans le prochain frame, le même objet aura

une représentation quasi similaire avec lequel le classifieur va réaliser une tentative de reconnaissance grâce à la base d'apprentissage

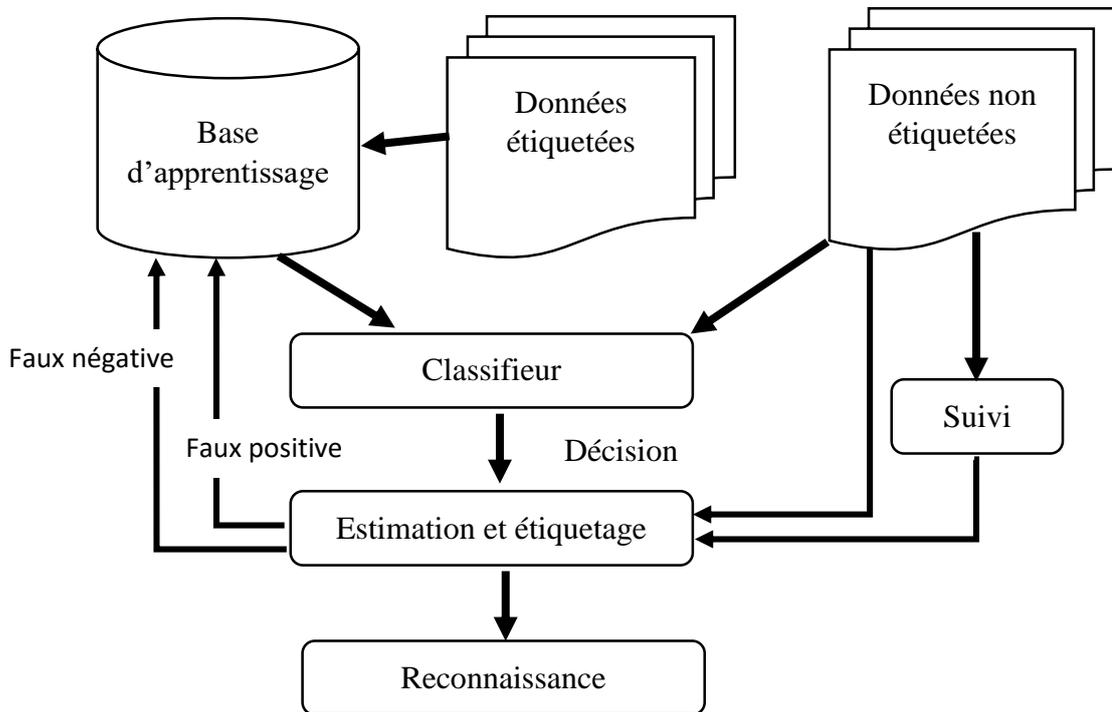


FIGURE 8.2 – Schéma du processus d'apprentissage et de décision

Afin de pouvoir détecter et corriger les erreurs commises par le classifieur, nous avons devisé les données non étiquetées en deux parties selon la classification actuelle. L'idée principale est de réinjecter les données non étiquetées mal classées avec une nouvelle estimation et une étiquette en utilisant les informations obtenues du suivi. L'algorithme étant itératif permet de réadapter le classifieur afin de ne plus commettre les mêmes erreurs dans les frames suivants.

8.6.4 Suivi d'objets

Cette section présente la mise en place du filtrage prédictif(cf. chapitre 1) pour l'estimation de mouvement.

8.6.4.1 Modélisation et initialisation

La configuration d'un objet est représentée par son vecteur d'état contenant sa position actuelle et son vecteur de vitesse. A l'instant t , le vecteur d'état X_t^o pour un objet s'écrit

selon l'équation suivante :

$$X_t = \begin{bmatrix} x_t \\ y_t \\ \bar{x}_t \\ \bar{y}_t \end{bmatrix} = \begin{bmatrix} 1 & 0 & \Delta_t & 0 \\ 0 & 1 & 0 & \Delta_t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_{t-1} \\ y_{t-1} \\ \bar{x}_{t-1} \\ \bar{y}_{t-1} \end{bmatrix} + \Delta_t \cdot \begin{bmatrix} \frac{1}{2}\bar{\bar{x}}_{t-1}\Delta_t \\ \frac{1}{2}\bar{\bar{y}}_{t-1}\Delta_t \\ \bar{\bar{x}}_{t-1} \\ \bar{\bar{y}}_{t-1} \end{bmatrix} \quad (8.2)$$

Où $(x, y), (\bar{x}_t, \bar{y}_t), (\bar{\bar{x}}_t, \bar{\bar{y}}_t)$ représente respectivement les coordonnées actuelles de l'objet, le vecteur de vitesse et le vecteur d'accélération de l'objet. Δ_t représente l'intervalle de temps entre deux prédictions consécutives.

8.6.4.2 Evolution linéaire du vecteur d'état

La description de l'évolution du vecteur d'état est effectuée via l'algorithme de Kalman. Pour cela, un modèle autorégressif de mouvement est défini. L'état estimé d'un objet est une extrapolation linéaire de l'état précédent à laquelle est ajouté un bruit gaussien représentant l'incertitude du modèle. Ceci revient à considérer le vecteur d'accélération $(\bar{\bar{x}}_t, \bar{\bar{y}}_{t-1})$ comme étant un bruit blanc gaussien noté w_t et centré en zéro. L'équation précédente peut être réécrite de la manière suivante :

$$X_t = F_{t-1} \cdot X_{t-1} + W_t \quad (8.3)$$

8.6.4.3 Mesure de l'observation

Les équations de mesures traduisent la relation entre l'observation et le vecteur d'état. Le vecteur d'observation est composé des positions x et y auxquelles est ajouté un bruit de mesure supposé gaussien.

8.6.4.4 Statut du suivi d'un objet

Nous avons défini quatre états pour décrire la configuration d'un objet durant le processus de suivi.

Nouveau : Un objet est assigné à l'état nouveau lorsqu'il apparaît pour la première fois dans la scène. Il peut s'agir d'un véritable objet en mouvement ou d'un faux positif, nous introduisons une mesure de confiance permettant de mesurer la fiabilité de l'objet.

Suivi : L'objet est suivi et est considéré comme étant un véritable objet.

Occulté : Il peut arriver qu'un objet suivi ne soit pas détecté à l'image suivante due à une occlusion ou une fusion avec d'autres objets.

Perdu : Un objet suivi n'a pas été détecté dans la nouvelle image à cause d'une erreur d'extraction de l'arrière-plan ou une sortie totale de la scène.

8.7 Cohérence spatio-temporelle

Plusieurs scénarios peuvent se présenter lors du suivi tel que l'occlusion ou la fusion de deux ou plusieurs objets en mouvement. Les informations temporelles extraites de l'étape d'extraction et de classification sont utilisées pour résoudre ce type d'ambiguïtés. En effet, lorsqu'une ambiguïté se présente un mécanisme implanté dans le système va permettre de conserver les états prédits des objets qui ont eu un problème de suivi. Dans ce cas, le système va suspendre la mise à jour des paramètres de suivi jusqu'à ce que cette situation se débloque soit par la séparation des objets fusionnés ou par une apparition complète de l'objet occulté.

Lorsque plusieurs objets se croisent, le système essaye d'estimer les nouvelles trajectoires en utilisant les résultats de la reconnaissance des postures précédentes. Ceci permet de conserver une cohérence spatiale et temporelle de l'évolution des objets. Le système peut se mettre dans un échec total lorsque les objets changent de posture et de trajectoire en même temps. Dans ce cas, l'opération de suivi est initialisée en affectant l'état « nouveau » à tous les objets en mouvement non reconnu.

8.8 Cohérence structurelle

La performance de la classification des objets dépend fortement de la qualité de l'étape de segmentation. À cause de plusieurs facteurs liés à la nature de l'environnement et au comportement des objets en mouvement dans la scène, le contour des régions détectées ne reflète pas le réel format de la silhouette de l'objet. En effet, un objet peut être obstrué par des objets de premier plan stationnaires ou en raison du fait que seule une partie de l'objet est capturée dans la scène. Dans ce type de cas, l'algorithme de classification échoue pour attribuer la bonne étiquette pour l'objet.

Pour éviter une situation similaire, nous avons utilisé l'analyse des histogrammes de contour. Chaque histogramme est initialisé et maintenu pour un objet O détecté dans la scène. La taille de cet histogramme est égale au nombre de types d'objets différents et chaque bloc de cet histogramme conserve le nombre d'apparitions de l'objet "O" dans ce type de posture. Avec l'aide de ces hypothèses, les types possibles d'un objet peuvent être accumulés sur une période de temps prédéfini et la véritable décision de ce type peut être faite avec plus de précision en sélectionnant le type du bac avec la plus grande valeur que le type de l'objet.

8.9 Conclusion

Nous avons présenté dans ce chapitre les différentes modalités relatives à l'analyse comportementale des objets en mouvement. Nous avons commencé par un petit aperçu sur les méthodes qui ont des points communs avec les algorithmes proposés dans notre système. Nous avons ensuite présenté l'architecture globale du système et essayé de détailler chaque processus.

Pour aboutir à un système de suivi performant, nous avons proposé une combinaison entre la classification, l'apprentissage et le suivi. L'utilisation de cette combinaison a permis de contourner les problèmes les plus difficiles du suivi tels que la gestion des occlusions et la fusion des objets en mouvement.

L'apprentissage en ligne, avec les données étiquetées et les données non étiquetées, a permis d'augmenter la fiabilité du classifieur grâce à un mécanisme adaptatif et dynamique pour la gestion du changement de postures des objets en mouvement. En effet, les estimations des emplacements obtenus par l'algorithme de suivi des objets en mouvement ont permis au classifieur de corriger certaines décisions mal prises et d'utiliser ces informations pour éviter une erreur de classification dans les frames suivantes.

Le module de suivi est un filtre à particule qui utilise les estimations de Kalman pour prédire les prochains emplacements possibles des objets en mouvement. Nous avons choisi comme méthode de classification l'algorithme *K Plus Proches Voisins* qui permet par conception, la possibilité d'effectuer un apprentissage en ligne adaptatif des paramètres du classifieur.

Afin d'alléger le processus de suivi, nous avons utilisé les caractéristiques géométriques comme un critère de discrimination des objets en mouvement. Ces primitives se résument dans le calcul du centre de gravité, l'histogramme de couleur, la taille de la composante connexe et l'utilisation des contours de l'objet. Nous avons également mis en point une stratégie pour assurer à la fois une cohérence temporelle et une cohérence structurelle afin de maximiser les performances du système de suivi. La cohérence structurelle est assurée par l'analyse des histogrammes de contour tandis que l'analyse temporelle est assurée par un travail collaboratif entre le module de détection, le classifieur et le module de suivi.

9

Expérimentations et résultats

9.1 Introduction

L'objectif de l'expérimentation est d'approuver par un protocole de validation les différentes propositions utilisées pour la résolution de problèmes. Nous avons remarqué que dans cette dernière décennie, plusieurs systèmes de vidéosurveillance ont été proposés afin d'augmenter la robustesse, la performance et l'efficacité de ces derniers par rapport aux difficultés rencontrées dans le domaine. Donc, une étude comparative est également indispensable afin de bien situer la performance des méthodes proposées par rapport à ce qui existe dans l'état de l'art.

L'expérimentation est utilisée entre autres pour déterminer la puissance et la faiblesse des algorithmes par rapport à un type spécifique d'environnement. Cependant, lorsque le système est de nature complexe et contenant plusieurs modules à valider, la mise en place de l'évaluation des performances devient une étape difficile et ardue.

9.2 Procédé de validation

Plusieurs travaux de recherche se sont focalisés non pas sur la résolution de problème, mais sur la préparation d'une plateforme adéquate pour permettre une validation correcte et uniforme pour toutes les propositions effectuée dans ce domaine. Cette plateforme est constituée de plusieurs bases sous forme de fichier vidéo ou des fragments d'images d'une vidéo dans divers environnements. La plateforme contient également un ensemble de mé-

triques caractérisant des aspects particuliers d'une tâche demandée. Chaque métrique exprime une contrainte particulière sur le comportement du système et l'utilisation d'une seule métrique n'est pas suffisante pour valider les caractéristiques d'un système.

L'utilisation d'un grand nombre de métriques n'est pas une pratique aisée puisque cette dernière nécessite un temps supplémentaire pour effectuer l'expertise des résultats obtenus. Pour cela, le choix des métriques doit être un compromis entre le besoin d'une validation particulière des performances du système et l'existence de travaux similaires utilisant les mêmes métriques choisies pour une éventuelle étude comparative.

9.3 Vérité-terrain

Afin d'utiliser une métrique particulière, il faut vérifier l'existence de ce qu'on appelle une vérité-terrain. Cette notion permet d'exprimer ce que le système devrait faire. La réalisation d'une base pour les tests n'est pas difficile. Cependant, la génération des vérités-terrain est une tâche vraiment lourde à cause de l'absence d'un mécanisme automatique qui permet de réaliser cette tâche. En effet, l'opérateur humain doit segmenter les images de la base vidéo et étiqueter chacune d'elle manuellement en utilisant un ensemble de règles d'annotation.

Pour une meilleure crédibilité, il est également intéressant d'utiliser plusieurs vidéos vu que l'application d'un algorithme sur des séquences différentes donnera forcément des performances différentes.

9.4 Corpus de données

L'évaluation des performances est effectuée sur un ensemble de vidéos issues d'acquisitions de scènes dans des conditions réelles. Parmi ces vidéos, des séquences ont été sélectionnées pour l'évaluation des performances représentant certaines conditions particulières.

Nous avons choisi des bases de données vidéos prises à partir de différents contextes en tenant compte des fluctuations de l'intensité lumineuse, la richesse avec des objets, le nombre et le type d'objet, les mouvements provoqués par des éléments de la nature, tels que : les nuages, la poussière, les bruits et les mouvements de la caméra.

Pour évaluer les performances du système, nous avons utilisé en plus de nos bases de données, des trois bases de données publiques.

La première (DBA) contient six vidéos marquées A1, A2 ... A6 prises dans quatre environnements représentant respectivement : Campus, Highway I, Highway I2, Highway II, Intelligent room, Laboratory [PRA 03b].

La deuxième (DBB) est constituée de neuf vidéos marquées B1, B2 ... B9 représentant

Base	Libellé	Nombre de Frame	Résolution	Durée (mn)	Fréquence (fps)	Environnement
DBA	A1	1178	352 X 288	01 :57	10	Campus
	A2	439	320 X 240	00 :44	10	Highway I
	A3	439	320 X 240	00 :31	14	Highway I2
	A4	499	320 X 240	00 :35	14	Highway II
	A5	299	320 X 240	00 :30	10	Intelligent room
	A6	886	320 X 240	01 :28	10	Laboratory
DBB	B1	3054	160 X 120	/	/	Bootstrap
	B2	2438	160 X 120	/	/	Campus
	B3	23963	160 X 120	/	/	Curtain
	B4	4814	160 X 130	/	/	Escalator
	B5	1522	160 X 128	/	/	Fountain
	B6	4583	176 X 144	/	/	Hall
	B7	2545	160 X 128	/	/	Lobby
	B8	2285	320 X 256	/	/	Shopping Mall
	B9	1632	160 X 128	/	/	Water surface
DBC	C1	1799	320 X 240	03 :00	10	Hallway
	C2	2055	320 X 240	03 :25	10	Highway III
DBD	D1-1	526	320 X 240	00 :17	30	Campus
	D1-2	896		00 :30		
	D1-3	896		00 :30		
	D2-1	616	640 X 480	00 :21	29	Hallway I
	D2-2	893		00 :30		
	D2-3	382		00 :13		
	D3-1	1210	640 X 480	00 :41	29	Highway IV
	D3-2	420		00 :14		
	D3-3	896		00 :30		
	D4-1	559	640 X 480	00 :19	29	Public Park
	D4-2	585		00 :20		
	D4-3	555		00 :19		
	D5-1	860	640 X 480	00 :29	29	Garden
	D5-2	930		00 :32		
	D5-3	413		00 :14		
D6-1	1260	640 X 480	00 :43	29	a living room	
D6-2	1123		00 :38			
D6-3	413		00 :14			

TABLE 9.1 – Corpus de tests détaillé

respectivement Bootstrap, Campus, Curtain, Escalator, Fountain, Hall, Lobby, Shopping Mall et Water surface [MAR 07]. La troisième (DBC) contient deux vidéos marquées C1, C2 représentant respectivement Hallway, Highway III [LI 04a]. Notre base de données (DBD) est constituée de six environnements marqués D1, D2 ... D6 représentant respectivement Campus, Hallway I, Highway IV, Public Park, Garden et a living room où chaque environnement est pris avec au moins trois vidéos avec des perspectives différentes. Le

tableau 9.1 montre quelques détails sur les vidéos utilisés dans le test.

9.5 Évaluation quantitative

Afin d'évaluer correctement le système proposé, nous avons utilisé plusieurs métriques. L'absence d'un standard pour l'évaluation des performances a rendu le choix des métriques propices très difficile, malgré l'existence de nombreux projets de mise en œuvre d'une plateforme d'évaluation de performance pour les systèmes de vidéosurveillance. Le choix des métriques dépend du problème traité, de la base de vidéos utilisée et de la disponibilité de résultats pour les mêmes métriques dans l'état de l'art pour une éventuelle étude comparative. Ces mesures permettent de caractériser un algorithme vis-à-vis du nombre de réussites et d'échecs ainsi que le comportement global du système face à un problème donné.

On trouve dans la littérature des dizaines de métriques d'évaluation, et même si plusieurs d'entre eux se ressemblent dans la forme, elle se diffère dans la manière d'interpréter les résultats. Nous rappelons que nous nous sommes intéressés à l'évaluation des performances pour l'extraction de l'arrière-plan, la détection d'objets en mouvement, la détection et la suppression d'ombres portées et le suivi d'objets.

Nous allons présenter dans cette section les métriques les plus communément utilisées dans l'évaluation des systèmes de vidéosurveillance.

9.5.1 Notations standards

Avant de définir les mesures d'évaluation, nous allons d'abord présenter quelques définitions usuelles qui sont utilisées dans le calcul de nombreuses métriques.

Vrai Positif (TP) : indique que le système a détecté un évènement qui existe aussi bien dans le résultat de l'algorithme que dans la vérité-terrain.

Vrai Négatif (TN) : indique que le système a détecté un évènement qui n'existe pas dans la vérité-terrain.

Faux Négatif (FN) : indique que ni le système ni la vérité-terrain n'ont détecté d'évènement cependant, l'évènement existe dans la réalité.

Faux Positif (FP) : indique que le système n'a pas détecté un évènement qui existe en réalité dans la vérité-terrain.

9.5.2 Métriques d'évaluation

La première métrique utilisée est le taux de détection connu également comme rappel, il donne le taux des pixels qui ont été correctement classés par rapport aux données de

réalité terrain.

$$Recall = \frac{TP}{TP + FN} \quad (9.1)$$

La seconde est la prédiction positive aussi connue comme précision. Elle donne le taux des vrais positifs détectés par rapport au nombre total de régions détectées.

$$Precision = \frac{TP}{TP + FP} \quad (9.2)$$

Le taux de fausse alarme FAR consiste à comptabiliser le nombre de pixels incorrectement détectés parmi l'ensemble des pixels détectés par l'algorithme. Cette valeur reflète le nombre de faux positifs détecté par l'algorithme.

$$FAR = \frac{FP}{TP + FP} = 1 - Precision \quad (9.3)$$

La F-mesure appelée également F-Score permet de fournir une mesure scalaire obtenue en combinant les valeurs de la Precision et du Recall. Cette valeur doit être aussi élevée que possible Pour refléter à la fois un bon taux de détection et une bonne précision.

$$F - mesure = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (9.4)$$

En plus des mesures présentées précédemment, nous avons eu à définir de nouvelles métriques relatives à la détection des objets en mouvement et à la détection et la suppression des ombres. Ces derniers permettent d'attribuer plus d'information sur le comportement du système.

9.5.3 Mesures ajoutées pour l'extraction de l'arrière-plan

La Sensibilité (SE) concerne la capacité du test à détecter correctement les objets en mouvement.

$$SE = \frac{\text{Nombre de pixels détectés}}{\text{Nombre total de pixels}} \quad (9.5)$$

L'amélioration (AC) est une mesure directe avec un système de référence. Elle permet de mesurer le taux de détection sans se préoccuper des résultats de la vérité-terrain.

$$AC = \frac{\text{Nombre de pixels correctement détectés par notre système}}{\text{Nombre de pixels correctement détectés par un système de référence}} \quad (9.6)$$

La spécificité (SP) se rapporte à la capacité du test à détecter correctement les objets en mouvement sans condition.

$$SP = \frac{\text{Nombre de frames correctement traités}}{\text{Nombre total de frames}} \quad (9.7)$$

9.5.4 Mesures ajoutées pour la détection des ombres

Afin d'évaluer quantitativement la détection de l'ombre, nous avons utilisé deux mesures [PRA 03b]. Le taux de détection de l'ombre η , qui indique à quel point l'algorithme détecte les ombres.

$$\eta = \frac{TP_S}{TP_S + FN_S} \quad (9.8)$$

Le taux de discrimination de l'ombre ξ , qui décrit comment le système peut différencier entre les ombres et les pixels du premier plan.

$$\xi = \frac{\overline{TP_F}}{TP_F + FN_F} \quad (9.9)$$

Les indices S et F désignent respectivement l'ombre et le premier plan. $\overline{TP_F}$ est le nombre exact de points des objets du premier plan moins le nombre de points des objets du premier plan marqué comme l'ombre. η et ξ sont définis par rapport deux vérités-terrain, le premier concerne l'ombre et le deuxième les objets. TP_S et FN_S sont calculés selon la vérité-terrain d'ombre tandis que TP_F et FN_F sont calculées à l'aide de la vérité-terrain d'objets.

9.5.5 Modalité d'utilisation de la vérité-terrain

Toutes les métriques ont une relation directe ou indirecte avec la vérité-terrain. Cependant, l'utilisation de la vérité-terrain est conditionnée par un certain nombre de contraintes et ne peut être effectuée de manière informelle. Pour cela, plusieurs mesures de comparaison entre les résultats du système (S) et la vérité-terrain (VT) ont été proposées [NGH 07] pour déterminer le nombre de succès et d'échecs d'un algorithme. Les distances utilisées doivent prendre en compte à la fois l'aspect spatial et l'aspect temporel des observations fournies par l'algorithme. L'association est déterminée à l'aide du recouvrement spatial et temporel entre l'observation et la vérité-terrain. Nous allons présenter quatre mesures notées M_1, M_2, M_3 et M_4

$$M_1 = \frac{2 \times \text{Card}(S \cap VT)}{\text{Card}(S) + \text{Card}(VT)} \quad (9.10)$$

$$M_2 = \frac{\text{Card}(S \cap VT)}{\text{Card}(VT)} \quad (9.11)$$

$$M_3 = \frac{\text{Card}(S \cap VT)^2}{\text{Card}(S) \times \text{Card}(VT)} \quad (9.12)$$

$$M_4 = \max \left(\frac{\text{Card}(S \cap VT)}{\text{Card}(S)}, \frac{\text{Card}(S \cap VT)}{\text{Card}(VT)} \right) \quad (9.13)$$

Où $\text{Card}(E)$ représente le nombre d'éléments de l'ensemble E . Dans le cas de notre étude, une décision binaire est préférable. Pour cela, nous avons appliqué un seuillage sur ces

mesures pour la mise en correspondance.

