

UNIVERSITE BLIDA 1

Faculté de Technologie

Département d'Electronique

THESE DE DOCTORAT-LMD

Spécialité : Electronique

FUSION D'INFORMATIONS POUR LA VIDEOSURVEILLANCE INTELLIGENTE

Par

BENRAYA Imane

Devant le jury composé de :

A. GUESSOUM	Professeur	U. Blida 1	Président
N. BENBLIDIA	Professeur	U. Blida 1	Directrice de thèse
S. OUKID	Professeur	U. Blida1	Examinatrice
A. MEZIANE	MRA	CERIST	Examineur
F.Z. REGUIEG	MCB	U