9.6 Résultats relatifs à l'extraction de l'arrière-plan

9.6.1 Configuration des paramètres

Pour assurer la stabilité du système lors de tests de performance, le nombre gaussiennes (K), le taux d'apprentissage (α) et le seuil fixant la portion de mesure minimale (B) ont été empiriquement fixés à 5, 0.001 et 0.3.

9.6.2 Résultats qualitatifs

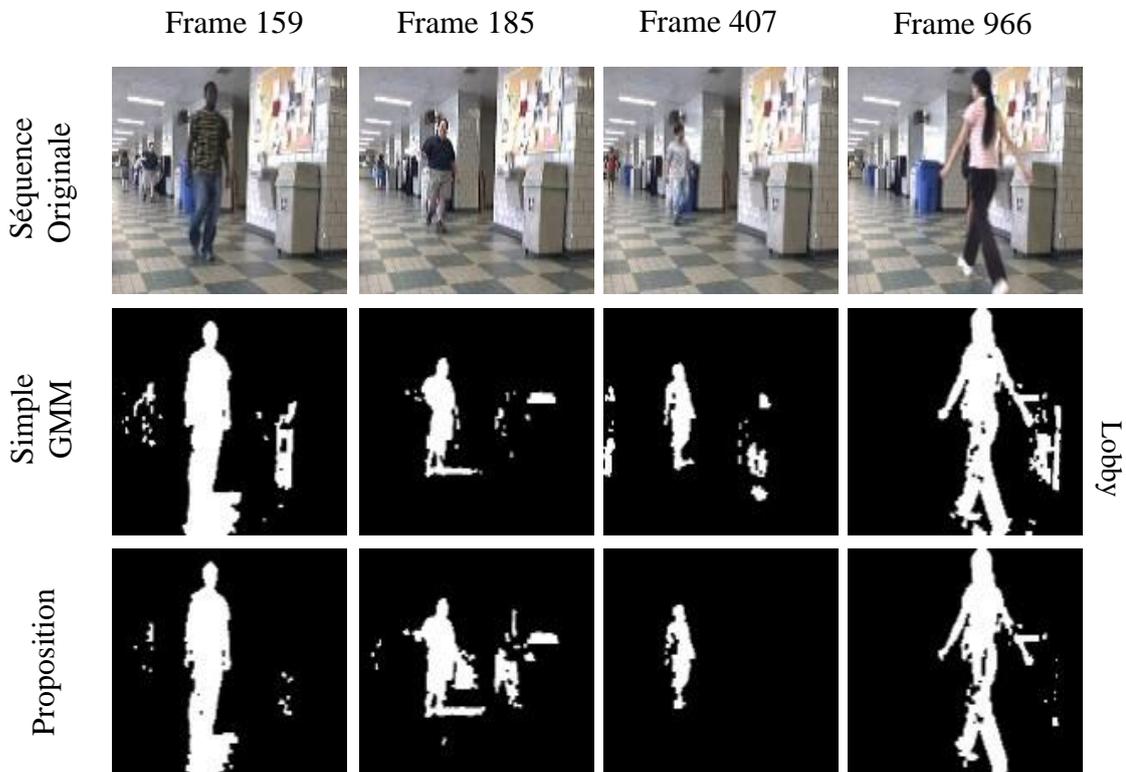


FIGURE 9.1 – Résultats de la soustraction de l'arrière-plan dans la vidéo B7

La figure 9.1 montre que le système proposé a effectué une légère correction par rapport à une simple GMM dans des environnements à faible variation.

La figure 9.2 montre clairement que l'approche proposée donne de meilleurs résultats par rapport à une simple GMM. Elle montre que notre système est en mesure de réduire considérablement le bruit causé par la variation instantanée de la luminosité et de garder une qualité acceptable dans des environnements stables. La figure 9.2 montre l'efficacité du système proposé dans un environnement avec de multiples objets en mouvement. Il

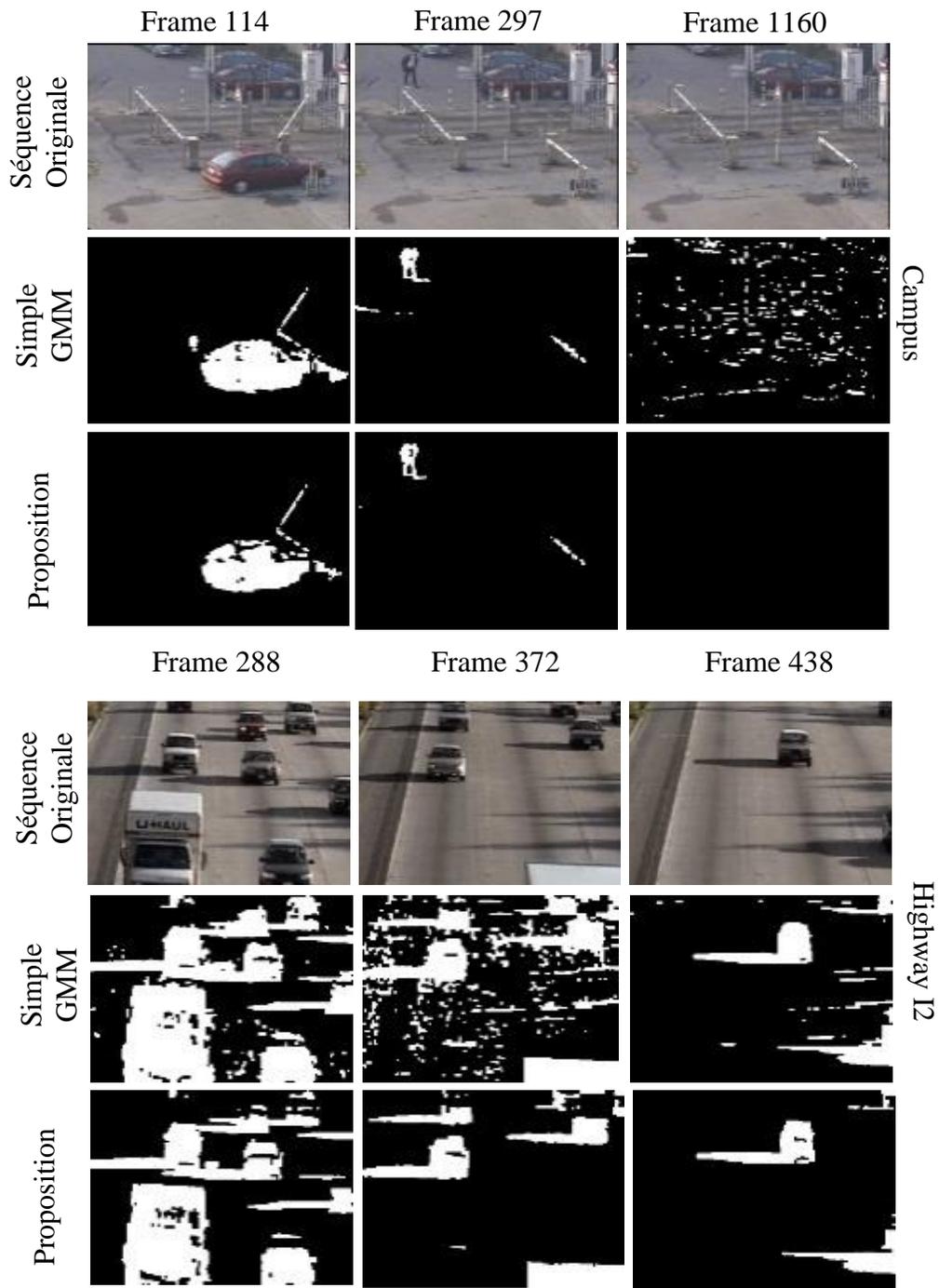


FIGURE 9.2 – Résultats de la soustraction de l’arrière-plan dans les vidéos A1 et A3

montre également la faible influence de l’effet de la poussière causée par les voitures en mouvement.

La figure 9.3 montre l’efficacité de notre système dans les environnements contenant une faible source de luminosité. Les environnements intérieurs posent énormément de problèmes de luminosité à cause de l’influence des lumières réfléchies sur les objets. En effet, les environnements extérieurs sont éclairés par le soleil qui a une forte luminosité et qui rend la lumière réfléchie pratiquement négligeable devant cette lumière instance. Tan-

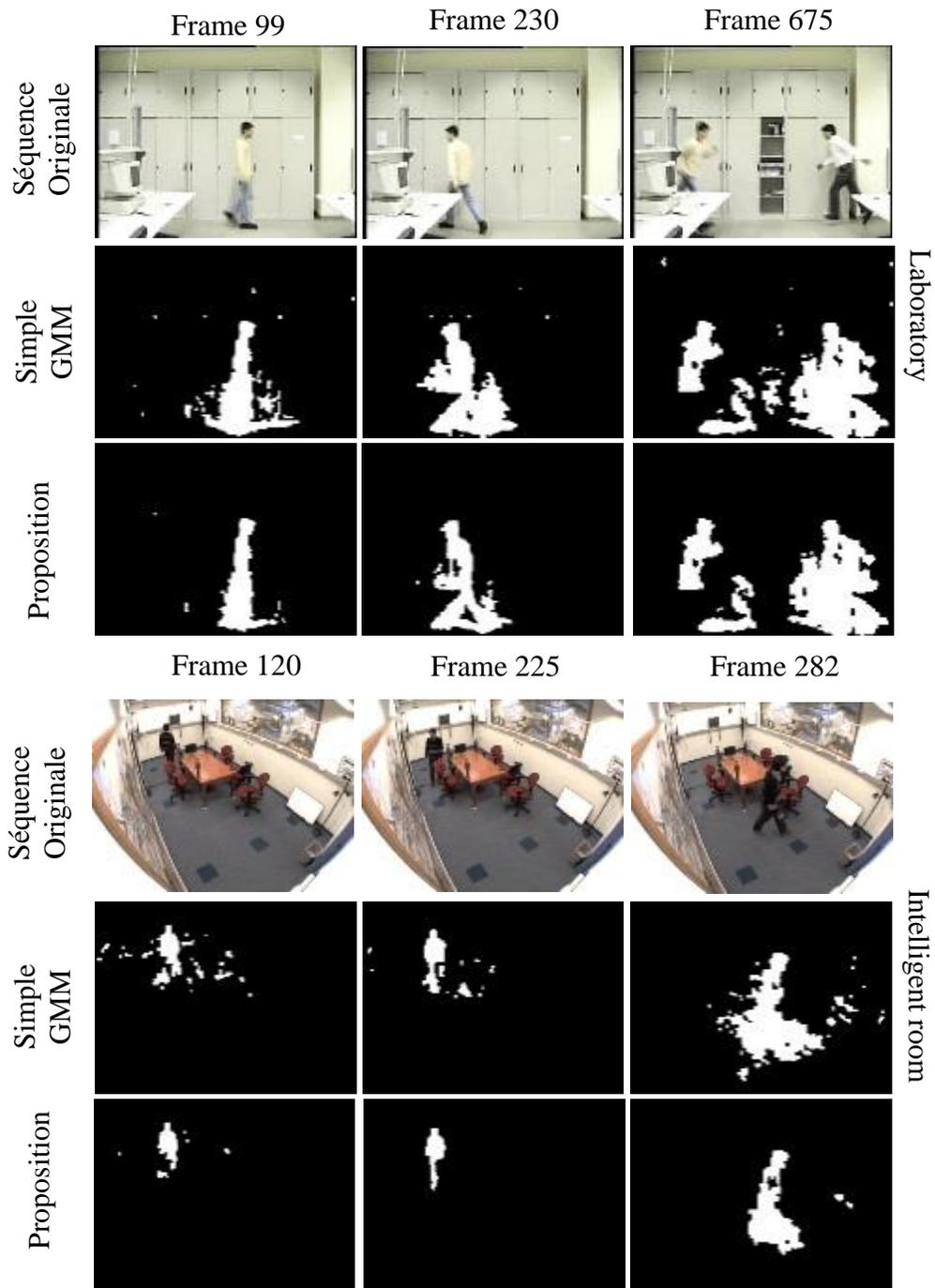


FIGURE 9.3 – Résultats de la soustraction de l’arrière-plan dans les vidéos A5 et A6

dis que les environnements intérieurs se perturbent facilement par la réflectance des objets et l’apparition d’un seul objet en mouvement peut changer l’effet global de la lumière sur la scène ce qui génère une très forte perturbation sur le comportement des systèmes d’extraction de l’arrière-plan.

La figure 9.4 montre la stabilité de notre système dans un environnement poussiéreux causé principalement par le vent.

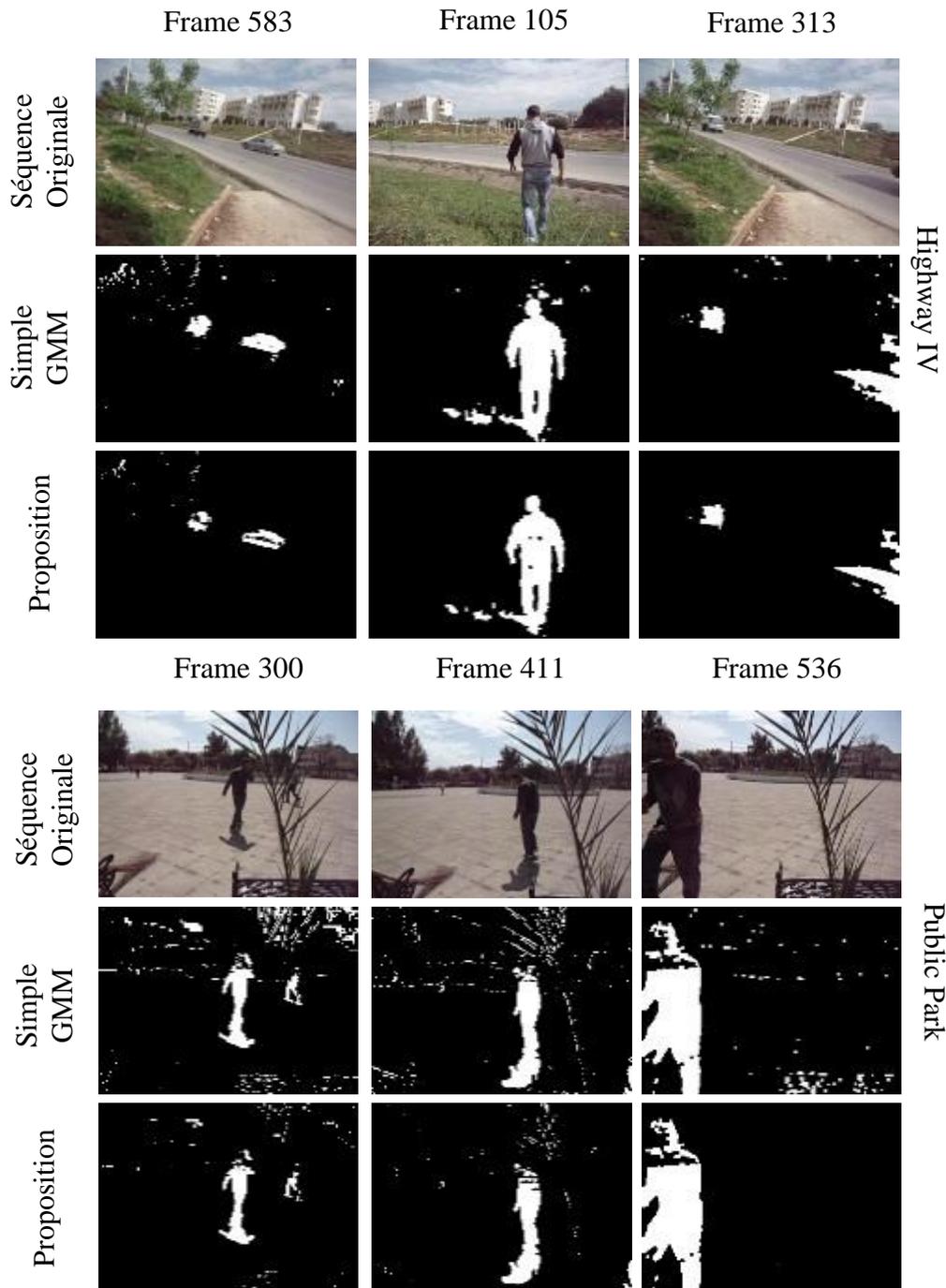


FIGURE 9.4 – Résultats de la soustraction de l’arrière-plan dans les vidéos D3-1 et D4-2

Dans la figure 9.5, nous pouvons voir clairement la contribution de notre approche dans les vidéos prises au hasard et sans aucune contrainte dans un environnement complexe. La première partie de la figure 9.5 (campus) est prise dans des conditions météorologiques instables lorsque le passage des nuages dans une journée ensoleillée génère un changement de luminosité instantanée. Le frame 196 montre l’effet indésirable causé par cette variation sur une simple GMM tandis que notre système a réduit considérablement cette perturbation. La deuxième partie de la figure 9.5 est prise dans un couloir sombre

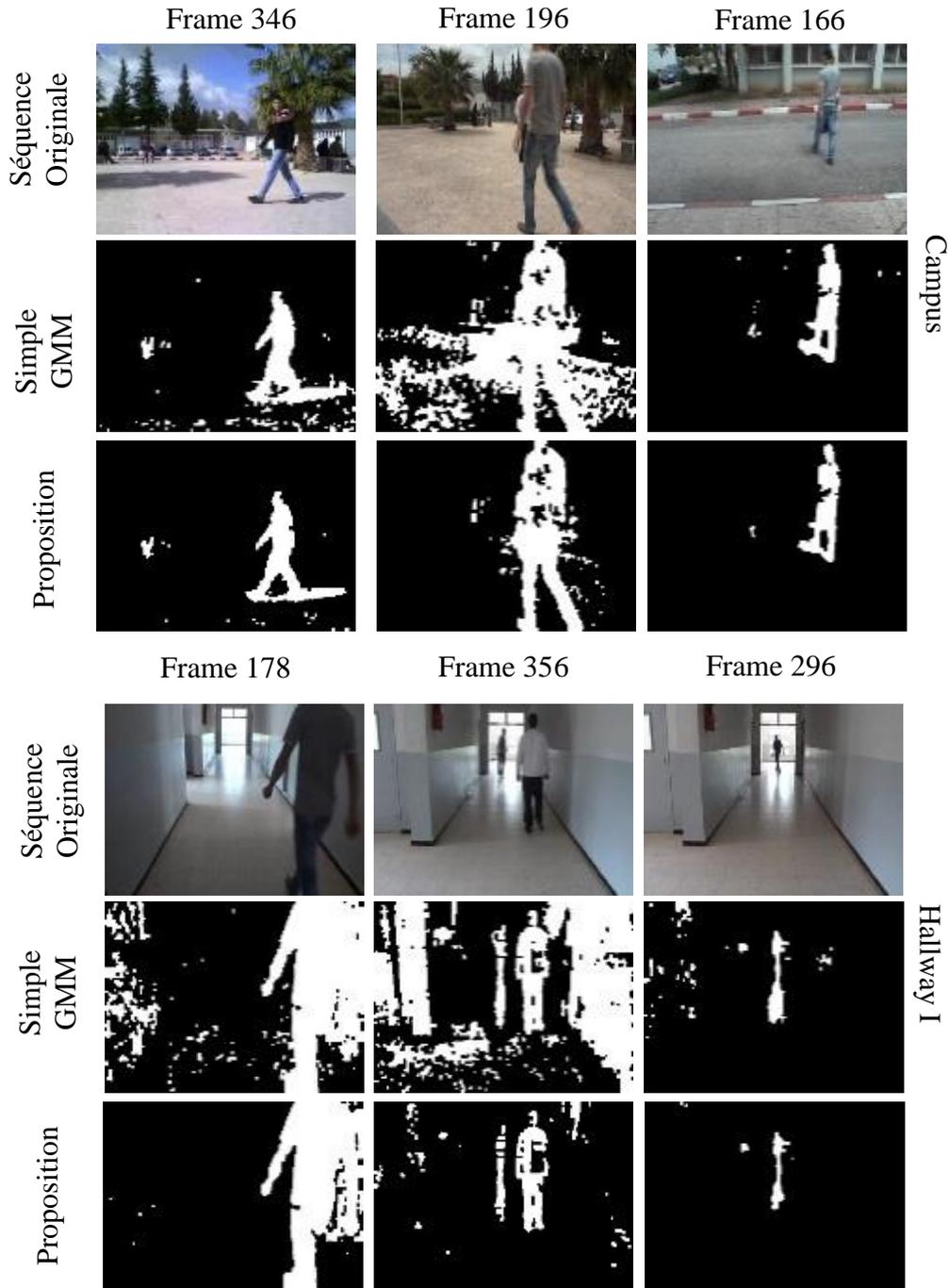


FIGURE 9.5 – Résultats de la soustraction de l’arrière-plan dans les vidéos D1-1, D1-2, D1-3 et D2-2

avec une fenêtre. Elle montre l’efficacité de notre système par rapport à de très grandes variations de l’intensité lumineuse.

9.6.3 Résultats quantitatifs

Afin d’évaluer les performances du système quantitativement, nous avons utilisé les métriques, présentées dans la section précédente. Nous avons également calculé le temps

d'exécution qui exprime la complexité algorithmique des méthodes proposées. Nous avons implémenté une simple GMM basée sur le modèle de Stauffer et Grimson [STA 99b] pour effectuer une étude comparatif. Le tableau 9.2 montre les résultats obtenus par notre système et celui d'un simple GMM et cela dans les bases vidéo publiques et personnelles déjà présentées.

Vidéos	SE		AC	SP%		PT(s)	
	Proposée	simple GMM		Proposée	simple GMM	Proposée	simple GMM
A1	02.49	04.34	0.57	08.66	12.46	0.46	0.92
A2	16.50	21.72	0.76	08.15	12.79	0.41	0.89
A3	15.02	20.27	0.74	08.09	11.82	0.38	0.87
A4	05.13	05.76	0.89	08.14	12.31	0.38	0.87
A5	01.77	03.28	0.54	08.34	12.09	0.41	0.89
A6	08.12	10.18	0.80	08.69	12.55	0.41	0.89
B1	19.18	24.60	0.78	14.51	18.66	0.31	0.45
B2	24.21	34.10	0.71	15.13	18.33	0.31	0.45
B3	21.80	32.54	0.67	15.44	17.14	0.31	0.45
B4	14.58	29.16	0.50	15.01	17.53	0.31	0.46
B5	13.61	24.76	0.55	15.41	18.48	0.32	0.46
B6	28.28	35.35	0.80	15.55	17.88	0.34	0.48
B7	19.75	34.06	0.58	15.28	17.77	0.32	0.46
B8	13.77	21.19	0.65	14.49	17.89	0.42	0.91
B9	24.23	35.64	0.68	15.66	18.65	0.32	0.45
C1	08.04	09.28	0.87	21.47	21.85	0.44	1.19
C2	07.50	90.06	0.83	21.56	21.82	0.43	1.17
D1	06.93	13.09	0.53	08.12	11.82	0.41	0.89
D2	08.75	22.25	0.39	08.04	11.53	0.57	1.46
D3	03.82	04.81	0.79	08.18	12.01	0.54	1.43
D4	06.04	07.60	0.79	08.92	11.44	0.57	1.48
Moy	12.83	19.99	0.68	12.51	15.56	0.39	0.83

TABLE 9.2 – Résultats obtenus par les critères SE, AC, SP et PT

Dans le tableau 9.2, le critère SE montre que l'approche proposée a une faible sensibilité au bruit par rapport à une simple GMM dans toutes les bases de données utilisées avec un écart moyen de 6,36%, tandis que le critère AC montre une meilleure qualité avec un taux moyen de 68%. Le critère SP montre que notre système a un retard de 3,05%. Ceci est principalement dû au non-déclenchement des observateurs chargés de certaines régions de l'arrière-plan. En effet, les observateurs de l'arrière-plan ne pouvaient pas être déclenchés parce que la forme de l'histogramme n'a pas subi de changement significatif. Cette erreur est due à l'utilisation du modèle HSV qui donne parfois une certaine équivalence après un transfert de l'espace RGB vers l'espace de couleur HSV. Le critère PT montre que malgré l'ajout de certaines opérations, le temps de calcul de notre système est

beaucoup plus faible par rapport à une simple GMM avec un écart moyen de 0,44 seconde. Ce gain de performance est dû au déclenchement sélectif des mises à jour dans le modèle de fond. La majorité des travaux proposés pour améliorer les performances des GMM utilisent des bases de données vidéo personnelles (pour une application très spécifique), qui rendent difficile la tâche de comparaison et parfois impossible. L'absence de mesure de résultat unifié et l'absence de code source pour les autres méthodes sont également des problèmes pour situer notre travail par rapport à d'autres méthodes.

Le tableau 9.3 montre également les améliorations apportées par notre système sur le rendement de l'extraction de l'arrière-plan par un gain de 4% dans le taux de détection et un gain de 8% dans la précision soit un apport total de 6% selon la métrique F-score.

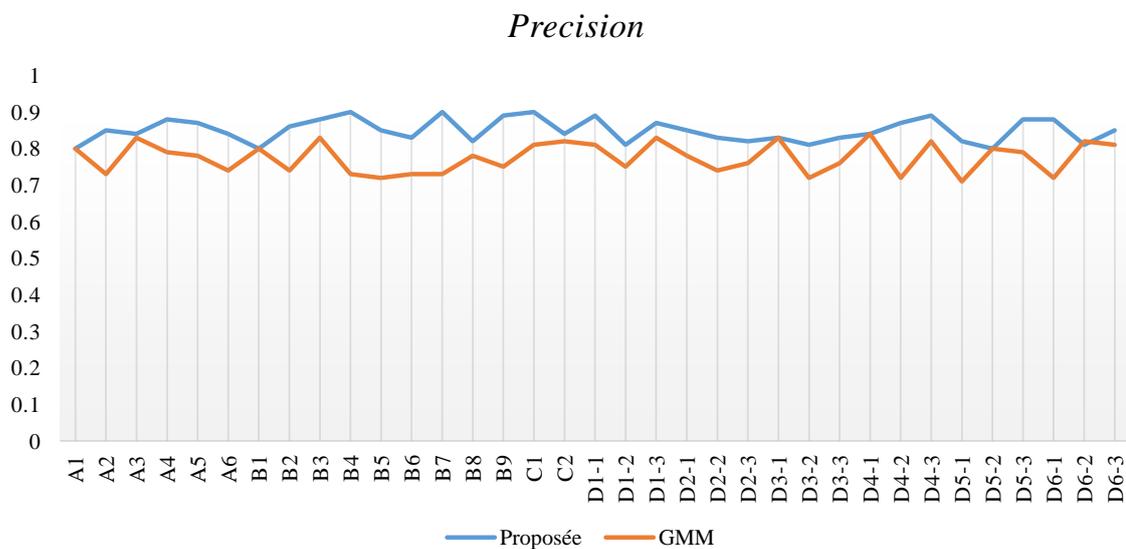


FIGURE 9.6 – Résultats de l'évaluation des performances en termes de Precision

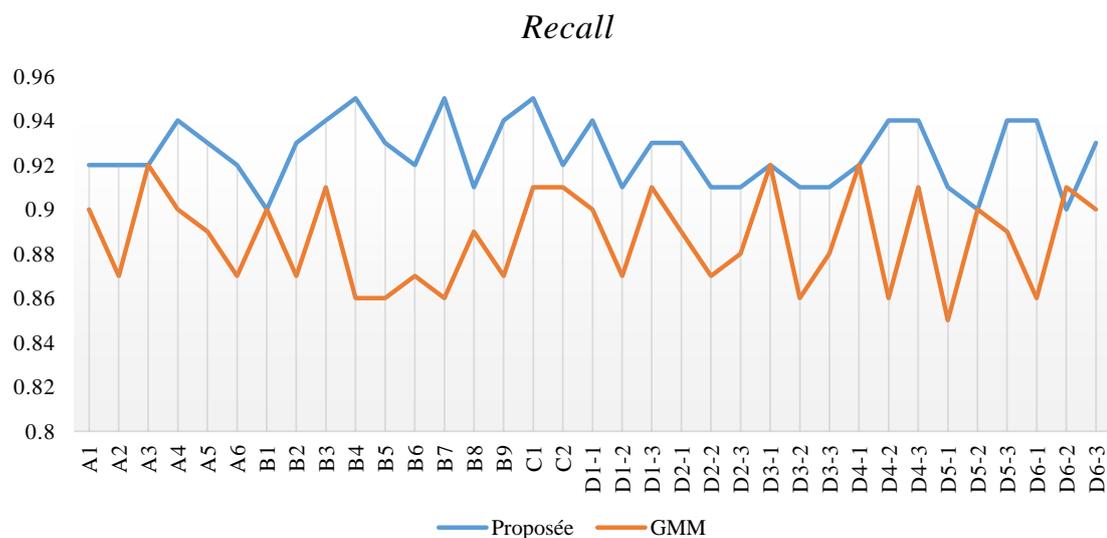


FIGURE 9.7 – Résultats de l'évaluation des performances en termes de Rappel

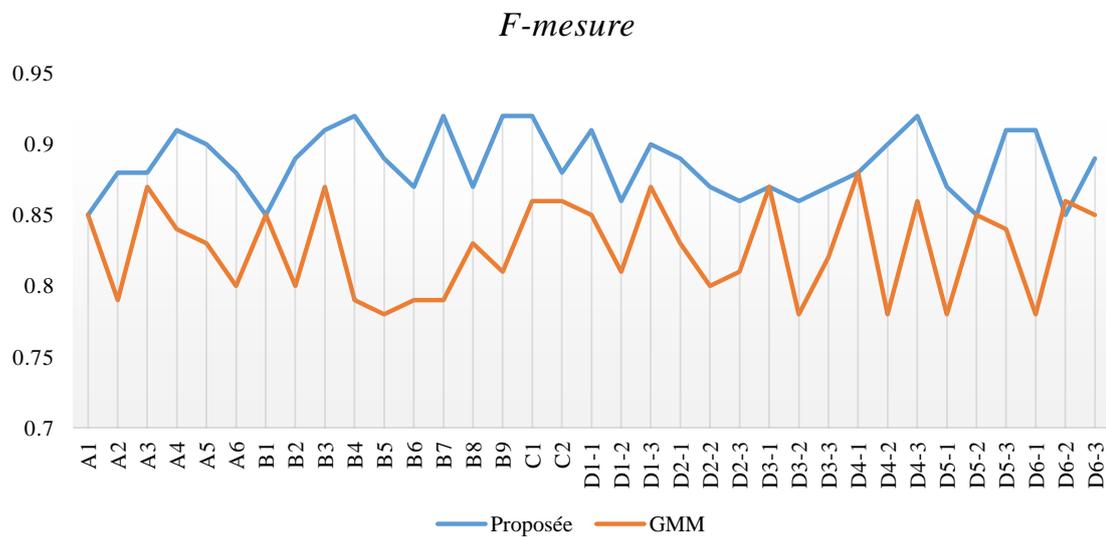


FIGURE 9.8 – Résultats de l'évaluation des performances en termes F-mesure.

Vidéos	Recall		Precision		F-Score	
	Proposée	GMM	Proposée	GMM	Proposée	GMM
A1	0.92	0.90	0.80	0.80	0.85	0.85
A2	0.92	0.87	0.85	0.73	0.88	0.79
A3	0.92	0.92	0.84	0.83	0.88	0.87
A4	0.94	0.90	0.88	0.79	0.91	0.84
A5	0.93	0.89	0.87	0.78	0.90	0.83
A6	0.92	0.87	0.84	0.74	0.88	0.80
B1	0.90	0.90	0.80	0.80	0.85	0.85
B2	0.93	0.87	0.86	0.74	0.89	0.80
B3	0.94	0.91	0.88	0.83	0.91	0.87
B4	0.95	0.86	0.90	0.73	0.92	0.79
B5	0.93	0.86	0.85	0.72	0.89	0.78
B6	0.92	0.87	0.83	0.73	0.87	0.79
B7	0.95	0.86	0.90	0.73	0.92	0.79
B8	0.91	0.89	0.82	0.78	0.87	0.83
B9	0.94	0.87	0.89	0.75	0.92	0.81
C1	0.95	0.91	0.90	0.81	0.92	0.86
C2	0.92	0.91	0.84	0.82	0.88	0.86
D1-1	0.94	0.90	0.89	0.81	0.91	0.85
D1-2	0.91	0.87	0.81	0.75	0.86	0.81
D1-3	0.93	0.91	0.87	0.83	0.90	0.87
D2-1	0.93	0.89	0.85	0.78	0.89	0.83
D2-2	0.91	0.87	0.83	0.74	0.87	0.80
D2-3	0.91	0.88	0.82	0.76	0.86	0.81
D3-1	0.92	0.92	0.83	0.83	0.87	0.87
D3-2	0.91	0.86	0.81	0.72	0.86	0.78
D3-3	0.91	0.88	0.83	0.76	0.87	0.82
D4-1	0.92	0.92	0.84	0.84	0.88	0.88
D4-2	0.94	0.86	0.87	0.72	0.90	0.78
D4-3	0.94	0.91	0.89	0.82	0.92	0.86
D5-1	0.91	0.85	0.82	0.71	0.87	0.78
D5-2	0.90	0.90	0.80	0.80	0.85	0.85
D5-3	0.94	0.89	0.88	0.79	0.91	0.84
D6-1	0.94	0.86	0.88	0.72	0.91	0.78
D6-2	0.90	0.91	0.81	0.82	0.85	0.86
D6-3	0.93	0.90	0.85	0.81	0.89	0.85
Moy	0.93	0.89	0.85	0.77	0.89	0.83

TABLE 9.3 – Résultats obtenus par les critères Recall, Precision et F-Score

9.7 Résultats relatifs à la détection des ombres

Les travaux comparatifs approfondis effectués par Andres Sanin et al. [SAN 12], ont montré que la méthode basée sur les textures à régions larges (LR) donne de meilleurs résultats par rapport aux autres méthodes « *In all cases, the large region texture-based method performs considerably better than all the others, obtaining high values for both the detection and discrimination rates in all sequences* ». Pour mieux situer notre approche par rapport à ce qui existe nous avons implémenté la méthode LR qui sera utilisée comme une méthode de référence.

9.7.1 Résultats qualitatifs

Nous avons choisi quelques résultats qualitatifs des tests effectués sur les bases vidéo présentées précédemment. Les pixels colorés en bleu indiquent les zones dans lequel nous avons détecté une ombre. Les figures montrent également le masque d'ombre pour une bonne comparaison visuelle des résultats obtenus. Nous avons choisi une large gamme de vidéos dans un environnement extérieur vu que ces derniers comportent une forte variation d'intensité lumineuse et de bruit.

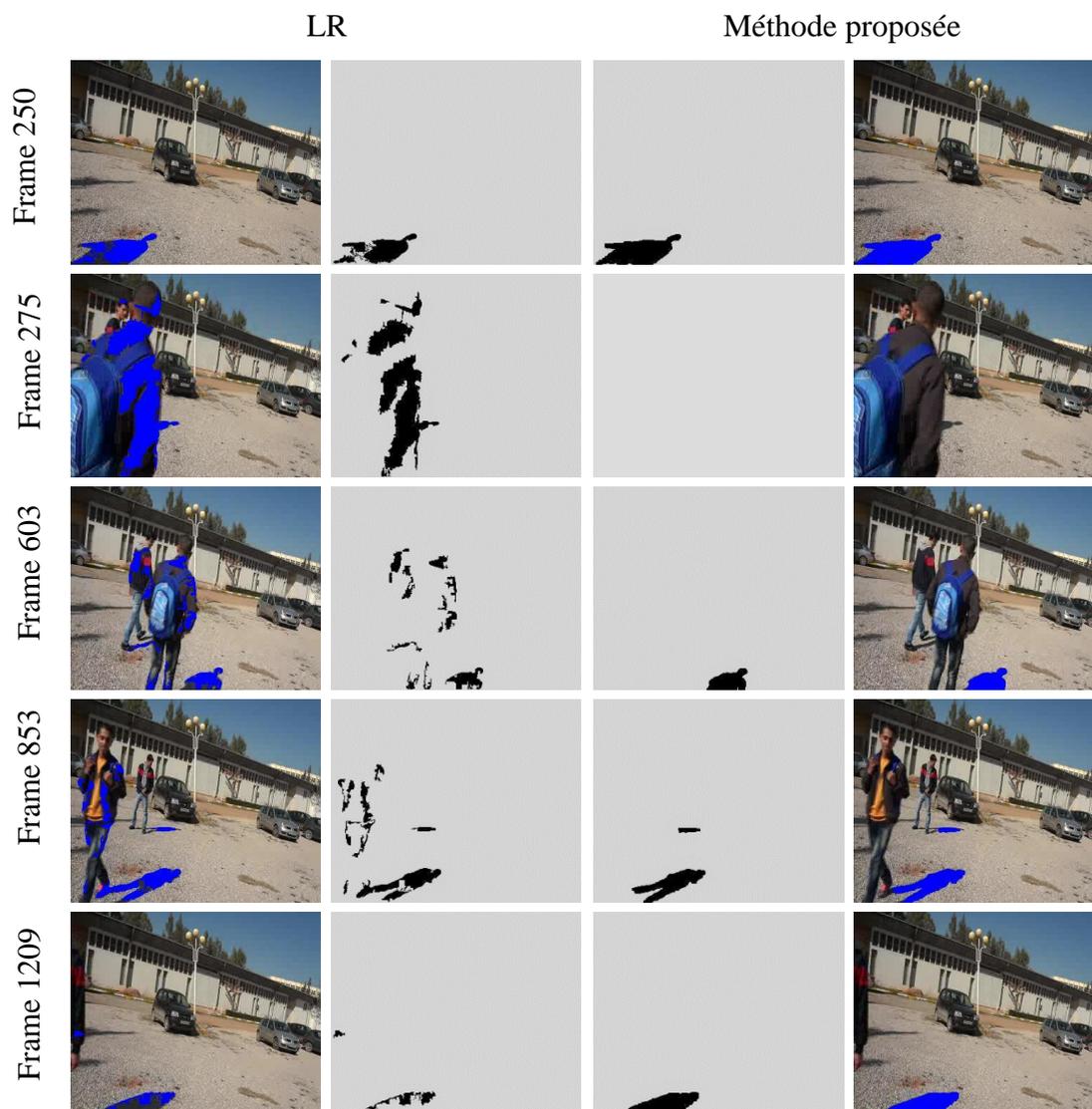


FIGURE 9.9 – Résultats de la détection de l’ombre dans la vidéo D4-2 de la base DBD

La figure 9.9 montre clairement que notre approche permet de détecter les ombres pour tous les objets en mouvements sans aucune supposition préalable sur la nature de l’environnement ou sur les objets en mouvement.

Elle montre également que notre approche donne de meilleurs résultats par rapport à la méthode des textures larges.

Dans le frame 275, on remarque clairement que les textures n’ont pas permis de faire la différence entre l’objet et son ombre.

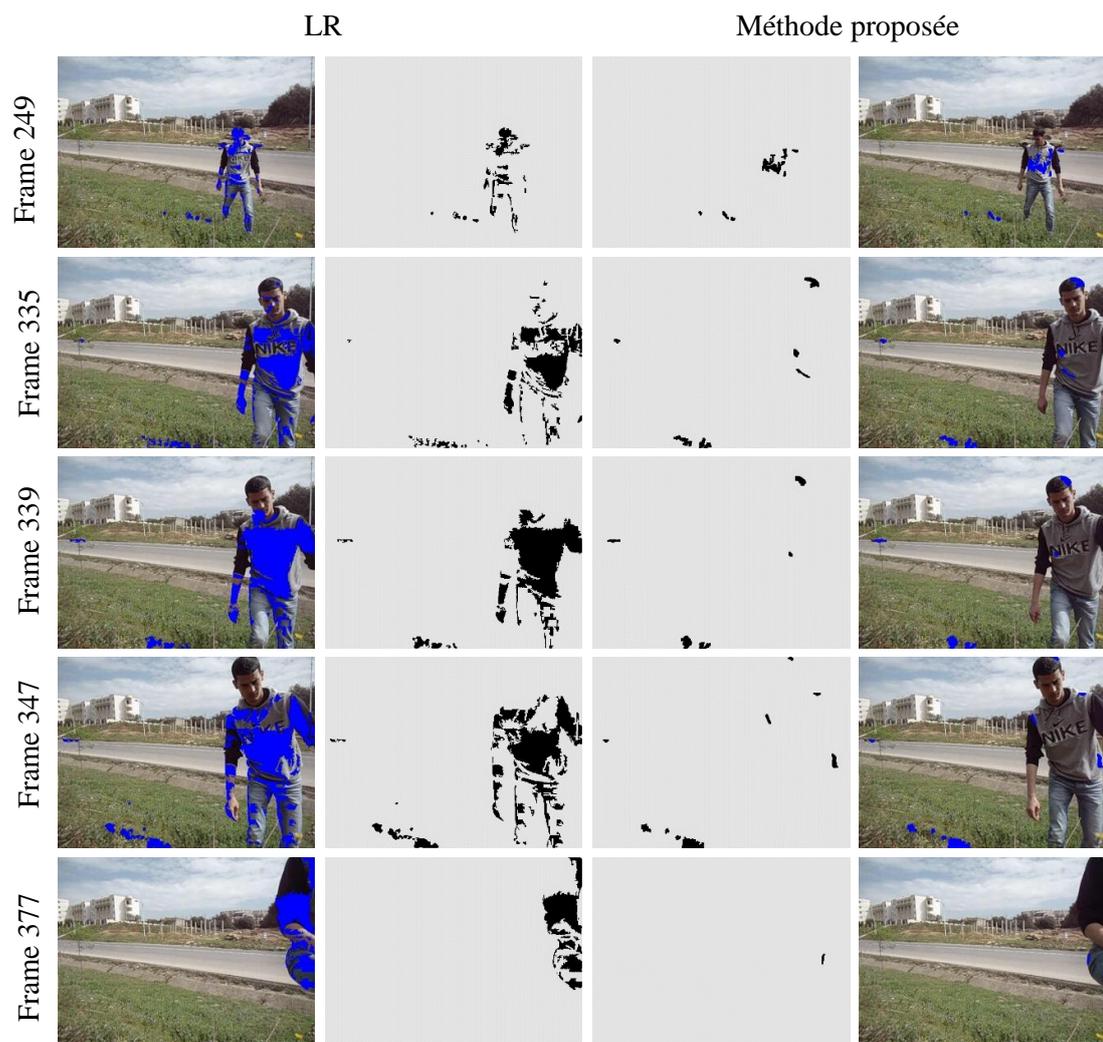


FIGURE 9.10 – Résultats de la détection de l'ombre dans la vidéo D3-2

La vidéo de la figure 9.10 a été prise dans une autoroute dans un jour ensoleillé avec un passage fréquent de nuage et de vent.

Les frames 339, 347 montrent que la non-distinction entre les ombres portées et l'ombre propre a totalement déformé l'objet en mouvement. Tandis que notre système a parfaitement différencié entre les deux types d'ombre.

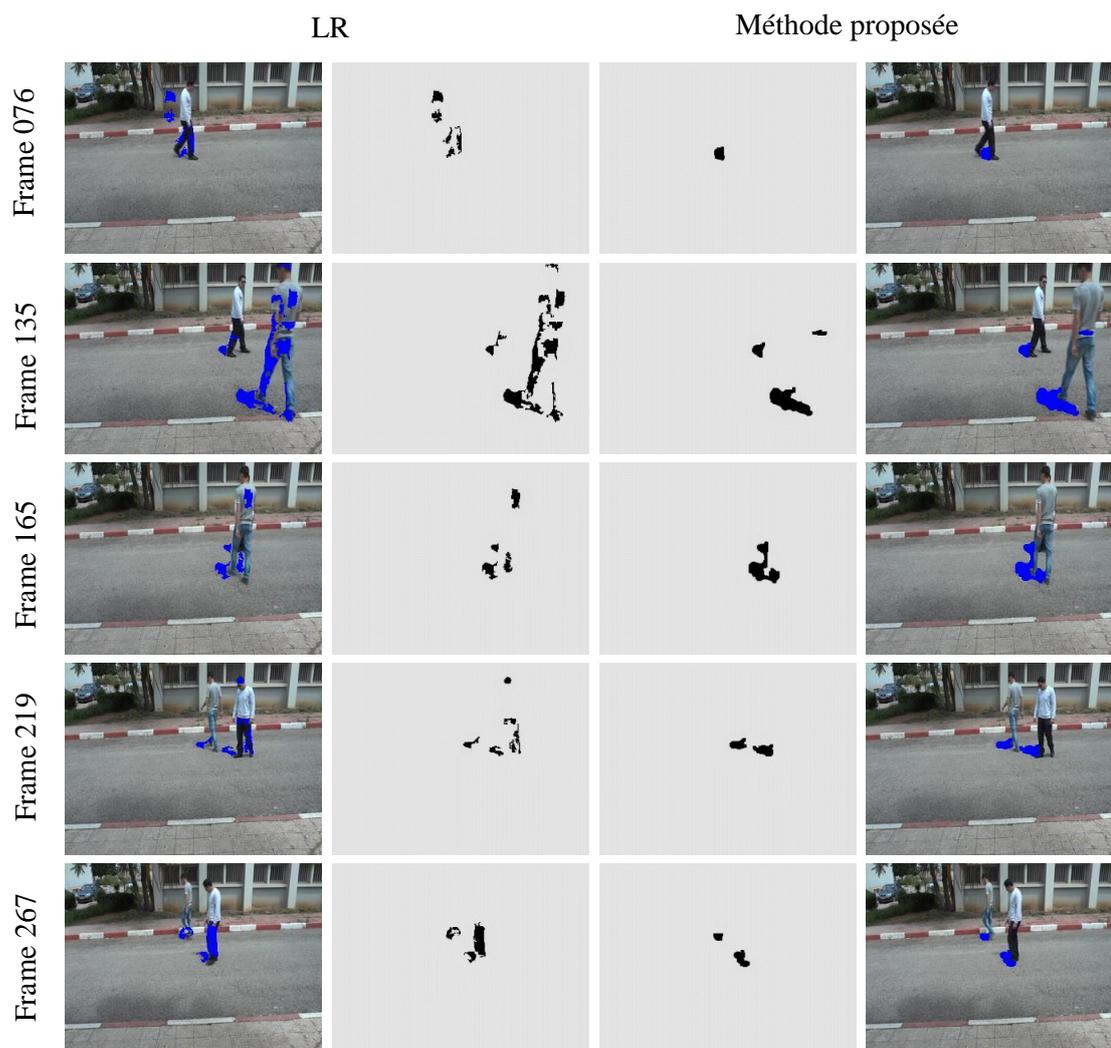


FIGURE 9.11 – Résultats de la détection de l'ombre dans la vidéo D1-3

Le frame 135 de la figure 9.11 montre que notre système permet de faire la différence entre l'ombre portée et l'ombre propre ce qui n'est pas le cas dans les systèmes à base de textures larges.

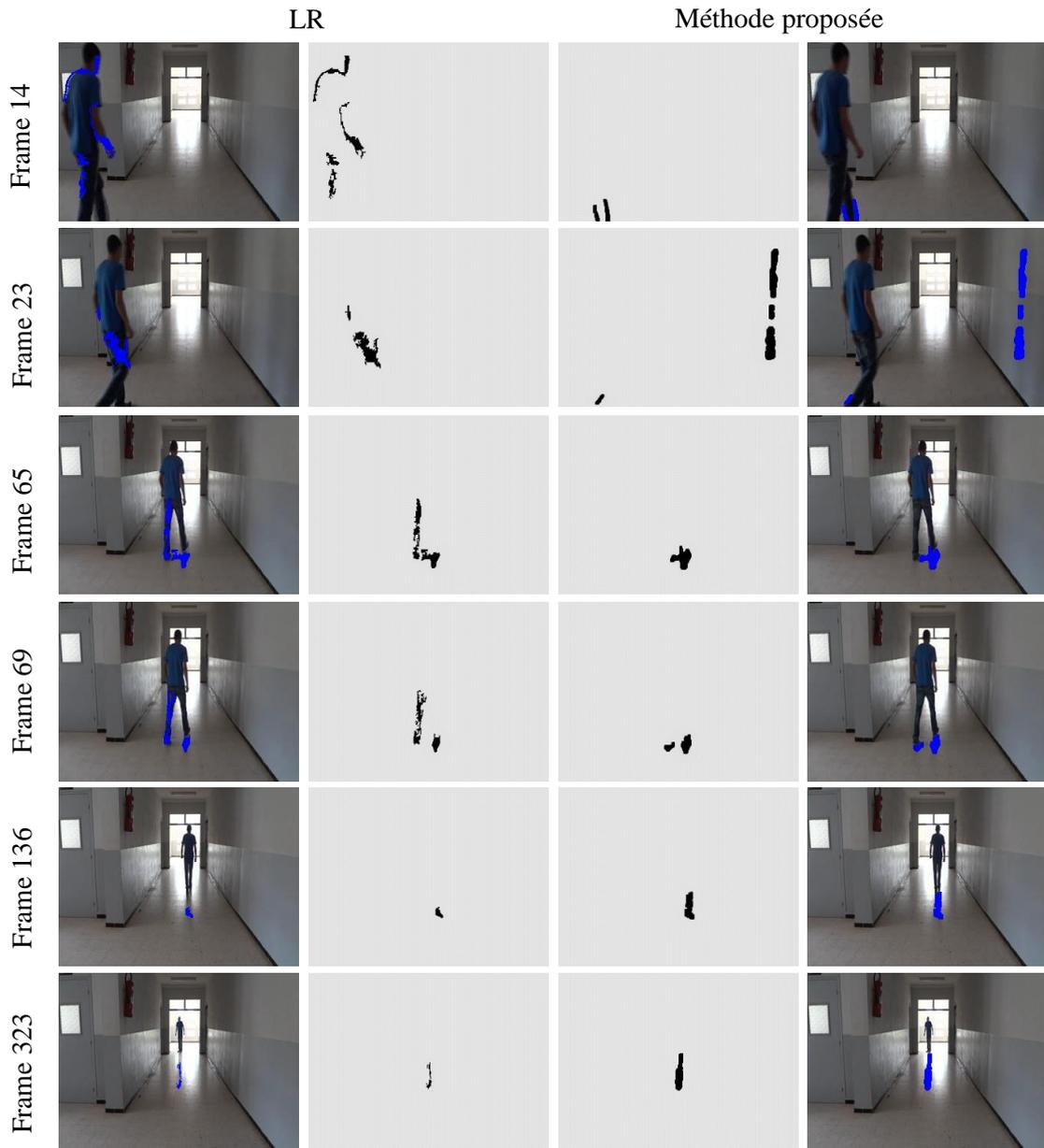


FIGURE 9.12 – Résultats de la détection de l’ombre dans la vidéo D2-1

La figure 9.12 montre l’efficacité de notre système dans les environnements partiellement sombre avec des effets de réflexion de lumière sur les murs et sur le sol. Notre système a été capable de détecter l’ombre portée sur le sol et sur les murs. Le frame 23 montre également l’avantage de notre système par rapport aux méthodes des textures larges dans la séparation entre l’ombre propre et l’ombre portée.

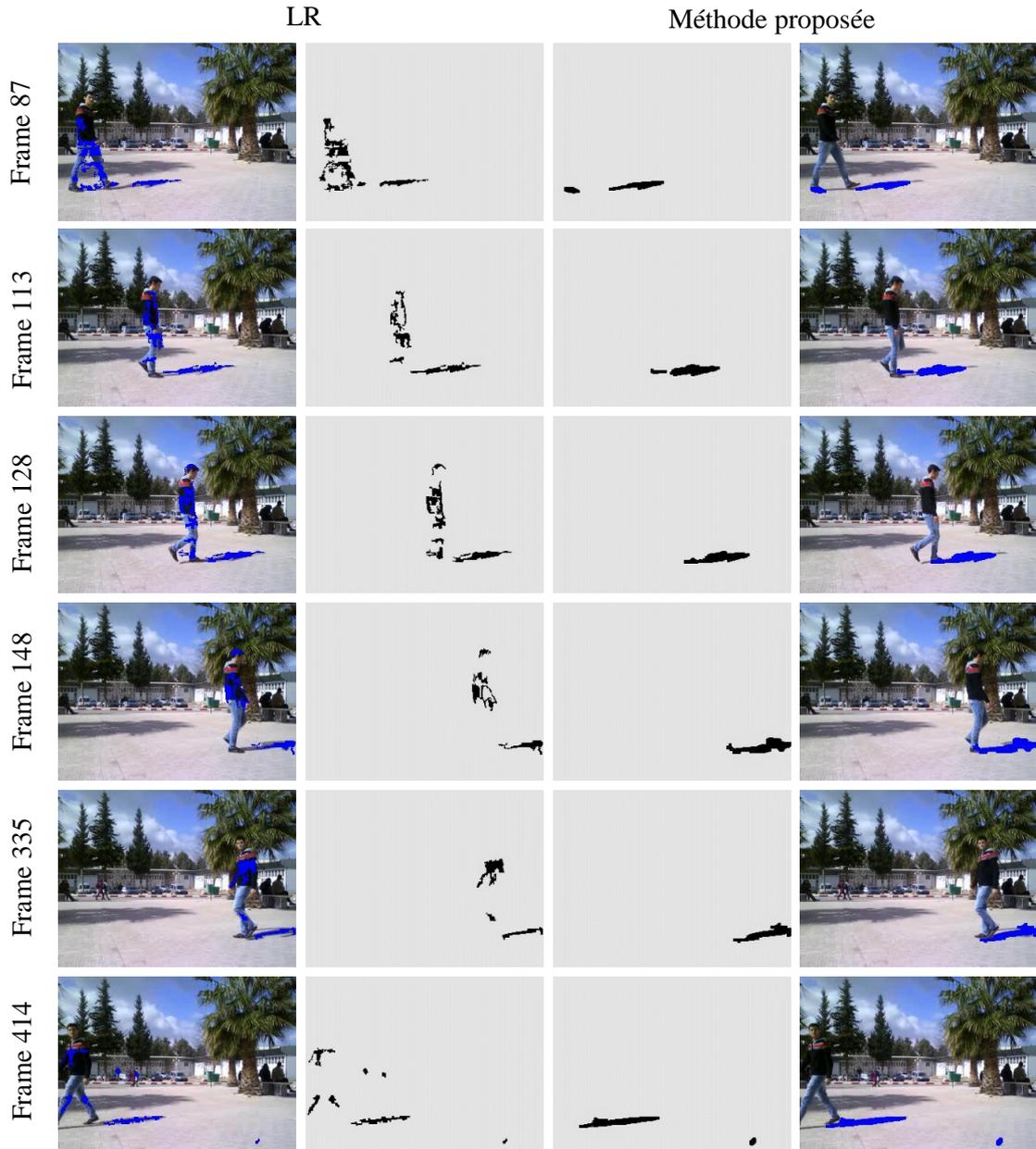


FIGURE 9.13 – Résultats de la détection de l’ombre dans la vidéo D1-1

La figure 9.13 montre les résultats extraits à partir d’une vidéo prise dans le campus de l’université de Guelma dans un jour ensoleillé avec un passage fréquent de nuage. Ce phénomène permet de créer un changement instantané de luminosité affectant toutes les couleurs de l’arrière-plan. Elle montre la puissance du système proposé même dans des conditions d’éclairage variable et aléatoire.

Elle montre aussi la capacité d’adaptation du système avec les changements multiples dans les couleurs de l’arrière-plan.

On remarque aussi que notre système a permis d’éliminer les faiblesses de la méthode des textures large en faisant la différence entre l’ombre protégée et l’ombre propre.

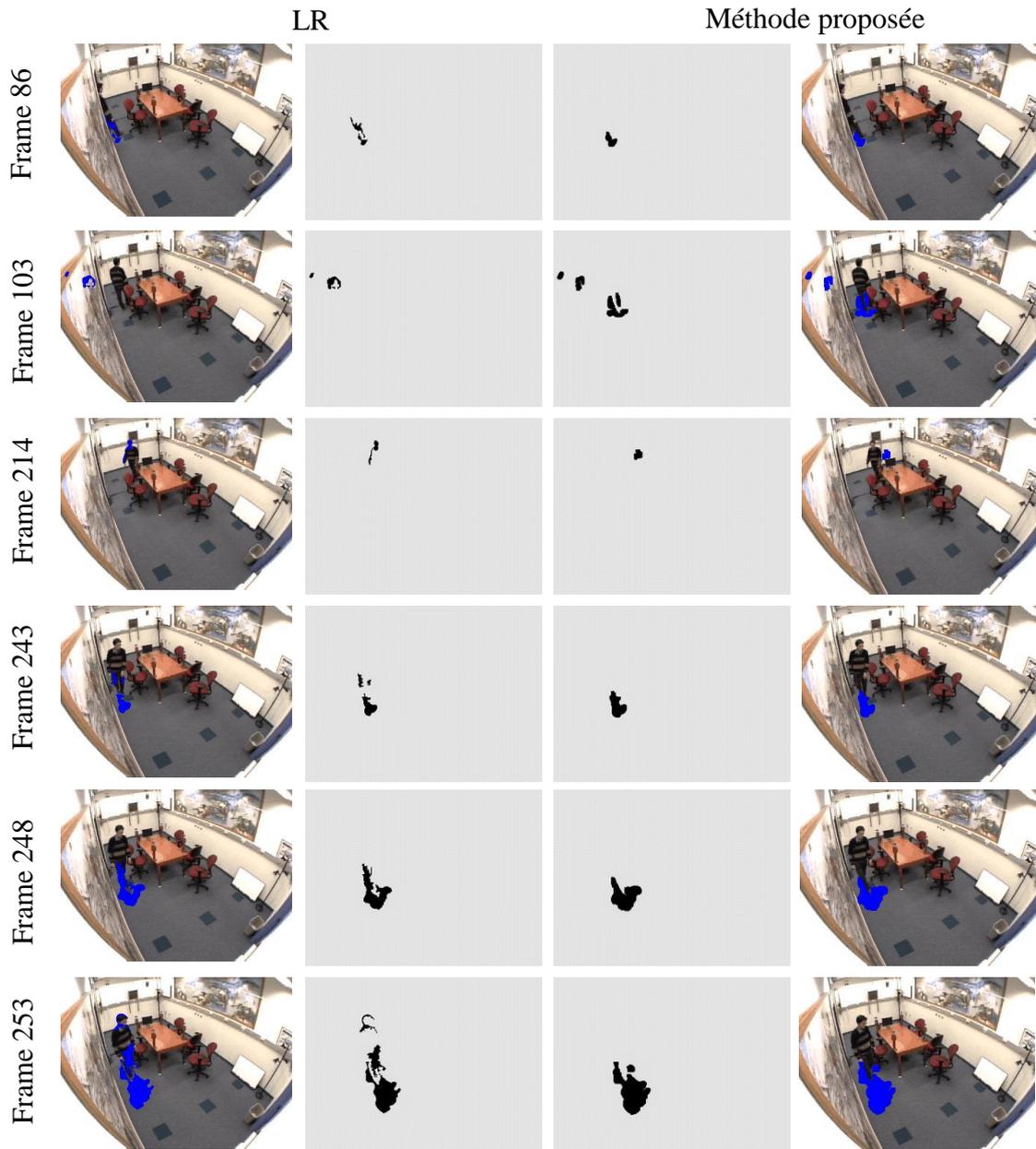


FIGURE 9.14 – Résultats de la détection de l’ombre dans la vidéo A5

La figure 9.14 montre également la capacité de notre système à détecter les ombres dans les bases vidéo publiques.

Le frame 248 de la figure montre que notre système a réussi à détecter intégralement l’ombre même dans les pires scénarios vu que la couleur du sol est pratiquement similaire à la couleur de l’ombre. En effet, cette situation est parmi les inconvénients des méthodes basées sur la chromaticité.

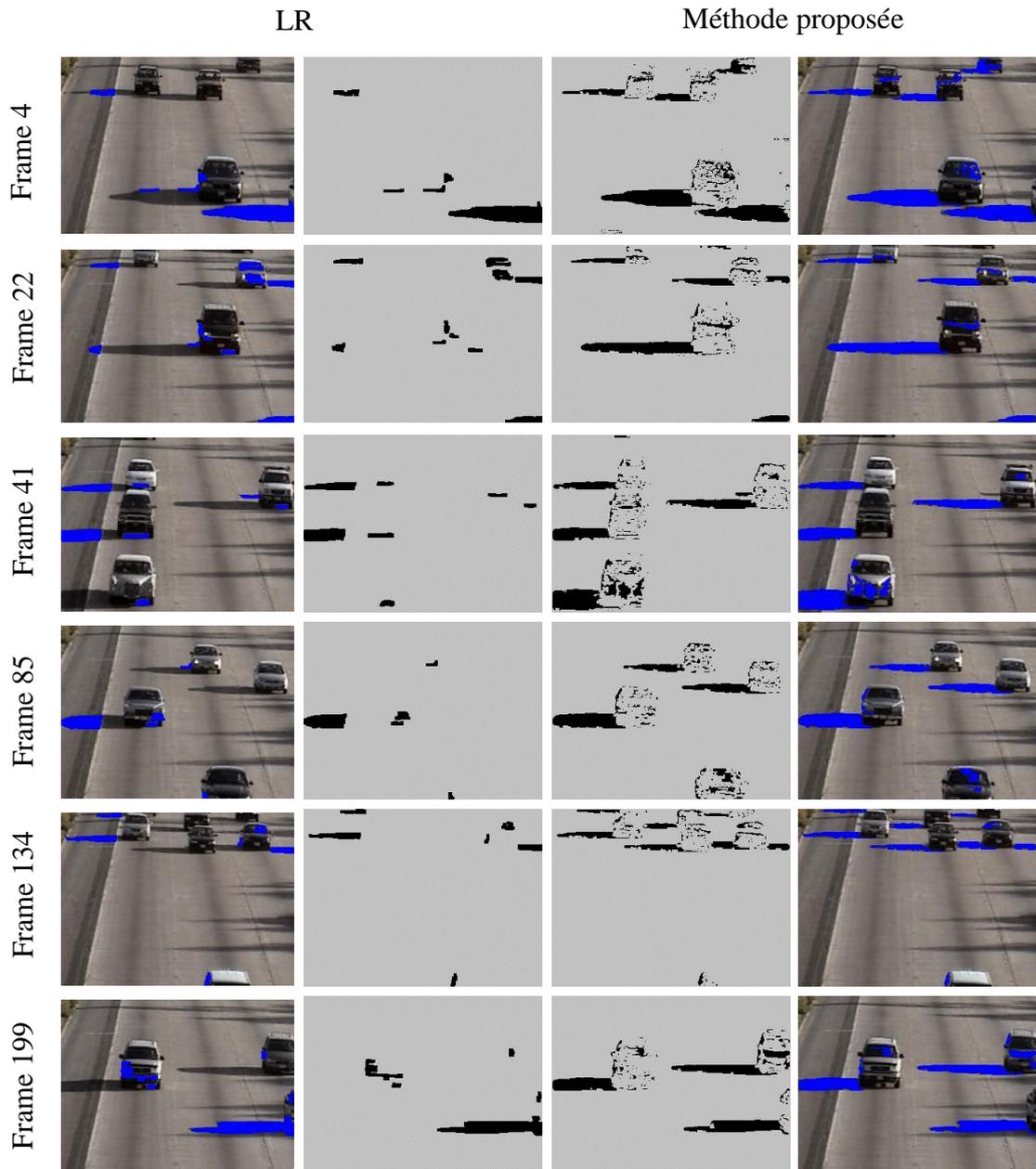


FIGURE 9.15 – Résultats de la détection de l’ombre dans la vidéo A2

En résumé, les résultats qualitatifs présentés ont montré que notre approche donne un résultat meilleur que celui des textures larges et cela presque dans toutes les vidéos utilisées. Ils montrent également que le système permet de séparer entre l’ombre propre et l’ombre portée. On peut conclure aussi que l’approche proposée est efficace sans aucune condition préalable sur la nature de l’environnement, la qualité des images, les variations de lumière ou de couleur.

9.7.2 Résultats quantitatifs

Nous avons utilisé une large sélection des méthodes de détection et d'extraction de l'ombre pour bien situer notre travail. Ce large choix de méthodes est dû à l'utilisation des mêmes bases de vidéos et l'utilisation des mêmes métriques d'évaluation à savoir le taux de détection η et le taux de discrimination ξ . Le tableau 9.4 compare les résultats obtenus par notre système avec les résultats obtenus par d'autres méthodes.

Méthodes	Highway I		Intelligent Room		Laboratory		Campus	
	$\eta\%$	$\xi\%$	$\eta\%$	$\xi\%$	$\eta\%$	$\xi\%$	$\eta\%$	$\xi\%$
Mikic[MIK 00]	59.59	84.70	76.27	90.74	64.85	95.39	72.43	74.08
Haritaoglu [HAR 00]	81.59	63.76	72.82	88.90	84.03	92.35	82.87	86.65
Cucchiara[CUC 02]	69.72	76.93	78.61	90.29	76.26	89.87	82.87	86.65
Stander [STA 99a]	75.49	62.38	62.00	93.89	60.34	81.57	69.10	62.96
Salvador [SAL 04]	71.82	79.29	73.45	86.52	88.24	93.57	72.4	72.4
Martel-B [MAR 05]	75.43	74.67	73.60	79.10	76.62	75.14	66.2	72.3
Al-Najdawi [ALN 06]	N/A	N/A	87.24	95.85	90.22	92.83	90.67	93.34
Horprasert [HOR 99]	N/A	N/A	72.82	88.90	84.03	92.35	80.58	69.37
Joshi [JOS 08]	88.21	97.00	91.02	97.66	N/A	N/A	N/A	N/A
Jung [JUN 09]	N/A	N/A	97.67	86.21	85.84	95.1	87.69	92.18
Zhang [ZHA 07b]	67.17	90.19	88.63	88.97	86.28	92.64	87.95	97.74
Siala [SIA 04]	83.30	68.92	N/A	N/A	N/A	N/A	N/A	N/A
Song [SON 07]	76.86	80.52	N/A	N/A	N/A	N/A	N/A	N/A
Martel-B [MAR 08]	72.10	79.70	N/A	N/A	N/A	N/A	N/A	N/A
Celik [CEL 08]	79.74	90.07	86.24	98.96	67.18	96.52	N/A	N/A
Choi [CHO 10]	84.98	88.97	95.01	91.39	90.63	94.00	N/A	N/A
Méthode proposée	79.29	98,78	93,15	96,39	95,39	98,92	95,55	96,68

TABLE 9.4 – Évaluations comparatives des résultats obtenus

Les résultats, présentés dans le tableau 9.4, montrent que notre système fournit à la fois un bon taux de détection et un bon taux de discrimination par rapport aux autres méthodes. Cependant, ces résultats ne reflètent pas la réalité. En effet, l'amélioration des performances de la détection de l'ombre est proportionnelle à l'amélioration des performances de la soustraction de l'arrière-plan. Donc pour tirer des conclusions approfondies, nous avons implémenté la méthode LR avec les mêmes prétraitements et avec le même module d'extraction de l'arrière-plan.

Les résultats quantitatifs, présentés dans le tableau 9.5, montrent clairement que le système proposé permet à la fois une bonne détection avec une moyenne de 91,37 % et une meilleure discrimination avec une moyenne de 95,98 % ce qui signifie que notre méthode a obtenu un gain de détection de 3,44 % et un gain de discrimination de 7,98 % par rapport à la méthode LR. Nous remarquons qu'il y a une chute de performance

notamment dans les vidéos Highway I et Highway III due à une mauvaise segmentation des contours.

Base de vidéos	Taux de détection (η %)		Taux de discrimination (ξ %)	
	LR	Proposée	LR	Proposée
Campus	89.63	95,55	92,93	96,68
Highway I	87,93	79.29	93,25	98,78
Intelligent room	90.19	93,15	88,14	96,39
Laboratory	87.13	95,39	92,67	98,92
Bootstrap	77.86	84,06	91,89	96,73
Campus	88.63	88,93	91,95	96,99
Fountain	88.91	90,73	89,94	97,46
Hall	86.27	89,66	82,91	95,07
Hallway	91.75	91,39	90,92	93,01
Highway III	66.60	79,08	89,94	94,68
Campus	88.2	96,77	77,88	97,98
Hallway I	86.8	91,47	95,80	96,57
Highway IV	86.52	94,28	68,95	89,52
Public Park	89.17	97,44	84,96	95,05
Moyenne	87,93	91,37	88,00	95,98

TABLE 9.5 – Comparaison de la méthode proposée avec la méthode LR dans les bases de données vidéo publiques et privées

Base de vidéos	LR (ms/frame)	Proposée (ms/frame)
Campus	10.76	4.72
Highway I	27.71	7.73
Intelligent room	6.25	4.14
Laboratory	12.73	4.95
Bootstrap	7.35	2.56
Campus	8.69	2.76
Fountain	8.27	2.23
Hall	4.8	1.98
Hallway	13.59	4.13
Highway III	4.75	2.82
Campus	23.79	7.2
Hallway	22.07	5.67
Highway	22.66	7.01
Public Park	22.43	6.44
Moyenne	13.99	4.6

TABLE 9.6 – Le temps de traitement calculé en millisecondes par image

Les résultats qualitatifs ont montré également cette diminution de performance vu que

notre système a très mal détecté les bords entre l'objet et son ombre. On remarque aussi que la méthode LR donne de bons résultats avec une stabilité de performance dans tous les cas de figure. En revanche, la discrimination est un peu faible forcément à cause de non-disponibilité de texture discriminative dans les cas traités.

En plus des résultats quantitatifs et qualitatifs, nous avons calculé le temps de traitement. Le tableau 9.6 montre que notre système consomme moins de temps par rapport à la méthode LR bien que les deux algorithmes ont la même complexité des opérations. Cela est dû à des étapes supplémentaires requises par la méthode LR pour produire les zones candidates d'ombre et pour calculer les gradients pour chaque pixel. La durée moyenne est réduite à presque un tiers.

9.8 Résultats relatifs à la classification et le suivi

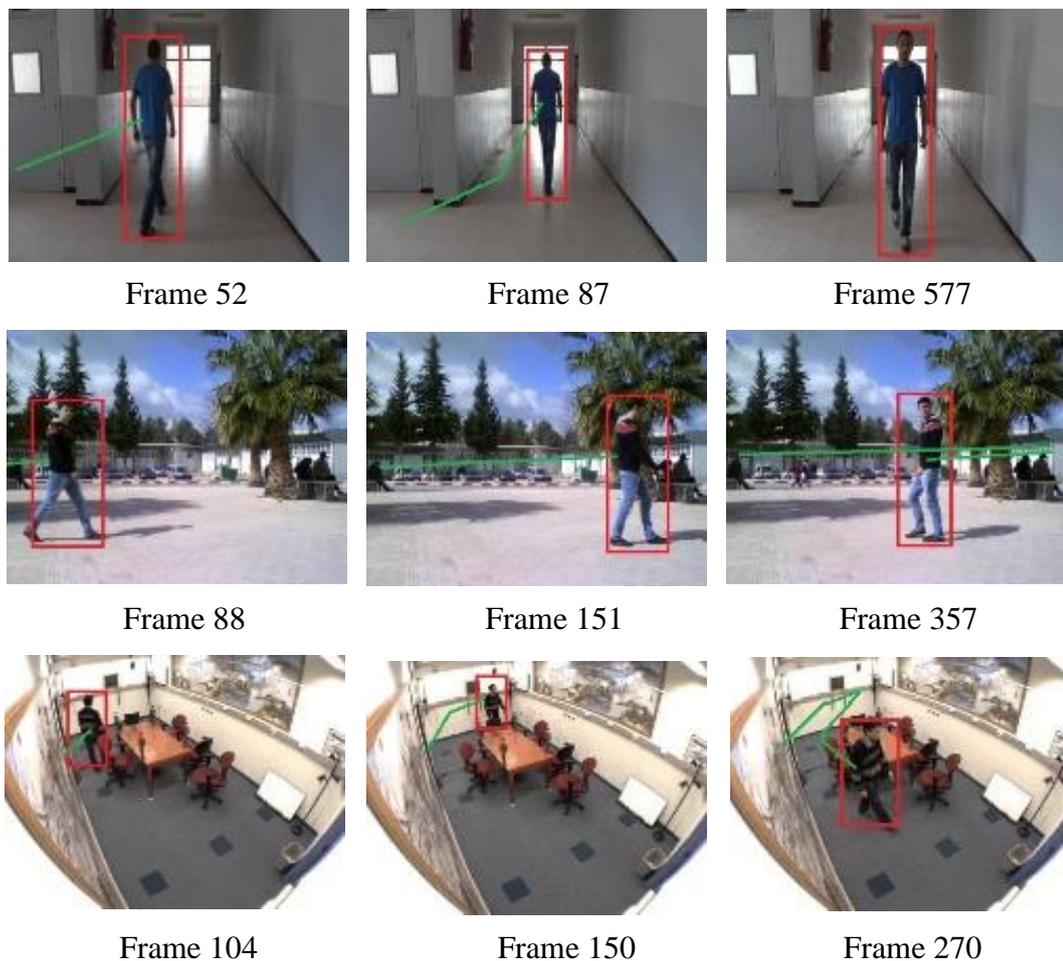


FIGURE 9.16 – Résultats du suivi sur des scènes monocible

Nous avons sélectionné des frames aléatoires pour présenter les résultats du suivi. Les vidéos utilisées pour l'évaluation se diffèrent par rapport à l'environnement en question,

par rapport au nombre d'objets en mouvement et par rapport à la nature des trajectoires des objets. Les cadres colorés en rouge représentent l'objet suivi tandis que le trait vert représente la trajectoire mémorisée du même objet à travers les frames précédents.

Les figures 9.16 et 9.17 montrent les résultats obtenus dans le cas d'une seule cible et dans le cas de plusieurs cibles. La figure 9.17 contient les situations les plus difficiles dans le suivi des objets en mouvement à savoir les collisions et la fusion des objets en mouvements.

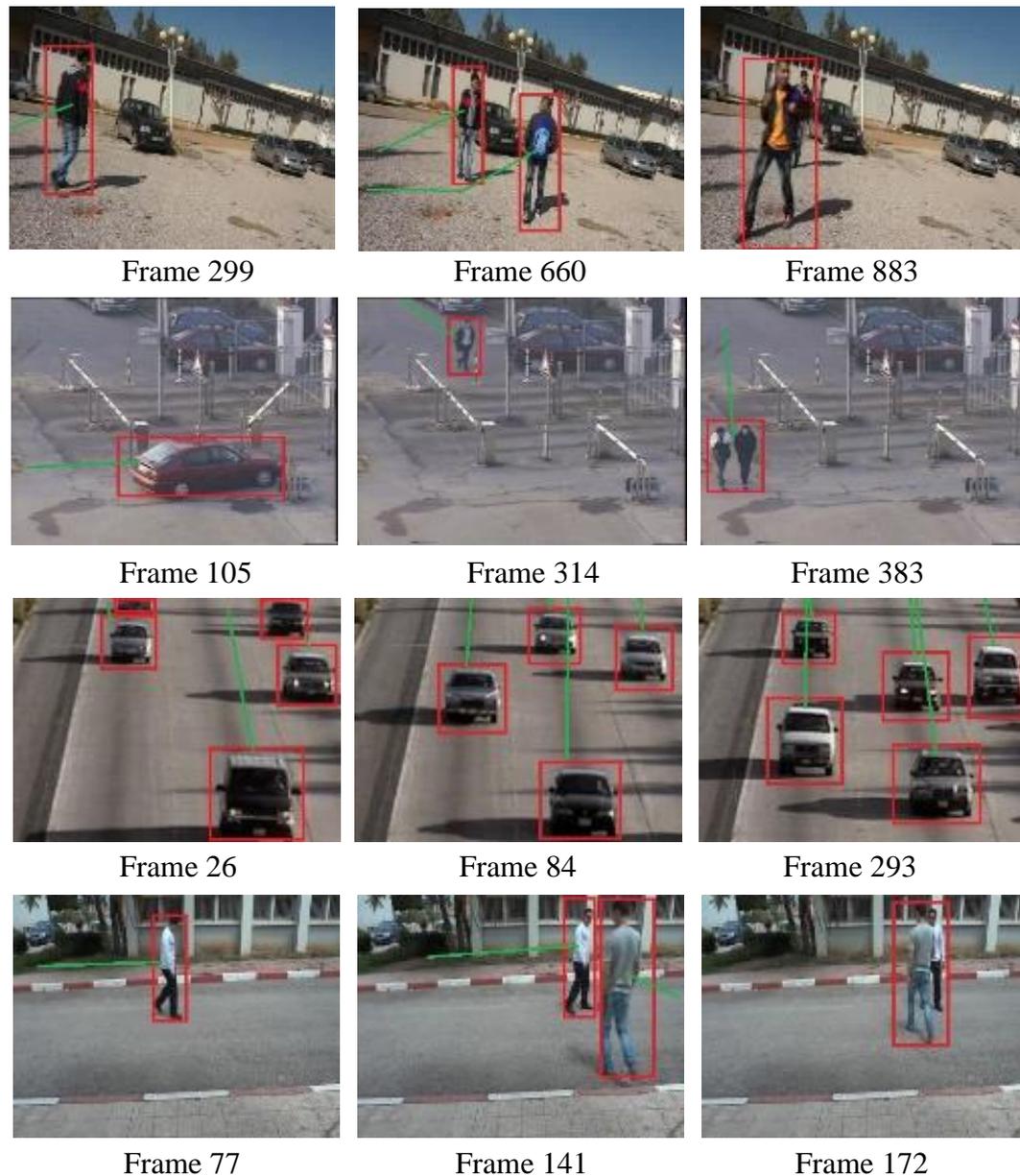


FIGURE 9.17 – Résultats du suivi sur des scènes multicibles

9.9 Conclusion

Nous avons présenté dans ce chapitre quelques résultats relatifs aux contributions proposées dans les étapes liées à l'extraction de l'arrière-plan, à la détection des ombres portées engendrées par les objets en mouvement et à la reconnaissance et le suivi des objets à travers les frames des vidéos. Nous avons proposé des résultats de nature qualitatifs et d'autres de nature quantitatifs pour une meilleure visibilité des améliorations apportées par les méthodes proposées.

Pour bien situer notre travail par rapport à ce qui existe dans l'état de l'art, nous avons également présenté une étude comparative avec plusieurs méthodes qui traitent les mêmes problèmes. Les résultats quantitatifs ont été calculés à partir des métriques d'évaluation et les vérités-terrains obtenues soit par les créateurs des bases vidéo utilisées ou par une annotation manuelle des vidéos personnelles.

Pour donner plus de crédibilité aux résultats, nous avons utilisé trois bases publiques et une personnelle avec un total de 35 vidéos représentant des environnements intérieurs et extérieurs avec des conditions aléatoires sur la nature du mouvement, la nature de l'environnement ainsi que le nombre d'objets en mouvement. Nous pouvons dire que les méthodes proposées ont dépassé en terme de qualité des résultats et temps d'exécution sur la majorité des méthodes proposées dans l'état de l'art et cela dans toutes les bases vidéos utilisées. Cependant, nos méthodes ont eu quelques difficultés à s'adapter avec certaines situations. Ces faiblesses sont dues principalement à des choix imposés par les contraintes de temps dans les systèmes de vidéosurveillance.

Conclusion Générale

Notre travail porte sur les techniques du Datamining relatifs à la détection et la reconnaissance des formes dans une vidéo. Les travaux présentés dans cette thèse tournent autour d'un type spécifique de vidéo et décrivent un système de vidéosurveillance intelligent destiné à la détection, la reconnaissance et le suivi des objets mobiles par une analyse automatique et autonome de la scène. Nous allons tenter à travers cette conclusion de dégager les contributions présentées et de proposer quelques perspectives envisageables.

La première partie de cette thèse a été consacrée à un état de l'art sur les différents travaux réalisés dans le cadre des systèmes de vidéosurveillance et cela dans tous les niveaux. Cette étude nous a permis de comprendre les problèmes relatifs à ce domaine et d'avoir une vue d'ensemble à la fois sur les méthodes qui ont donné des améliorations significatives et sur les axes de recherches qui n'ont pas eu un succès dans ce domaine malgré une très grande réussite dans d'autre domaine relatif à la reconnaissance des formes. Nous avons également critiqué les approches proposées selon la qualité des résultats obtenus, l'adaptabilité avec plusieurs environnements, la vitesse d'exécution, la robustesse et enfin la capacité mémoire utilisée.

Pour être le plus possible fidèle à l'architecture d'un système de vidéosurveillance, la première partie a été divisée en quatre chapitres représentant chacun un module principal dans un système de vidéosurveillance. Nous avons commencé par l'étape la plus importante dans un tel système à savoir l'extraction de l'arrière-plan. Sans être extravagant, l'extraction de l'arrière-plan est le pilier des systèmes de vision en général et plus particulièrement la vidéosurveillance, car il permet de différencier entre ce qui est considéré comme objet faisant partie de la structure de la scène et entre un objet intrus. Le deuxième chapitre porte sur la détection et la suppression des ombres portées engendrées par les objets en mouvement. En général, ce module ne fait pas partie de l'architecture des systèmes de vidéosurveillance. Cependant, c'est un prétraitement spécifique à certains types d'application qui permet de filtrer les données et faciliter les étapes ultérieures. En effet, la nature de l'ombre le rend indétectable par les méthodes d'extractions de l'arrière-plan vu qu'il partage les mêmes habitudes de déplacement que l'objet porteur. Pour cela, la détection et la suppression de l'ombre permettent d'obtenir un gain non négligeable dans la qualité des résultats obtenus.

Le domaine de la reconnaissance des formes a été introduit dans le troisième chapitre ou nous avons essayé de présenter les approches les plus communément utilisées dans les systèmes de vidéosurveillance. Selon le domaine d'application de la vidéosurveillance, la classification et la reconnaissance peuvent être soit une fin en soi ou une étape intermédiaire pour un éventuel suivi des objets détectés.

Le quatrième chapitre est un état de l'art sur les méthodes de suivi qui permettent d'analyser le comportement des objets en mouvement dans la scène. L'objectif du suivi se diffère d'une application à une autre et une méthode fonctionnelle sur un environnement particulier peut ne pas donner de bons résultats dans un autre environnement. Nous avons également présenté les caractéristiques les plus utilisées pour la représentation des objets et discuté leurs avantages et faiblesses selon le domaine d'application.

La deuxième partie de ce manuscrit est dédié à la proposition d'un système de reconnaissance et de suivi des formes pour le Datamining dans un système de vidéosurveillance. Nous avons divisé cette partie en cinq chapitres.

Nous avons introduit dans le cinquième chapitre la notion de vidéosurveillance et les caractéristiques liées à cette dernière. Nous avons ensuite présenté l'architecture générale de notre système et expliqué le rôle de chaque module qui le compose.

Le sixième chapitre porte sur une analyse spatio-temporelle pour la détection et l'extraction des objets en mouvement. Nous avons d'abord proposé une nouvelle taxonomie relative à l'extraction de l'arrière-plan. L'objectif de cette taxonomie est de donner une nouvelle plateforme de classification des approches proposées dans le domaine. Ce chapitre contient une nouvelle méthode d'extraction de l'arrière-plan pour la détection des objets mobiles dans des séquences d'images extraites à partir d'une caméra fixe. Cette méthode est une collaboration entre les mixtures de gaussiennes et les observateurs de l'arrière-plan. Afin d'aboutir à une performance maximale, nous avons effectué un passage de l'espace de couleur RGB vers l'espace de couleur HSV pour canaliser la luminosité dans un seul composant. La plupart des méthodes d'extraction de l'arrière-plan succombent devant la présence des variations locales et des variations soudaines de luminosité. Pour cette raison, nous avons proposé une approche qui permet de détecter les variations soudaines de luminosité en utilisant un appariement entre l'image en cours de traitement et un modèle de la scène bâti sur la base des gradients. Pour surmonter le problème de la variation local, nous avons divisé l'image en N régions et attribué à chacune d'elle un observateur d'arrière-plan. Ce dernier permet de sélectionner uniquement les régions qui ont subi un très grand changement en se basant sur l'analyse des histogrammes de couleur et le coefficient de corrélation de Bhattacharyya. L'extraction de l'arrière-plan laisse quelques bruits indésirables générés essentiellement par la présence de poussière et l'existence d'une similitude de couleur entre l'objet en mouvement et l'arrière-plan. Pour remédier à ces problèmes, nous avons proposé un ensemble d'opérations morphomathématiques pour raffiner davantage les résultats de l'extraction de l'arrière-plan. Les

post-traitements proposés respectent les contraintes de temps réel et de portabilité en minimisant au maximum la complexité algorithmique et l'espace mémoire nécessaire pour l'application de ces derniers.

Le septième chapitre porte sur une analyse spectrale pour la détection et la suppression des ombres portées générées par les objets en mouvement. La suppression de l'ombre n'est rien d'autre qu'un post-traitement spécifique à un problème d'une complexité supérieure. Pour aboutir à une détection efficace et robuste, nous avons proposé une méthode basée sur les espaces de couleurs RGB, HSV et YUV. Le plus grand défi était de réaliser un système qui sera capable d'opérer à la fois dans les environnements intérieurs et extérieurs et sans aucune hypothèse préalable sur les conditions d'éclairage. La majorité des méthodes de traitement de l'ombre ne permet pas de faire la différence entre les ombres portées et les ombres propres. La suppression des ombres propres déforme la forme de l'objet ce qui nous ramène au point de départ. Pour cette raison, nous avons appliqué le filtre de Canny pour détecter les limites entre les objets et les ombres. Ces informations sont utilisées comme critère pour distinguer entre les ombres portées et les ombres attachées. Nous avons utilisé trois critères calculés dans les espaces de couleur RGB, HSV et YUV pour décider si les pixels font partie des ombres ou des objets. Les pixels qui ont satisfait les trois critères sont classés en tant qu'ombre et seront supprimés de l'image par une mise à zéro.

Le huitième chapitre est une analyse comportementale sur les objets en mouvement détectés dans les étapes précédentes. Nous avons commencé par présenter les différentes modalités relatives à l'analyse comportementale des objets en mouvement. Dans la plupart des systèmes étudiés dans l'état de l'art, les processus de suivi et de reconnaissance sont totalement séparés. Cette architecture limite le bon fonctionnement du système pour une application longue durée et dans des environnements de nature aléatoire. Pour aboutir à un système de suivi performant, nous avons proposé une combinaison entre la classification, l'apprentissage et le suivi. L'utilisation de cette combinaison a permis de contourner les problèmes les plus difficiles du suivi tels que la gestion des occlusions et la fusion des objets en mouvement. L'apprentissage en ligne, avec les données étiquetées et les données non étiquetées, a permis d'augmenter la fiabilité du classifieur grâce à un mécanisme adaptatif et dynamique pour la gestion du changement de postures des objets en mouvement. Pour suivre les objets en mouvement à travers les frames de la vidéo, nous avons utilisé un filtre à particule qui se base sur les estimations de Kalman pour prédire les prochains emplacements possibles des objets en mouvement. Nous avons choisi comme méthode de classification l'algorithme des KPPV qui permet par conception, la possibilité d'effectuer un apprentissage en ligne adaptatif des paramètres du classifieur. Afin d'alléger le processus de suivi, nous avons utilisé les caractéristiques géométriques comme un critère de discrimination des objets en mouvement. Nous avons également mis en point une stratégie pour assurer à la fois une cohérence temporelle et une cohérence structurelle

afin de maximiser les performances du système de suivi.

Le dernier chapitre porte sur une étude expérimentale des approches proposées dans chaque niveau du système de vidéosurveillance proposé. Pour une meilleure visibilité des améliorations apportées, nous avons présenté des résultats visuels et métriques relatifs aux contributions proposées dans les étapes liées à l'extraction de l'arrière-plan, à la détection des ombres, à la reconnaissance et au suivi des objets. Ce chapitre contient également une étude comparative entre les approches proposées et les méthodes existantes dans l'état de l'art. Nous pouvons dire que les méthodes proposées ont dépassé, en terme de qualité des résultats et temps d'exécution, la majorité des méthodes proposées dans l'état de l'art et cela dans toutes les bases vidéos utilisées. Cependant, nos méthodes ont eu quelques difficultés à s'adapter avec certaines situations. Ces faiblesses sont dues principalement à des choix imposés par les contraintes de temps dans les systèmes de vidéosurveillance.

Malgré les efforts effectués afin de créer un système de vidéosurveillance intelligent et totalement autonome, ce dernier reste toujours loin de la perfection voulue et la présence d'un opérateur humain est indispensable pour le bon fonctionnement de ce dernier. Cependant, les systèmes de vidéosurveillance intelligents ont énormément contribué dans l'augmentation du rendement et l'amélioration de la qualité des résultats en secondant les opérateurs humains dans leurs tâches quotidiennes et répétitives. Comme perspective, nous proposons de sortir de la programmation ordinaire CPU vers une programmation spécifique GPU ou la puissance des cartes graphiques sera directement accessible via des langages de programmation dédiée et afin de tirer avantage de l'utilisation des algorithmes complexes qui sont, dans la situation actuelle, inutilisables à cause des contraintes de temps imposé par les systèmes de vidéosurveillance.

Bibliographie

- [ABD 06] Mohamed F ABDELKADER et al. « Integrated motion detection and tracking for visual surveillance ». In : *Computer Vision Systems, 2006 ICVS'06. IEEE International Conference on*. IEEE. 2006, p. 28–28.
- [ADA 06] Amit ADAM, Ehud RIVLIN et Ilan SHIMSHONI. « Robust fragments-based tracking using the integral histogram ». In : *Computer vision and pattern recognition, 2006 IEEE Computer Society Conference on*. T. 1. IEEE. 2006, p. 798–805.
- [AHM 11] KA AHMAD et al. « Improvement moving vehicle detection using RGB removal shadow segmentation ». In : *Control System, Computing and Engineering (ICCSCE), 2011 IEEE International Conference on*. IEEE. 2011, p. 22–26.
- [AKM 07] O AKMAN. « Multi-camera video surveillance detection, occlusion handling, tracking and event recognition ». Thèse de doct. Middle East Technical University, 2007.
- [ALB 12] John E ALBUS et al. *Syntactic pattern recognition, applications*. T. 14. Springer Science & Business Media, 2012.
- [ALI 01] Anjum ALI et JK AGGARWAL. « Segmentation and recognition of continuous human activity ». In : *Detection and recognition of events in video, 2001. Proceedings. IEEE Workshop on*. IEEE. 2001, p. 28–35.
- [ALL 07] Mohand Said ALLILI, Nizar BOUGUILA et Djemel ZIOU. « Finite generalized Gaussian mixture modeling and applications to image and video foreground segmentation ». In : *Computer and Robot Vision, 2007. CRV'07. Fourth Canadian Conference on*. IEEE. 2007, p. 183–190.
- [ALN 06] N AL-NAJDAWI, HE BEZ et EA EDIRISINGHE. « A novel approach for cast shadow modelling and detection ». In : (2006).
- [ALN 12a] Nijad AL-NAJDAWI et al. « A survey of cast shadow detection algorithms ». In : *Pattern Recognition Letters* 33.6 (2012), p. 752–764.
- [ALN 12b] Nijad AL-NAJDAWI et al. « An automated real-time people tracking system based on KLT features detection. » In : *Int. Arab J. Inf. Technol.* 9.1 (2012), p. 100–107.
- [AMI 13] Rahul AMIN et al. « Optical algorithm for cloud shadow detection over water ». In : *Geoscience and Remote Sensing, IEEE Transactions on* 51.2 (2013), p. 732–741.
- [ANG 01] Larisa ANGSTENBERGER. *Dynamic fuzzy pattern recognition with applications to finance and engineering*. Springer Science & Business Media, 2001.

- [ANG 04] Plamen P ANGELOV et Dimitar P FILEV. « Flexible models with evolving structure ». In : *International Journal of Intelligent Systems* 19.4 (2004), p. 327–340.
- [ANG 10] Plamen ANGELOV, Dimitar P FILEV et Nik KASABOV. *Evolving intelligent systems : methodology and applications*. T. 12. John Wiley & Sons, 2010.
- [AS 07] MA AS' ARI, UU SHEIKH et SAR ABU-BAKAR. « Object's shadow removal with removal validation ». In : *Signal Processing and Information Technology, 2007 IEEE International Symposium on*. IEEE. 2007, p. 841–845.
- [AVI 04] Shai AVIDAN. « Support vector tracking ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 26.8 (2004), p. 1064–1072.
- [AVI 07] Shai AVIDAN. « Ensemble tracking ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 29.2 (2007), p. 261–271.
- [AZA 10] Maha M AZAB, Howida SHEDEED, Ashraf S HUSSEIN et al. « A new technique for background modeling and subtraction for motion detection in real-time videos ». In : *Image Processing (ICIP), 2010 17th IEEE International Conference on*. IEEE. 2010, p. 3453–3456.
- [BAB 09] Boris BABENKO, Ming-Hsuan YANG et Serge BELONGIE. « Visual tracking with online multiple instance learning ». In : *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE. 2009, p. 983–990.
- [BAG 90] Giulia BAGALLO et David HAUSSLER. « Boolean feature discovery in empirical learning ». In : *Machine learning* 5.1 (1990), p. 71–99.
- [BEL 92] Abdel BELAID et Yolande BELAID. « Reconnaissance des formes : méthodes et applications ». In : (1992).
- [BER 00] Marcelo BERTALMIO, Guillermo SAPIRO et Gregory RANDALL. « Morphing active contours ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 22.7 (2000), p. 733–737.
- [BEZ 13] James C BEZDEK. *Pattern recognition with fuzzy objective function algorithms*. Springer Science & Business Media, 2013.
- [BIB 10] Charles BIBBY et Ian REID. « Real-time tracking of multiple occluding objects using level sets ». In : *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE. 2010, p. 1307–1314.
- [BIR 98] Stan BIRCHFIELD. « Elliptical head tracking using intensity gradients and color histograms ». In : *Computer Vision and Pattern Recognition, 1998. Proceedings. 1998 IEEE Computer Society Conference on*. IEEE. 1998, p. 232–237.
- [BIS 10] Adrian N BISHOP, Andrey V SAVKIN et Pubudu N PATHIRANA. « Vision-based target tracking and surveillance with robust set-valued state estimation ». In : *Signal Processing Letters, IEEE* 17.3 (2010), p. 289–292.
- [BLA 96] Michael J BLACK et Paul ANANDAN. « The robust estimation of multiple motions : Parametric and piecewise-smooth flow fields ». In : *Computer vision and image understanding* 63.1 (1996), p. 75–104.

- [BLA 98] Michael J BLACK et Allan D JEPSON. « Eigenttracking : Robust matching and tracking of articulated objects using a view-based representation ». In : *International Journal of Computer Vision* 26.1 (1998), p. 63–84.
- [BLU 98] Avrim BLUM et Tom MITCHELL. « Combining labeled and unlabeled data with co-training ». In : *Proceedings of the eleventh annual conference on Computational learning theory*. ACM. 1998, p. 92–100.
- [BOR 03] Silvio BORER. « New support vector algorithms for multicategorical data applied to real-time object recognition ». Thèse de doct. Citeseer, 2003.
- [BOS 92] Bernhard E BOSER, Isabelle M GUYON et Vladimir N VAPNIK. « A training algorithm for optimal margin classifiers ». In : *Proceedings of the fifth annual workshop on Computational learning theory*. ACM. 1992, p. 144–152.
- [BOU 99] Constantinos BOUKOUVALAS et al. « Color grading of randomly textured ceramic tiles using color histograms ». In : *IEEE Transactions on industrial electronics* 46.1 (1999), p. 219–226.
- [BOW 99] Kevin BOWYER, Christine KRANENBURG et Sean DOUGHERTY. « Edge detector evaluation using empirical ROC curves ». In : *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on*. T. 1. IEEE. 1999.
- [BRE 09] Michael D BREITENSTEIN et al. « Robust tracking-by-detection using a detector confidence particle filter ». In : *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE. 2009, p. 1515–1522.
- [BRE 84] Leo BREIMAN et al. *Classification and regression trees*. CRC press, 1984.
- [BRE 97] Leonard A BRESLOW et David W AHA. « Simplifying decision trees : A survey ». In : *The Knowledge Engineering Review* 12.01 (1997), p. 1–40.
- [BUE 08] Patrick BUEHLER et al. « Long term arm and hand tracking for continuous sign language TV broadcasts ». In : *Proceedings of the 19th British Machine Vision Conference*. BMVA Press. 2008, p. 1105–1114.
- [CAN 03] Erick CANTU-PAZ et Chandrika KAMATH. « Inducing oblique decision trees with evolutionary algorithms ». In : *Evolutionary Computation, IEEE Transactions on* 7.1 (2003), p. 54–68.
- [CAS 10] Rui CASEIRO, Joao F HENRIQUES et Jorge BATISTA. « Foreground segmentation via background modeling on Riemannian manifolds ». In : *Pattern Recognition (ICPR), 2010 20th International Conference on*. IEEE. 2010, p. 3570–3574.
- [CAV 05] Andrea CAVALLARO, Elena SALVADOR et Touradj EBRAHIMI. « Shadow-aware object-based video processing ». In : *Vision, Image and Signal Processing, IEE Proceedings*. T. 152. 4. IET. 2005, p. 398–406.
- [CAY 82] M CAYROL, H FARRENY et H PRADE. « Fuzzy pattern matching ». In : *Kybernetes* 11.2 (1982), p. 103–116.
- [CEL 08] Hasan CELIK et al. « Autonomous and adaptive learning of shadows for surveillance ». In : *Image Analysis for Multimedia Interactive Services, 2008. WIAMIS'08. Ninth International Workshop on*. IEEE. 2008, p. 59–62.

- [CHA 06] Olivier CHAPELLE, Bernhard SCHOLKOPF, Alexander ZIEN et al. « Semi-supervised learning ». In : (2006).
- [CHA 10] Theekapun CHAROENPONG, Ajaree SUPASUTTEKUL et Chaiwat NUTHONG. « Adaptive background modeling from an image sequence by using K-Means clustering ». In : *Electrical Engineering/Electronics Computer Telecommunications and Information Technology (ECTI-CON), 2010 International Conference on*. IEEE. 2010, p. 880–883.
- [CHA 99] Tat-Jen CHAM et James M REHG. « A multiple hypothesis approach to figure tracking ». In : *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on*. T. 2. IEEE. 1999.
- [CHE 01] Yunqiang CHEN, Yong RUI et Thomas S HUANG. « JPDAF based HMM for real-time contour tracking ». In : *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. T. 1. IEEE. 2001, p. I–543.
- [CHE 07] Tao CHEN, Julian MORRIS et Elaine MARTIN. « Gaussian process regression for multivariate spectroscopic calibration ». In : *Chemometrics and Intelligent Laboratory Systems* 87.1 (2007), p. 59–71.
- [CHE 10a] Chia-Chih CHEN et Jake K AGGARWAL. « Human shadow removal with unknown light source ». In : *Pattern Recognition (ICPR), 2010 20th International Conference on*. IEEE. 2010, p. 2407–2410.
- [CHE 10b] Chun-Ting CHEN, Chung-Yen SU et Wen-Chung KAO. « An enhanced segmentation on vision-based shadow removal for vehicle detection ». In : *Green Circuits and Systems (ICGCS), 2010 International Conference on*. IEEE. 2010, p. 679–682.
- [CHE 11] Fan-Chieh CHENG, Shih-Chia HUANG et Shanq-Jang RUAN. « Illumination-sensitive background modeling approach for accurate moving object detection ». In : *Broadcasting, IEEE Transactions on* 57.4 (2011), p. 794–801.
- [CHO 10] JinMin CHOI, Yung Jun YOO et Jin Young CHOI. « Adaptive shadow estimator for removing shadow of moving object ». In : *Computer Vision and Image Understanding* 114.9 (2010), p. 1017–1029.
- [COL 05] Robert T COLLINS, Yanxi LIU et Marius LEORDEANU. « Online selection of discriminative tracking features ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 27.10 (2005), p. 1631–1643.
- [COM 02a] Dorin COMANICIU. « Bayesian kernel tracking ». In : *Pattern Recognition*. Springer, 2002, p. 438–445.
- [COM 02b] Dorin COMANICIU et Peter MEER. « Mean shift : A robust approach toward feature space analysis ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24.5 (2002), p. 603–619.
- [COM 03] Dorin COMANICIU, Visvanathan RAMESH et Peter MEER. « Kernel-based object tracking ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 25.5 (2003), p. 564–577.
- [COR 11] Antoine CORNUÉJOLS et Laurent MICLET. *Apprentissage artificiel : concepts et algorithmes*. Editions Eyrolles, 2011.

- [COV 67] Thomas M COVER et Peter E HART. « Nearest neighbor pattern classification ». In : *Information Theory, IEEE Transactions on* 13.1 (1967), p. 21–27.
- [COX 96] Lngemar J COX et Sunita L HINGORANI. « An efficient implementation of Reid’s multiple hypothesis tracking algorithm and its evaluation for the purpose of visual tracking ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 18.2 (1996), p. 138–150.
- [COZ 03] Fabio Gagliardi COZMAN, Ira COHEN, Marcelo Cesar CIRELO et al. « Semi-supervised learning of mixture models ». In : *ICML. 2003*, p. 99–106.
- [CRE 03] Daniel CREMERS et Christoph SCHNORR. « Statistical shape knowledge in variational motion segmentation ». In : *Image and Vision Computing* 21.1 (2003), p. 77–86.
- [CUC 02] Rita CUCCHIARA et al. « The Sakbot system for moving object detection and tracking ». In : *Video-Based Surveillance Systems*. Springer, 2002, p. 145–157.
- [CUC 03] Rita CUCCHIARA et al. « Detecting moving objects, ghosts, and shadows in video streams ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 25.10 (2003), p. 1337–1342.
- [CVE 06] Sascha CVETKOVIC, Peter BAKKER, J SEHIRRIS et al. « Background estimation and adaptation model with light-change removal for heavily down-sampled video surveillance signals ». In : *Image Processing, 2006 IEEE International Conference on*. IEEE. 2006, p. 1829–1832.
- [DAS 91] Belur V DASARATHY. « Nearest neighbor norms : pattern classification techniques ». In : (1991).
- [DEN 10] Jinhao DENG et Juan ZHU. « Research on pedestrian detection algorithms based on video ». In : *Computer Design and Applications (ICCD), 2010 International Conference on*. T. 5. IEEE. 2010, p. V5–474.
- [DJO 90] Abdelhamid DJOUADI, OO SNORRASON et FD GARBER. « The quality of training sample estimates of the bhattacharyya coefficient ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 12.1 (1990), p. 92–97.
- [DON 11] Y DONG et Guilherme N DESOUZA. « Adaptive learning of multi-subspace for foreground detection under illumination changes ». In : *Computer Vision and Image Understanding* 115.1 (2011), p. 31–49.
- [DOU 10] Anastasis DOULAMIS et al. « Self Adaptive background modeling for identifying persons’ falls ». In : *Semantic Media Adaptation and Personalization (SMAP), 2010 5th International Workshop on*. IEEE. 2010, p. 57–63.
- [DOW 05] Nicholas DH DOWSON et Richard BOWDEN. « Simultaneous modeling and tracking (smat) of feature sets ». In : *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. T. 2. IEEE. 2005, p. 99–105.
- [DUB 90] D DUBUISSON. « Diagnostic et reconnaissance des formes ». In : (1990).
- [DUD 12] Richard O DUDA, Peter E HART et David G STORK. *Pattern classification*. John Wiley & Sons, 2012.

- [DUF 09] Franck DUFRENOIS, Johan COLLIEZ et Denis HAMAD. « Bounded influence support vector regression for robust single-model estimation ». In : *Neural Networks, IEEE Transactions on* 20.11 (2009), p. 1689–1706.
- [DUM 86] Vincențiu DUMITRU et Florica LUBAN. « On some optimization problems under uncertainty ». In : *Fuzzy Sets and Systems* 18.3 (1986), p. 257–272.
- [EL 09] Fida EL BAF, Thierry BOUWMANS et Bertrand VACHON. « Fuzzy statistical modeling of dynamic backgrounds for moving object detection in infrared videos ». In : *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on*. IEEE. 2009, p. 60–65.
- [ELG 00] Ahmed ELGAMMAL, David HARWOOD et Larry DAVIS. « Non-parametric model for background subtraction ». In : *Computer Vision–ECCV 2000*. Springer, 2000, p. 751–767.
- [ELS 09] Lamiaa M ELSHENAWY et al. « Efficient recursive principal component analysis algorithms for process monitoring ». In : *Industrial & Engineering Chemistry Research* 49.1 (2009), p. 252–259.
- [FAN 08] Liu Zhi FANG, Wang Yun QIONG et You Zhi SHENG. « A method to segment moving vehicle cast shadow based on wavelet transform ». In : *Pattern Recognition Letters* 29.16 (2008), p. 2182–2188.
- [FAR 11] Brahim FAROU et Hamid SERIDI. « Système de détection et de suivi des objets en mouvement dans une vidéo de surveillance ». In : *Proceedings of the International Conference on Signal, Image, Vision and their Applications SIVA'2011*. 2011, p. 272–276.
- [FAR 12] Brahim FAROU, Hamid SERIDI et Herman AKDAG. « Approche basée sur les mixtures de gaussiennes et les agents guetteurs pour l'extraction des objets en mouvement ». In : *Proceedings of 9^{ème} Colloque sur l'Optimisation et les Systèmes d'Information*. 2012, p. 28.
- [FAR 13a] Brahim FAROU, Hamid SERIDI et Herman AKDAG. « A New Approach for the Extraction of Moving Objects ». In : *Modeling Approaches and Algorithms for Advanced Computer Applications*. Springer, 2013, p. 27–36.
- [FAR 13b] Brahim FAROU, Hamid SERIDI et Herman AKDAG. « Using Gaussian Mixture Models and HSV color space for background subtraction ». In : *Control Engineering (CONTECH), 2013 International conference on*. 2013.
- [FAR 15a] Brahim FAROU, Hamid SERIDI et Herman AKDAG. « Improved Parameters Updating Algorithm for the Detection of Moving Objects ». In : *Computer Science and Its Applications*. Springer, 2015, p. 527–537.
- [FAR 15b] Brahim FAROU et al. « Une nouvelle approche pour l'élimination des ombres portées engendrées par les objets mobiles dans une vidéo de surveillance ». In : *Applied Automation and Industrial Diagnostics(ICAID), 2015 The 1st International Conference on*. 2015.
- [FER 03] Robert FERGUS, Pietro PERONA et Andrew ZISSERMAN. « Object class recognition by unsupervised scale-invariant learning ». In : *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*. T. 2. IEEE. 2003, p. II–264.

- [FIE 97] Paul FIEGUTH et Demctri TERZOPOULOS. « Color-based tracking of heads and other mobile objects at video frame rates ». In : *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*. IEEE. 1997, p. 21–27.
- [FIN 02] Graham D FINLAYSON, Steven D HORDLEY et Mark S DREW. « Removing shadows from images ». In : *Computer Vision–ECCV 2002*. Springer, 2002, p. 823–836.
- [FOR 03] David A FORSYTH et Jean PONCE. « A Modern Approach ». In : *Computer Vision : A Modern Approach* (2003).
- [FRA 99] Alexandre RJ FRANÇOIS et Gérard G MEDIONI. « Adaptive color background modeling for real-time segmentation of video streams ». In : *Proceedings of the International Conference on Imaging Science, Systems, and Technology*. T. 1. 121. 1999, p. 227–232.
- [FRE 89] Laurence S FREEDMAN et David J SPIEGELHALTER. « Comparison of Bayesian with group sequential methods for monitoring clinical trials ». In : *Controlled clinical trials* 10.4 (1989), p. 357–367.
- [FRE 97] Yoav FREUND et Robert E SCHAPIRE. « A decision-theoretic generalization of on-line learning and an application to boosting ». In : *Journal of computer and system sciences* 55.1 (1997), p. 119–139.
- [FRI 97] Nir FRIEDMAN et Stuart RUSSELL. « Image segmentation in video sequences : A probabilistic approach ». In : *Proceedings of the Thirteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc. 1997, p. 175–181.
- [FUN 01] George SK FUNG et al. « Effective moving cast shadow detection for monocular color image sequences ». In : *Image Analysis and Processing, 2001. Proceedings. 11th International Conference on*. IEEE. 2001, p. 404–409.
- [GAB 00] Bogdan GABRYS et Andrzej BARGIELA. « General fuzzy min-max neural network for clustering and classification ». In : *Neural Networks, IEEE Transactions on* 11.3 (2000), p. 769–783.
- [GAL 10] Jaime GALLEGO et Montse PARDAS. « Enhanced bayesian foreground segmentation using brightness and color distortion region-based model for shadow removal ». In : *Image Processing (ICIP), 2010 17th IEEE International Conference on*. IEEE. 2010, p. 3449–3452.
- [GAO 01] Dashan GAO et Jie ZHOU. « Adaptive background estimation for real-time traffic monitoring ». In : *Intelligent Transportation Systems, 2001. Proceedings. 2001 IEEE*. IEEE. 2001, p. 330–333.
- [GAO 08] Dongfa GAO, Zhuolin JIANG et Ming YE. « A new approach of dynamic background modeling for surveillance information ». In : *Computer Science and Software Engineering, 2008 International Conference on*. T. 1. IEEE. 2008, p. 850–855.
- [GAO 10] Ping GAO, Xiangju SUN et Wei WANG. « Moving object detection based on kirsch operator combined with Optical Flow ». In : *Image Analysis and Signal Processing (IASP), 2010 International Conference on*. IEEE. 2010, p. 620–624.

- [GER 86] Ron GERSHON, Allan D JEPSON et John K TSOTSOS. « Ambient illumination and the determination of material changes ». In : *JOSA A* 3.10 (1986), p. 1700–1707.
- [GEV 99] Theo GEVERS et Arnold WM SMEULDERS. « Color-based object recognition ». In : *Pattern recognition* 32.3 (1999), p. 453–464.
- [GRA 06] Helmut GRABNER et Horst BISCHOF. « On-line boosting and vision ». In : *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*. T. 1. IEEE. 2006, p. 260–267.
- [GRA 08] Helmut GRABNER, Christian LEISTNER et Horst BISCHOF. « Semi-supervised on-line boosting for robust tracking ». In : *Computer Vision–ECCV 2008*. Springer, 2008, p. 234–247.
- [GRA 10] Helmut GRABNER et al. « Tracking the invisible : Learning where the object might be ». In : *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE. 2010, p. 1285–1292.
- [GRE 94] Hayit GREENSPAN et al. « Overcomplete steerable pyramid filters and rotation invariance ». In : *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on*. IEEE. 1994, p. 222–228.
- [GRO 88] Stephen GROSSBERG. « Nonlinear neural networks : Principles, mechanisms, and architectures ». In : *Neural networks* 1.1 (1988), p. 17–61.
- [GUE 02] Yann GUERMEUR. « Combining discriminant models with new multi-class SVMs ». In : *Pattern Analysis & Applications* 5.2 (2002), p. 168–179.
- [HAF 10] Fadhlán HAFIZ et al. « Foreground segmentation-based human detection with shadow removal ». In : *Computer and Communication Engineering (ICCCE), 2010 International Conference on*. IEEE. 2010, p. 1–6.
- [HAN 07] Bohyung HAN et Ramesh JAIN. « Real-time subspace-based background modeling using multi-channel data ». In : *Advances in Visual Computing*. Springer, 2007, p. 162–172.
- [HAQ 10] Anwaar-ul HAQ, Iqbal GONDAL et Manzur MURSHED. « Automated multi-sensor color video fusion for nighttime video surveillance ». In : *Computers and Communications (ISCC), 2010 IEEE Symposium on*. IEEE. 2010, p. 529–534.
- [HAR 00] Ismail HARITAOGLU, David HARWOOD et Larry S DAVIS. « W 4 : Real-time surveillance of people and their activities ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 22.8 (2000), p. 809–830.
- [HAR 73] Robert M HARALICK, Karthikeyan SHANMUGAM et Its' Hak DINSTEIN. « Textural features for image classification ». In : *Systems, Man and Cybernetics, IEEE Transactions on* 6 (1973), p. 610–621.
- [HED 10] M HEDAYATI, WMDW ZAKI et Aini HUSSAIN. « Real-time background subtraction for video surveillance : From research to reality ». In : *Signal Processing and Its Applications (CSPA), 2010 6th International Colloquium on*. IEEE. 2010, p. 1–6.

- [HER 09] José Antonio Medina HERNÁNDEZ, FG CASTAEDA et José Antonio Moreno CADENAS. « An evolving fuzzy neural network based on the mapping of similarities ». In : *Fuzzy Systems, IEEE Transactions on* 17.6 (2009), p. 1379–1396.
- [HON 03] Dongpyo HONG et Woontack WOO. « A background subtraction for a vision-based user interface ». In : *Information, Communications and Signal Processing, 2003 and Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint Conference of the Fourth International Conference on*. T. 1. IEEE. 2003, p. 263–267.
- [HOP 86] John J HOPFIELD, David W TANK et al. « Computing with neural circuits-A model ». In : *Science* 233.4764 (1986), p. 625–633.
- [HOR 81] Berthold K HORN et Brian G SCHUNCK. « Determining optical flow ». In : *1981 Technical symposium east*. International Society for Optics et Photonics. 1981, p. 319–331.
- [HOR 99] Thanarat HORPRASERT, David HARWOOD et Larry S DAVIS. « A statistical approach for real-time robust background subtraction and shadow detection ». In : *IEEE ICCV*. T. 99. 1999, p. 1–19.
- [HSI 03] Jun-Wei HSIEH et al. « Shadow elimination for effective moving object detection by Gaussian shadow modeling ». In : *Image and Vision Computing* 21.6 (2003), p. 505–516.
- [HU 06] Jia-kai HU, Jian-guo JIANG et Mei-bin QI. « A Motion Detection Approach Based on Context Modeling ». In : *Video Engineering* 3 (2006), p. 023.
- [HUA 09] Jia-Bin HUANG et Chu-Song CHEN. « Moving cast shadow detection using physics-based features ». In : *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*. IEEE. 2009, p. 2310–2317.
- [HUE 02] Carine HUE, Jean-Pierre LE CADRE et Patrick PÉREZ. « Sequential Monte Carlo methods for multiple target tracking and data fusion ». In : *Signal Processing, IEEE Transactions on* 50.2 (2002), p. 309–325.
- [HUT 93] Daniel P HUTTENLOCHER, Jae J NOH et William J RUCKLIDGE. « Tracking non-rigid objects in complex scenes ». In : *Computer Vision, 1993. Proceedings., Fourth International Conference on*. IEEE. 1993, p. 93–101.
- [IAN 05] Codrut IANASI et al. « A fast algorithm for background tracking in video surveillance, using nonparametric kernel density estimation ». In : *Facta universitatis-series : Electronics and Energetics* 18.1 (2005), p. 127–144.
- [INT 97] Stephen S INTILLE, James W DAVIS et Aaron E BOBICK. « Real-time closed-world tracking ». In : *Computer Vision and Pattern Recognition, 1997. Proceedings., 1997 IEEE Computer Society Conference on*. IEEE. 1997, p. 697–703.
- [IRI 08] Kenji IRIE et al. « An investigation into noise-bound shadow detection and removal ». In : *Image and Vision Computing New Zealand, 2008. IVCNZ 2008. 23rd International Conference*. IEEE. 2008, p. 1–6.
- [ISA 01] Michael ISARD et John MACCORMICK. « BraMBLe : A Bayesian multiple-blob tracker ». In : *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*. T. 2. IEEE. 2001, p. 34–41.

- [ISA 98] Michael ISARD et Andrew BLAKE. « Condensation–conditional density propagation for visual tracking ». In : *International journal of computer vision* 29.1 (1998), p. 5–28.
- [JAV 02] Omar JAVED et Mubarak SHAH. « Tracking and object classification for automated surveillance ». In : *Computer Vision–ECCV 2002*. Springer, 2002, p. 343–357.
- [JAV 05] Omar JAVED, Saad ALI et Mubarak SHAH. « Online detection and classification of moving objects using progressively improving detectors ». In : *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. T. 1. IEEE. 2005, p. 696–701.
- [JEP 03] Allan D JEPSON, David J FLEET et Thomas F EL-MARAGHI. « Robust online appearance models for visual tracking ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 25.10 (2003), p. 1296–1311.
- [JIA 12] Bian JIANYONG, Yang RUNFENG et Yang YANG. « A novel vehicle’s shadow detection and removal algorithm ». In : *Consumer Electronics, Communications and Networks (CECNet), 2012 2nd International Conference on*. IEEE. 2012, p. 822–826.
- [JOL 02] Ian JOLLIFFE. *Principal component analysis*. Wiley Online Library, 2002.
- [JOS 08] Ajay J JOSHI et Nikolaos P PAPANIKOLOPOULOS. « Learning to detect moving shadows in dynamic environments ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 30.11 (2008), p. 2055–2063.
- [JUN 09] Cláudio Rosito JUNG. « Efficient Background Subtraction and Shadow Removal for Monochromatic Video Sequences. » In : *IEEE Transactions on Multimedia* 11.3 (2009), p. 571–577.
- [KAM 05] A Homayoun KAMKAR-PARSI, Robert LAGANIÈRE et Martin BOUCHARD. « A multi-criteria model for robust foreground extraction ». In : *Proceedings of the third ACM international workshop on Video surveillance and sensor networks*. ACM. 2005, p. 67–70.
- [KAM 96] Yoshinari KAMEDA et Michihiko MINOH. « A human motion estimation method using 3-successive video frames ». In : *International conference on virtual systems and multimedia*. 1996, p. 135–140.
- [KAN 03] Jinman KANG, Isaac COHEN et Gerard MEDIONI. « Continuous tracking within and across camera streams ». In : *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*. T. 1. IEEE. 2003, p. I–267.
- [KAN 04] Jinman KANG, Isaac COHEN et Gerard MEDIONI. « Object Reacquisition Using Invariant Appearance Model ». In : *Proceedings of the Pattern Recognition, 17th International Conference on (ICPR’04) Volume 4 - Volume 04*. ICPR ’04. Washington, DC, USA : IEEE Computer Society, 2004, p. 759–762. ISBN : 0-7695-2128-2. DOI : 10 . 1109 / ICPR . 2004 . 633. URL : <http://dx.doi.org/10.1109/ICPR.2004.633>.
- [KAN 10] Jiangming KAN et al. « Background modeling method based on improved multi-Gaussian distribution ». In : *Computer Application and System Modeling (ICASM), 2010 International Conference on*. T. 2. IEEE. 2010, p. V2–214.

- [KAW 06] Satoshi KAWABATA, Shinsaku HIURA et Kosuke SATO. « Real-time detection of anomalous objects in dynamic scene ». In : *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*. T. 3. IEEE. 2006, p. 1171–1174.
- [KEE 03] S Sathiya KEERTHI et Chih-Jen LIN. « Asymptotic behaviors of support vector machines with Gaussian kernel ». In : *Neural computation* 15.7 (2003), p. 1667–1689.
- [KEL 85] James M KELLER, Michael R GRAY et James A GIVENS. « A fuzzy k-nearest neighbor algorithm ». In : *Systems, Man and Cybernetics, IEEE Transactions on* 4 (1985), p. 580–585.
- [KIF 04] Daniel KIFER, Shai BEN-DAVID et Johannes GEHRKE. « Detecting change in data streams ». In : *Proceedings of the Thirtieth international conference on Very large data bases-Volume 30*. VLDB Endowment. 2004, p. 180–191.
- [KIM 02] Kwang In KIM, Keechul JUNG et Hang Joon KIM. « Face recognition using kernel principal component analysis ». In : *Signal Processing Letters, IEEE* 9.2 (2002), p. 40–42.
- [KIM 04] Kyungnain KIM et al. « Background modeling and subtraction by codebook construction ». In : *Image Processing, 2004. ICIP'04. 2004 International Conference on*. T. 5. IEEE. 2004, p. 3061–3064.
- [KIM 08] Hansung KIM et al. « Background subtraction using generalised Gaussian family model ». In : *Electronics letters* 44.3 (2008), p. 189–190.
- [KIM 10] Gyu-Jin KIM et al. « Automated measurement of crowd density based on edge detection and optical flow ». In : *Industrial Mechatronics and Automation (ICIMA), 2010 2nd International Conference on*. T. 2. IEEE. 2010, p. 553–556.
- [KOH 82] Teuvo KOHONEN. « Self-organized formation of topologically correct feature maps ». In : *Biological cybernetics* 43.1 (1982), p. 59–69.
- [KUN 00] Murat KUNT. *Reconnaissance des formes et analyse de scènes*. T. 3. PPUR presses polytechniques, 2000.
- [KWO 10] Junseok KWON et Kyoung Mu LEE. « Visual tracking decomposition ». In : *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE. 2010, p. 1269–1276.
- [LAI 10] Jin-Cyuan LAI, Shih-Shinh HUANG et Chien-Cheng TSENG. « Image-based vehicle tracking and classification on the highway ». In : *Green Circuits and Systems (ICGCS), 2010 International Conference on*. IEEE. 2010, p. 666–670.
- [LAW 80] Kenneth I LAWS. *Textured image segmentation*. Rapp. tech. DTIC Document, 1980.
- [LEI 07] Bastian LEIBE, Konrad SCHINDLER et Luc VAN GOOL. « Coupled detection and trajectory estimation for multi-object tracking ». In : *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. IEEE. 2007, p. 1–8.

- [LEI 09] Tao LEI, Yangyu FAN et Liandong LI. « The algorithm of moving human body detection based on region background modeling ». In : *Computer Network and Multimedia Technology, 2009. CNMT 2009. International Symposium on*. IEEE. 2009, p. 1–4.
- [LEI 10] Fei LEI et Xiaoxia ZHAO. « Adaptive background estimation of underwater using Kalman-Filtering ». In : *Image and Signal Processing (CISP), 2010 3rd International Congress on*. T. 1. IEEE. 2010, p. 64–67.
- [LEL 03] Boudewijn PF LELIEVELDT et al. « Multi-view active appearance models for consistent segmentation of multiple standard views : application to long- and short-axis cardiac MR images ». In : *International Congress Series*. T. 1256. Elsevier. 2003, p. 1141–1146.
- [LEO 06] Alessandro LEONE, Cosimo DISTANTE et Francesco BUCCOLIERI. « A shadow elimination approach in video-surveillance context ». In : *Pattern Recognition Letters* 27.5 (2006), p. 345–355.
- [LEO 07] Alessandro LEONE et Cosimo DISTANTE. « Shadow detection for moving objects based on texture analysis ». In : *Pattern Recognition* 40.4 (2007), p. 1222–1233.
- [LEP 05] Vincent LEPETIT, Pascal LAGGER et Pascal FUA. « Randomized trees for real-time keypoint recognition ». In : *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. T. 2. IEEE. 2005, p. 775–781.
- [LEV 03] Anat LEVIN, Paul VIOLA et Yoav FREUND. « Unsupervised improvement of visual detectors using cotraining ». In : *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*. IEEE. 2003, p. 626–633.
- [LI 01] Baoxin LI et al. « Model-based temporal object verification using video ». In : *Image Processing, IEEE Transactions on* 10.6 (2001), p. 897–908.
- [LI 04a] Liyuan LI et al. « Statistical modeling of complex backgrounds for foreground object detection ». In : *Image Processing, IEEE Transactions on* 13.11 (2004), p. 1459–1472.
- [LI 04b] Yongmin LI. « On incremental and robust subspace learning ». In : *Pattern recognition* 37.7 (2004), p. 1509–1518.
- [LI 08] Yuan LI et al. « Tracking in low frame rate video : A cascade particle filter with discriminative observers of different life spans ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 30.10 (2008), p. 1728–1740.
- [LI 10] Zhen LI et al. « Pedestrian Detection Based on a New Two-Step Framework ». In : *Education Technology and Computer Science (ETCS), 2010 Second International Workshop on*. T. 3. IEEE. 2010, p. 56–59.
- [LI 11] Xuejiao LI et Xiaojun JING. « FPGA based mixture Gaussian background modeling and motion detection ». In : *Natural Computation (ICNC), 2011 Seventh International Conference on*. T. 4. IEEE. 2011, p. 2078–2081.
- [LIM 00] Tjen-Sien LIM, Wei-Yin LOH et Yu-Shan SHIH. « A comparison of prediction accuracy, complexity, and training time of thirty-three old and new classification algorithms ». In : *Machine learning* 40.3 (2000), p. 203–228.

- [LIN 02] Horng-Horng LIN, Tyng-Luh LIU et Jen-Hui CHUANG. « A probabilistic SVM approach for background scene initialization ». In : *Image Processing, 2002. Proceedings. 2002 International Conference on*. T. 3. IEEE. 2002, p. 893–896.
- [LIN 10] Chih-Yang LIN et al. « Real-time robust background modeling based on joint color and texture descriptions ». In : *Genetic and Evolutionary Computing (ICGEC), 2010 Fourth International Conference on*. IEEE. 2010, p. 622–625.
- [LIN 91] Jianhua LIN. « Divergence measures based on the Shannon entropy ». In : *Information Theory, IEEE Transactions on* 37.1 (1991), p. 145–151.
- [LIU 07] Zhou LIU et al. « Cast shadow removal combining local and global features ». In : *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE. 2007, p. 1–8.
- [LOP 11] Miguel LOPEZ-BENITEZ et Fernando CASADEVALL. « Versatile, accurate, and analytically tractable approximation for the Gaussian Q-function ». In : *Communications, IEEE Transactions on* 59.4 (2011), p. 917–922.
- [LOW 04] David G LOWE. « Distinctive image features from scale-invariant keypoints ». In : *International journal of computer vision* 60.2 (2004), p. 91–110.
- [LUC 81] Bruce D LUCAS, Takeo KANADE et al. « An iterative image registration technique with an application to stereo vision. » In : *IJCAI*. T. 81. 1981, p. 674–679.
- [LUO 06] Xingzhi LUO et al. « Nonparametric background modeling using the CONDENSATION algorithm ». In : *Video and Signal Based Surveillance, 2006. AVSS'06. IEEE International Conference on*. IEEE. 2006, p. 3–3.
- [MA 06] Jie MA et Shutao LI. « Moving target detection based on background modeling by multi-level median filter ». In : *Intelligent Control and Automation, 2006. WCICA 2006. The Sixth World Congress on*. T. 2. IEEE. 2006, p. 9974–9978.
- [MAC 98] David JC MACKAY. « Introduction to monte carlo methods ». In : *Learning in graphical models*. Springer, 1998, p. 175–204.
- [MAH 06] Shyjan MAHAMUD. « Comparing belief propagation and graph cuts for novelty detection ». In : *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*. T. 1. IEEE. 2006, p. 1154–1159.
- [MAH 99] MA MAHJOUB. « Application des modèles de Markov caché stationnaires et nonstationnaires à la reconnaissance en ligne des caractères arabes ». Thèse de doct. Universitee des sciences des techniques et de médecine de Tunis II, 1999.
- [MAK 11] Aliaksei MAKARAU et al. « Adaptive shadow detection using a blackbody radiator model ». In : *Geoscience and Remote Sensing, IEEE Transactions on* 49.6 (2011), p. 2049–2059.
- [MAL 89] Stephane G MALLAT. « A theory for multiresolution signal decomposition : the wavelet representation ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 11.7 (1989), p. 674–693.

- [MAN 02] Abdol-Reza MANSOURI. « Region tracking via level set PDEs without motion computation ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24.7 (2002), p. 947–961.
- [MAO 04] Yan-Fen MAO et Peng-Fei SHI. « Multimodal background model with noise and shadow suppression for moving object detection ». In : *Journal of Southeast University (English Edition)* 20.4 (2004), p. 423–426.
- [MAR 05] Nicolas MARTEL-BRISSON et Andre ZACCARIN. « Moving cast shadow detection from a gaussian mixture shadow model ». In : *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. T. 2. IEEE. 2005, p. 643–648.
- [MAR 07] Nicolas MARTEL-BRISSON et Andre ZACCARIN. « Learning and removing cast shadows through a multidistribution approach ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 29.7 (2007), p. 1133–1146.
- [MAR 08] Nicolas MARTEL-BRISSON et André ZACCARIN. « Kernel-based learning of cast shadows from a physical model of light sources and surfaces for low-level segmentation ». In : *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*. IEEE. 2008, p. 1–8.
- [MAT 04] Iain MATTHEWS, Takahiro ISHIKAWA et Simon BAKER. « The template update problem ». In : *IEEE Transactions on Pattern Analysis & Machine Intelligence* 6 (2004), p. 810–815.
- [MCF 11] R MCFEELY et al. « Removal of non-uniform complex and compound shadows from textured surfaces using adaptive directional smoothing and the thin plate model ». In : *Image Processing, IET* 5.3 (2011), p. 233–248.
- [MCF 95] Nigel JB MCFARLANE et C Paddy SCHOFIELD. « Segmentation and tracking of piglets in images ». In : *Machine vision and applications* 8.3 (1995), p. 187–193.
- [MEN 11] Ainhoa MENDIZABAL et Luis SALGADO. « A region based approach to background modeling in a wavelet multi-resolution framework ». In : *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. IEEE. 2011, p. 929–932.
- [MEY 12] Sean P MEYN et Richard L TWEEDIE. *Markov chains and stochastic stability*. Springer Science & Business Media, 2012.
- [MIC 84] Laurent MICLET. « Méthodes structurelles pour la reconnaissance des formes ». In : (1984).
- [MIG 06] Davide A MIGLIORE, Matteo MATTEUCCI et Matteo NACCARI. « A re-valuation of frame difference in fast and robust motion detection ». In : *Proceedings of the 4th ACM international workshop on Video surveillance and sensor networks*. ACM. 2006, p. 215–218.
- [MIK 00] Ivana MIKIĆ et al. « Moving shadow and object detection in traffic scenes ». In : *Pattern Recognition, 2000. Proceedings. 15th International Conference on*. T. 1. IEEE. 2000, p. 321–324.
- [MOR 63] James N MORGAN et John A SONQUIST. « Problems in the analysis of survey data, and a proposal ». In : *Journal of the American statistical association* 58.302 (1963), p. 415–434.

- [MOR 73] James N MORGAN et Robert C MESSENGER. « THAID : A sequential analysis program for the analysis of nominal scale dependent variables ». In : (1973).
- [MUR 94] Sreerama K. MURTHY, Simon KASIF et Steven SALZBERG. « A system for induction of oblique decision trees ». In : *Journal of artificial intelligence research* (1994).
- [NAD 02] Sohail NADIMI et Bir BHANU. « Moving shadow detection using a physics-based approach ». In : *Pattern Recognition, 2002. Proceedings. 16th International Conference on*. T. 2. IEEE. 2002, p. 701–704.
- [NAD 04] Sohail NADIMI et Bir BHANU. « Physical models for moving shadow and object detection in video ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 26.8 (2004), p. 1079–1087.
- [NGH 07] Anh T NGHIEM et al. « ETISEO, performance evaluation for video surveillance systems ». In : *Advanced Video and Signal Based Surveillance, 2007. AVSS 2007. IEEE Conference on*. IEEE. 2007, p. 476–481.
- [NIC 06] Henri NICOLAS et Jean-Marie PINEL. « Joint moving cast shadows segmentation and light source detection in video sequences ». In : *Signal processing : Image communication* 21.1 (2006), p. 22–43.
- [NIG 00] Kamal NIGAM et al. « Text classification from labeled and unlabeled documents using EM ». In : *Machine learning* 39.2-3 (2000), p. 103–134.
- [NOO 11] Mohd NOOR et al. « Moving vehicle segmentation in a dynamic background using self-adaptive kalman background method ». In : *2011 IEEE 7th International Colloquium on Signal Processing and its Applications*. 2011, p. 439–442.
- [OBD 05] Stepan OBDRZALEK et Jiri MATAS. « Sub-linear Indexing for Large Scale Object Recognition. » In : *BMVC*. 2005, p. 1–10.
- [OKU 04] Kenji OKUMA et al. « A boosted particle filter : Multitarget detection and tracking ». In : *Computer Vision-ECCV 2004*. Springer, 2004, p. 28–39.
- [OLA 03] Cristina OLARU et Louis WEHENKEL. « A complete fuzzy decision tree technique ». In : *Fuzzy sets and systems* 138.2 (2003), p. 221–254.
- [OLI 00] Nuria M OLIVER, Barbara ROSARIO et Alex P PENTLAND. « A Bayesian computer vision system for modeling human interactions ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 22.8 (2000), p. 831–843.
- [PAN 08] Sinno Jialin PAN, James T KWOK et Qiang YANG. « Transfer Learning via Dimensionality Reduction. » In : *AAAI*. T. 8. 2008, p. 677–682.
- [PAN 13] Athanasios PANAGOPOULOS et al. « Simultaneous cast shadows, illumination and geometry inference using hypergraphs ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 35.2 (2013), p. 437–449.
- [PAR 02] Nikos PARAGIOS et Rachid DERICHE. « Geodesic active regions and level set methods for supervised texture segmentation ». In : *International Journal of Computer Vision* 46.3 (2002), p. 223–247.

- [PAS 01] George PASCHOS. « Perceptually uniform color spaces for color texture analysis : an empirical evaluation ». In : *Image Processing, IEEE Transactions on* 10.6 (2001), p. 932–937.
- [PIL 10] Julien PILET et Hideo SAITO. « Virtually augmenting hundreds of real pictures : An approach based on learning, retrieval, and tracking ». In : *Virtual Reality Conference (VR), 2010 IEEE*. IEEE. 2010, p. 71–78.
- [PIN 01] J PINE et Henri NICOLAS. « Estimation 2d illuminant direction and shadow segmentation in natural video sequences ». In : *Proceedings of VLAV*. Citeseer. 2001, p. 197.
- [POW 02] P Wayne POWER et Johann A SCHOONEES. « Understanding background mixture models for foreground segmentation ». In : *Proceedings image and vision computing New Zealand*. T. 2002. 2002, p. 10–11.
- [PRA 03a] Andrea PRATI et al. « Detecting moving shadows : algorithms and evaluation ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 25.7 (2003), p. 918–923.
- [PRA 03b] Andrea PRATI et al. « Detecting moving shadows : Formulation, algorithms and evaluation ». In : *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25 (2003), p. 2003.
- [QIN 10] Rui QIN et al. « Moving cast shadow removal based on local descriptors ». In : *Pattern Recognition (ICPR), 2010 20th International Conference on*. IEEE. 2010, p. 1377–1380.
- [QUI 11] Charles-Henri QUIVY et Itsuo KUMAZAWA. « Background images generation based on the nelder-mead simplex algorithm using the eigenbackground model ». In : *Image Analysis and Recognition*. Springer, 2011, p. 21–29.
- [RAG 93] Harish RAGAVAN et Larry A RENDELL. « Lookahead Feature Construction for Learning Hard Concepts. » In : *ICML*. 1993, p. 252–259.
- [RAH 08] Ali RAHIMI, Louis-Philippe MORENCY et Trevor DARRELL. « Reducing drift in differential tracking ». In : *Computer Vision and Image Understanding* 109.2 (2008), p. 97–111.
- [RAK 97] Ricco RAKOTOMALALA. « Graphes d’induction ». Thèse de doct. 1997.
- [RAM 07] Deva RAMANAN, David FORSYTH, Andrew ZISSERMAN et al. « Tracking people by learning their appearance ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 29.1 (2007), p. 65–81.
- [RAM 09] Chinta RAMBABU, KiyounG KIM et Woontack WOO. « Fast and accurate extraction of moving object silhouette for personalized Virtual Reality Studio at Home ». In : *Journal of Real-Time Image Processing* 4.4 (2009), p. 317–328.
- [RAS 01] Christopher RASMUSSEN et Gregory D HAGER. « Probabilistic data association methods for tracking complex visual objects ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 23.6 (2001), p. 560–576.

- [RID 95] Christof RIDDER, Olaf MUNKELT et Harald KIRCHNER. « Adaptive background estimation and foreground detection using kalman-filtering ». In : *Proceedings of International Conference on recent Advances in Mechatronics*. Citeseer. 1995, p. 193–199.
- [RIP 96] Brian D. RIPLEY. *Pattern recognition and neural networks*. Cambridge university press, 1996.
- [RON 94] Rémi RONFARD. « Region-based strategies for active contour models ». In : *International journal of computer vision* 13.2 (1994), p. 229–251.
- [ROS 05] Chuck ROSENBERG, Martial HEBERT et Henry SCHNEIDERMAN. « Semi-supervised self-training of object detection models ». In : (2005).
- [ROS 08] David A ROSS et al. « Incremental learning for robust visual tracking ». In : *International Journal of Computer Vision* 77.1-3 (2008), p. 125–141.
- [ROS 58] Frank ROSENBLATT. « The perceptron : a probabilistic model for information storage and organization in the brain. » In : *Psychological review* 65.6 (1958), p. 386.
- [ROS 99] Romer ROSALES et Stan SCLAROFF. « 3D trajectory recovery for tracking multiple objects and trajectory guided recognition of actions ». In : *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on*. T. 2. IEEE. 1999.
- [RUM 86] David E RUMELHART et James L MCCLELLAND. « The PDP Research Group : Parallel distributed processing : Explorations in the microstructure of cognition ». In : *Foundations* 1 (1986).
- [RUP 01] Stefan RUPING. « Incremental learning with support vector machines ». In : *icdm*. IEEE. 2001, p. 641.
- [SAL 01] Jesper SALOMON. « Support vector machines for phoneme classification ». Thèse de doct. 2001.
- [SAL 04] Elena SALVADOR, Andrea CAVALLARO et Touradj EBRAHIMI. « Cast shadow segmentation using invariant color features ». In : *Computer vision and image understanding* 95.2 (2004), p. 238–259.
- [SAL 10] Imran SALEEMI, Lance HARTUNG et Mubarak SHAH. « Scene understanding by statistical modeling of motion patterns ». In : *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE. 2010, p. 2069–2076.
- [SAL 90] V SALARI et Ishwar K. SETHI. « Feature point correspondence in the presence of occlusion ». In : *IEEE Transactions on Pattern Analysis & Machine Intelligence* 1 (1990), p. 87–91.
- [SAN 08] Peter SAND et Seth TELLER. « Particle video : Long-range motion estimation using point trajectories ». In : *International Journal of Computer Vision* 80.1 (2008), p. 72–91.
- [SAN 10] Andres SANIN, Conrad SANDERSON et Brian C LOVELL. « Improved shadow removal for robust person tracking in surveillance scenarios ». In : *Pattern Recognition (ICPR), 2010 20th International Conference on*. IEEE. 2010, p. 141–144.

- [SAN 12] Andres SANIN, Conrad SANDERSON et Brian C LOVELL. « Shadow detection : A survey and comparative evaluation of recent methods ». In : *Pattern recognition* 45.4 (2012), p. 1684–1695.
- [SAT 04] Koichi SATO et Jake K AGGARWAL. « Temporal spatio-velocity transform and its application to tracking and interaction ». In : *Computer Vision and Image Understanding* 96.2 (2004), p. 100–128.
- [SCH 02a] Oliver SCHREER et al. « Fast and robust shadow detection in videoconference applications ». In : *Video/Image Processing and Multimedia Communications 4th EURASIP-IEEE Region 8 International Symposium on VI-PromCom*. IEEE. 2002, p. 371–375.
- [SCH 02b] Haim SCHWEITZER, JW BELL et Feng WU. « Very fast template matching ». In : *Computer Vision–ECCV 2002*. Springer, 2002, p. 358–372.
- [SER 04] David SERBY, EK MEIER et Luc VAN GOOL. « Probabilistic object tracking using multiple features ». In : *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*. T. 2. IEEE. 2004, p. 184–187.
- [SET 87] Ishwar K SETHI et Ramesh JAIN. « Finding trajectories of feature points in a monocular image sequence ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 1 (1987), p. 56–73.
- [SHA 05] Khurram SHAFIQUE et Mubarak SHAH. « A noniterative greedy algorithm for multiframe point correspondence ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 27.1 (2005), p. 51–65.
- [SHA 11] Munir SHAH, Jeremiah DENG et Brendon WOODFORD. « Enhanced codebook model for real-time background subtraction ». In : *Neural Information Processing*. Springer. 2011, p. 449–458.
- [SHA 85] Steven A SHAFER. « Using color to separate reflection components ». In : *Color Research & Application* 10.4 (1985), p. 210–218.
- [SHE 05] Yaser SHEIKH et Mubarak SHAH. « Bayesian modeling of dynamic scenes for object detection ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 27.11 (2005), p. 1778–1792.
- [SHI 94] Jianbo SHI et Carlo TOMASI. « Good features to track ». In : *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on*. IEEE. 1994, p. 593–600.
- [SIA 04] Kais SIALA et al. « Moving shadow detection with support vector domain description in the color ratios space ». In : *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*. T. 4. IEEE. 2004, p. 384–387.
- [SKO 08] Danijel SKOČAJ et Aleš LEONARDIS. « Incremental and robust learning of subspace representations ». In : *Image and Vision Computing* 26.1 (2008), p. 27–38.
- [SOF 08] Michal SOFKA. « Commentary Paper on Shadow Removal in Indoor Scenes ». In : *Advanced Video and Signal Based Surveillance, 2008. AVSS'08. IEEE Fifth International Conference on*. IEEE. 2008, p. 299–300.

- [SON 07] Kai-Tai SONG et Jen-Chao TAI. « Image-based traffic monitoring with shadow suppression ». In : *Proceedings of the IEEE 95.2* (2007), p. 413–426.
- [SPA 06] Paolo SPAGNOLO, M LEO, A DISTANTE et al. « Moving object segmentation by background subtraction and temporal analysis ». In : *Image and Vision Computing 24.5* (2006), p. 411–423.
- [STA 99a] J STANDER, Roland MECH et Jorn OSTERMANN. « Detection of moving cast shadows for object segmentation ». In : *Multimedia, IEEE Transactions on 1.1* (1999), p. 65–76.
- [STA 99b] Chris STAUFFER et W Eric L GRIMSON. « Adaptive background mixture models for real-time tracking ». In : *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on. T. 2. IEEE. 1999.*
- [STE 00] Mikkel Bille STEGMANN. « Active appearance models : Theory, extensions and cases ». In : *Richard Petersens Plads, Building 321* (2000), p. 262.
- [STR 12] Matthias STRAKA et al. « Simultaneous shape and pose adaption of articulated models using linear optimization ». In : *Computer Vision–ECCV 2012. Springer, 2012, p. 724–737.*
- [STR 94] Roy L STREIT et Tod E LUGINBUHL. « Maximum likelihood method for probabilistic multihypothesis tracking ». In : *SPIE’s International Symposium on Optical Engineering and Photonics in Aerospace Sensing. International Society for Optics et Photonics. 1994, p. 394–405.*
- [SUN 10] Bangyu SUN et Shutao LI. « Moving cast shadow detection of vehicle using combined color models ». In : *Pattern Recognition (CCPR), 2010 Chinese Conference on. IEEE. 2010, p. 1–5.*
- [SUO 08] Peng SUO et Yanjiang WANG. « An improved adaptive background modeling algorithm based on Gaussian Mixture Model ». In : *Signal Processing, 2008. ICSP 2008. 9th International Conference on. IEEE. 2008, p. 1436–1439.*
- [SUY 99] Johan AK SUYKENS et Joos VANDEWALLE. « Least squares support vector machine classifiers ». In : *Neural processing letters 9.3* (1999), p. 293–300.
- [SZE 97] Richard SZELISKI et James COUGHLAN. « Spline-based image registration ». In : *International Journal of Computer Vision 22.3* (1997), p. 199–218.
- [TAN 07a] Tatsuya TANAKA et al. « A fast algorithm for adaptive background model construction using parzen density estimation ». In : *Advanced Video and Signal Based Surveillance, 2007. AVSS 2007. IEEE Conference on. IEEE. 2007, p. 528–533.*
- [TAN 07b] Feng TANG et al. « Co-tracking using semi-supervised support vector machines ». In : *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on. IEEE. 2007, p. 1–8.*
- [TAN 07c] Peng TANG, Lin GAO et Zhifang LIU. « Salient moving object detection using stochastic approach filtering ». In : *Image and Graphics, 2007. ICIG 2007. Fourth International Conference on. IEEE. 2007, p. 530–535.*

- [TAO 02] Hai TAO, Harpreet S SAWHNEY et Rakesh KUMAR. « Object tracking with bayesian estimation of dynamic layer representations ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 24.1 (2002), p. 75–89.
- [TAV 05] A TAVAKKOLI. « Automatic Video Object Plane Extraction using non-Parametric Kernel Density Estimation ». In : *Mathematical Methods in Computer Vision, University of Nevada, Reno, NV* (2005).
- [TAV 06a] Alireza TAVAKKOLI, Mircea NICOLESCU et George BEBIS. « A novelty detection approach for foreground region detection in videos with quasi-stationary backgrounds ». In : *Advances in Visual Computing*. Springer, 2006, p. 40–49.
- [TAV 06b] Alireza TAVAKKOLI, Mircea NICOLESCU et George BEBIS. « Automatic statistical object detection for visual surveillance ». In : *proceedings of IEEE Southwest Symposium on Image Analysis and Interpretation*. 2006, p. 144–148.
- [TAV 06c] Alireza TAVAKKOLI, Mircea NICOLESCU et George BEBIS. « Robust recursive learning for foreground region detection in videos with quasi-stationary backgrounds ». In : *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*. T. 1. IEEE. 2006, p. 315–318.
- [TAV 14] Mircea Nicolescu TAVAKKOLI, Junxian WANG et George BEBIS. « Applications in Video Surveillance ». In : *Background Modeling and Foreground Detection for Video Surveillance* (2014).
- [TAY 09] Simon TAYLOR et Tom DRUMMOND. « Multiple Target Localisation at over 100 FPS. » In : *BMVC*. 2009, p. 1–11.
- [TER 93] Demetri TERZOPOULOS et Richard SZELISKI. « Tracking with Kalman snakes ». In : *Active vision*. MIT press. 1993, p. 3–20.
- [TIA 05] Ying-Li TIAN, Max LU et Arun HAMPAPUR. « Robust and efficient foreground analysis for real-time video surveillance ». In : *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. T. 1. IEEE. 2005, p. 1182–1187.
- [TOY 99] Kentaro TOYAMA et al. « Wallflower : Principles and practice of background maintenance ». In : *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*. T. 1. IEEE. 1999, p. 255–261.
- [UZA 09] Muhammad UZAIR et al. « Background modeling using corner features : An effective approach ». In : *Multitopic Conference, 2009. INMIC 2009. IEEE 13th International*. IEEE. 2009, p. 1–5.
- [VAC 04] Luca VACCHETTI, Vincent LEPETIT et Pascal FUA. « Stable real-time 3d tracking using online and offline information ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 26.10 (2004), p. 1385–1391.
- [VAP 13] Vladimir VAPNIK. *The nature of statistical learning theory*. Springer Science & Business Media, 2013.
- [VAP 98] Vladimir Naumovich VAPNIK et Vladimir VAPNIK. *Statistical learning theory*. T. 1. Wiley New York, 1998.

- [VAS 03] Namrata VASWANI, Amit Roy CHOWDHURY et Rama CHELLAPPA. « Activity recognition using the dynamics of the configuration of interacting objects ». In : *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*. T. 2. IEEE. 2003, p. II–633.
- [VEE 01] Cor J VEENMAN, Marcel JT REINDERS et Eric BACKER. « Resolving motion correspondence for densely moving points ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 23.1 (2001), p. 54–72.
- [VEM 09] Raviteja VEMULAPALLI et R ARAVIND. « Spatio-temporal nonparametric background modeling and subtraction ». In : *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*. IEEE. 2009, p. 1145–1152.
- [VIL 86] Felicia M VILNROTTER, Ramakant NEVATIA et Keith E PRICE. « Structural analysis of natural textures ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 1 (1986), p. 76–89.
- [VIO 01] Paul VIOLA et Michael JONES. « Rapid object detection using a boosted cascade of simple features ». In : *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*. T. 1. IEEE. 2001, p. I–511.
- [WAN 03] Liang WANG, Weiming HU et Tieniu TAN. « Recent developments in human motion analysis ». In : *Pattern recognition* 36.3 (2003), p. 585–601.
- [WAN 08a] Huan WANG, Ming-wu REN et Jing-Yu YANG. « Background modeling method based on sequential kernel density approximation ». In : *Pattern Recognition, 2008. CCPR'08. Chinese Conference on*. IEEE. 2008, p. 1–6.
- [WAN 08b] Weiqiang WANG, Jie YANG et Wen GAO. « Modeling background and segmenting moving objects from compressed video ». In : *Circuits and Systems for Video Technology, IEEE Transactions on* 18.5 (2008), p. 670–681.
- [WAN 09] Junxian WANG et al. « Improving target detection by coupling it with tracking ». In : *Machine Vision and Applications* 20.4 (2009), p. 205–223.
- [WAN 11] Ning WANG, Congyan LANG et De XU. « Image-Based Shadow Removal via Illumination Chromaticity Estimation ». In : *Multimedia Information Networking and Security (MINES), 2011 Third International Conference on*. IEEE. 2011, p. 33–36.
- [WID 60] Bernard WIDROW, Marcian E HOFF et al. « Adaptive switching circuits. » In : (1960).
- [WIL 05] Oliver WILLIAMS, Andrew BLAKE et Roberto CIPOLLA. « Sparse bayesian learning for efficient visual tracking ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 27.8 (2005), p. 1292–1304.
- [WIT 07] Sue D WITHERSPOON et Ming ZHANG. « Negative coefficient polynomial kernel density estimation for visualization ». In : *Proceedings of the 18th conference on Proceedings of the 18th IASTED International Conference : modelling and simulation*. ACTA Press. 2007, p. 398–403.
- [WOO 10] Hyenkyun WOO et al. « Environmentally robust motion detection for video surveillance ». In : *Image Processing, IEEE Transactions on* 19.11 (2010), p. 2838–2848.

- [WRE 97] Christopher Richard WREN et al. « Pfinder : Real-time tracking of the human body ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 19.7 (1997), p. 780–785.
- [WU 10] Xiaoyu WU, Lei YANG et Cheng YANG. « Real-time foreground segmentation based on a fused background model ». In : *Computer and Automation Engineering (ICCAE), 2010 The 2nd International Conference on*. T. 2. IEEE. 2010, p. 585–588.
- [XU 05] Dong XU et al. « Cast shadow detection in video segmentation ». In : *Pattern Recognition Letters* 26.1 (2005), p. 91–99.
- [XU 06] Zhifei XU, Pengfei SHI et Irene Yu-Hua GU. « An eigenbackground subtraction method using recursive error compensation ». In : *Advances in Multimedia Information Processing-PCM 2006*. Springer, 2006, p. 779–787.
- [XU 08] Zhifei XU, Irene Yu-Hua GU et Pengfei SHI. « Recursive error-compensated dynamic eigenbackground learning and adaptive background subtraction in video ». In : *Optical Engineering* 47.5 (2008), p. 057001–057001.
- [XU 09] Shaoqiu XU. « Dynamic background modeling for foreground segmentation ». In : *Computer and Information Science, 2009. ICIS 2009. Eighth IEEE/ACIS International Conference on*. IEEE. 2009, p. 599–604.
- [XU 10] Huazhong XU, Pei LV et Lei MENG. « A people counting system based on head-shoulder detection and tracking in surveillance video ». In : *Computer Design and Applications (ICCD), 2010 International Conference on*. T. 1. IEEE. 2010, p. V1–394.
- [XU 12] Ming XU et al. « Cast shadow removal in motion detection by exploiting multiview geometry ». In : *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 2012.
- [XUE 08] Song XUEHUA et al. « Shadow removal of vehicles in a video system based on RGB chroma model ». In : *Computer Science and Software Engineering, 2008 International Conference on*. T. 1. IEEE. 2008, p. 977–980.
- [YAN 09] Ming YANG, Ying WU et Gang HUA. « Context-aware visual tracking ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 31.7 (2009), p. 1195–1209.
- [YAN 12] Qingxiong YANG, Kar-Han TAN et Narendra AHUJA. « Shadow removal using bilateral filtering ». In : *Image Processing, IEEE Transactions on* 21.10 (2012), p. 4361–4368.
- [YIL 04] Alper YILMAZ, Xin LI et Mubarak SHAH. « Contour-based object tracking with occlusion handling in video acquired using mobile cameras ». In : *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 26.11 (2004), p. 1531–1536.
- [YIL 06] Alper YILMAZ, Omar JAVED et Mubarak SHAH. « Object tracking : A survey ». In : *Acm computing surveys (CSUR)* 38.4 (2006), p. 13.
- [YU 08] Qian YU, Thang Ba DINH et Gérard MEDIONI. « Online tracking and reacquisition using co-trained generative and discriminative trackers ». In : *Computer Vision–ECCV 2008*. Springer, 2008, p. 678–691.

- [YU 10a] Hai-Yan YU et al. « MSER based shadow detection in high resolution remote sensing image ». In : *Machine Learning and Cybernetics (ICMLC), 2010 International Conference on*. T. 2. IEEE. 2010, p. 780–783.
- [YU 10b] Jin YU, Xuan ZHOU et Feng QIAN. « Object kinematic model : A novel approach of adaptive background mixture models for video segmentation ». In : *Intelligent Control and Automation (WCICA), 2010 8th World Congress on*. IEEE. 2010, p. 6225–6228.
- [YUK 11] Jacky S-C YUK et Kwan-Yee K WONG. « An efficient pattern-less background modeling based on scale invariant local states ». In : *Advanced Video and Signal-Based Surveillance (AVSS), 2011 8th IEEE International Conference on*. IEEE. 2011, p. 285–290.
- [ZHA 02a] M ZHAO, N LI et C CHEN. « Robust automatic video object segmentation technique ». In : *IEEE International Conference on Image Processing ICIP*. 2002.
- [ZHA 02b] Ming ZHAO, Jiajun BU et Chun CHEN. « Robust background subtraction in HSV color space ». In : *ITCom 2002 : The Convergence of Information Technologies and Communications*. International Society for Optics et Photonics. 2002, p. 325–332.
- [ZHA 04] Li ZHANG, Weida ZHOU, Licheng JIAO et al. « Wavelet support vector machine ». In : *IEEE Transactions on Systems, Man, and Cybernetics, Part B : Cybernetics* 34.1 (2004), p. 34–39.
- [ZHA 06] Wei ZHANG, Xiang Zhong FANG et Yi XU. « Detection of moving cast shadows using image orthogonal transform ». In : *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*. T. 1. IEEE. 2006, p. 626–629.
- [ZHA 07a] Jian ZHANG et Yueting ZHUANG. « Adaptive weight selection for incremental eigen-background modeling ». In : *Multimedia and Expo, 2007 IEEE International Conference on*. IEEE. 2007, p. 851–854.
- [ZHA 07b] Wei ZHANG et al. « Moving cast shadows detection using ratio edge ». In : *Multimedia, IEEE Transactions on* 9.6 (2007), p. 1202–1214.
- [ZHA 08] Youdong ZHAO et al. « Spatio-temporal patches for night background modeling by subspace learning ». In : *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on*. IEEE. 2008, p. 1–4.
- [ZHA 09] ZhaoHui ZHANG et al. « Moving Foreground Detection Based on Modified Codebook ». In : *Image and Signal Processing, 2009. CISP'09. 2nd International Congress on*. IEEE. 2009, p. 1–5.
- [ZHA 10] Lijing ZHANG et Yingli LIANG. « Motion human detection based on background subtraction ». In : *Education Technology and Computer Science (ETCS), 2010 Second International Workshop on*. T. 1. IEEE. 2010, p. 284–287.
- [ZHO 03] Jing ZHONG et Stan SCLAROFF. « Segmenting foreground objects from a dynamic textured background via a robust kalman filter ». In : *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*. IEEE. 2003, p. 44–50.

- [ZHU 09] Xiaojin ZHU et Andrew B GOLDBERG. « Introduction to semi-supervised learning ». In : *Synthesis lectures on artificial intelligence and machine learning* 3.1 (2009), p. 1–130.
- [ZHU 10] Zhou ZHU et Xiaobo LU. « An accurate shadow removal method for vehicle tracking ». In : *Artificial Intelligence and Computational Intelligence (AICI), 2010 International Conference on*. T. 2. IEEE. 2010, p. 59–62.
- [ZIV 06] Zoran ZIVKOVIC et Ferdinand van der HEIJDEN. « Efficient adaptive density estimation per image pixel for the task of background subtraction ». In : *Pattern recognition letters* 27.7 (2006), p. 773–780.

Publications de l'auteur

Publications Internationales

Brahim Farou, Seridi Hamid and Akdag Herman, "A New Approach for the Extraction of Moving Objects", Modeling Approaches and Algorithms for Advanced Computer Applications, Springer International Publishing, pp 27-36, 2013.

Brahim Farou, Hamid Seridi and Herman Akdag, "Improved Parameters Updating Algorithm for the Detection of Moving Objects", Computer Science and Its Applications, Springer International Publishing, pp 527-537, 2015.

Brahim Farou, Houssam Eddine Rouabhia, Hamid Seridi and Herman Akdag, "A novel approach for detection and removal of moving cast shadows based on RGB, HSV and YUV color spaces", Computing and Informatics Journal (CAI), paper accepted and to be published in subsequent issue of 2016.

Brahim Farou, Hamid Seridi and Herman Akdag, "Improved Gaussian Mixture Model With Background Spotter For The Extraction of Moving Objects", The International Arab Journal of Information Technology (IAJIT), to be published in, Volume 14, No. 2, March 2017.

Communications Internationales

Brahim Farou and Hamid Seridi, "Système de détection et de suivi des objets en mouvement dans une vidéo de surveillance", International Conference on Signal, Image, Vision and their Applications SIVA'11, 21-24 Novembre, Guelma, Algérie, 2011.

Brahim Farou and Hamid Seridi, "Approche basée sur les mixtures de gaussiennes et les agents guetteurs pour l'extraction des objets en mouvement", Colloque sur l'Optimisation et les Systèmes d'Information, COSI'2012, 12-15 mai, Telemcen, Algérie, 2012.

Brahim Farou and Hamid Seridi, "A system based on the analysis of the histograms and Gaussian Mixture Models for the extraction of moving objects", The First Conference on Theoretical and Applicative Aspects of Computer Sciences, CTAACS'2012, 25-26 Novembre, Skikda, Algérie, 2012.

Brahim Farou, Hamid Seridi and Herman Akdag, "A new approach for the extraction of moving objects", 4th International Conference on Computer Science and its Applications, CIIA'2013, 4-6 May 2013, Taher Moulay University, Saida, Algeria, 2013.

Brahim Farou, Hamid Seridi and Herman Akdag, "Using Gaussian Mixture Models and HSV color space for background subtraction", International conference on control engineering, CONTECH'13, 26-28 Décembre, Istanbul, Turkey, 2013.

Yakouta Zarouk, Ismahane Souici, **Brahim Farou** and Hamid Seridi, "Enhanced Evolutionary Algorithm for Partial Video Encryption", 2nd International Conference on Signal, Image, Vision and their Applications, SIVA'2013, 18-20 Novembre, Guelma, Algérie, 2013.

Brahim Farou, Houssam Rouabhia and Hamid Seridi, "Une nouvelle approche pour l'élimination des ombres portées engendrées par les objets mobiles dans une vidéo de surveillance", The 1st International Conference on Applied Automation and Industrial Diagnostics, ICAID'2015, 29-30 Mars, Djelfa, Algérie, 2015.

Brahim Farou, Houssam Rouabhia and Hamid Seridi, "Improved Parameters Updating Algorithm for the Detection of Moving Objects", 5th IFIP International Conference Computer Science and its Applications, CIIA'2015, 20-21 Mai, Saida, Algérie, 2015.

Hamid Seridi, Houssam Rouabhia and **Brahim Farou**, "Texture-based method for shadow detection", 5th International Conference on Information Systems and Technologies, ICIST'2015, 21-23 Mars, Istanbul, Turquie, 2015.

Communications Nationales

Brahim Farou and Hamid Seridi, "Modélisation multi gaussienne pour la détection et le suivi des objets en mouvements", première journée doctorale, JDI'2011, 12 Novembre, Guelma, Algérie, 2011.

Brahim Farou and Hamid Seridi, "Detection des objets en mouvement en temps réel", première journée Nationale des Sciences et technologies de l'information et de la Communication, JSTIC'2012, 01 Juillet, Guelma, Algérie, 2012.

Brahim Farou, Mohamed Nemissi and Hamid Seridi, "Approche multi gaussiennes utilisant une sélection automatique des zones d'intérêts pour la détection et le suivi des objets en mouvements", 2^{èmes} journée Doctorales en Informatique, JDI'2012, 18-19 Novembre, Guelma, Algérie, 2012.

Yakouta Zarouk, Ismahane Souici, **Brahim Farou** and Hamid Seridi, "Cryptage Partial pour un système de vidéo surveillance", deuxième journée nationale des sciences et technologies de l'information et de la communication, JSTIC'2013, 01 juillet, Guelma, Algérie, 2013.

Houssam Rouabhia, **Brahim Farou** and Hamid Seridi, "Real-Time People Dtection", 4^{èmes} Journées Doctorales en Informatique de Guelma, JDI'2014, 3-4 Décembre, Guelma, Algérie, 2014.

Houssam Rouabhia, **Brahim Farou** and Hamid Seridi, "People Dtection : State of The Art", 3^{èmes} journée des Sciences et Technologies de l'Information et de la Communication, JSTIC'2014, 22 Juin, Guelma, Algérie, 2014.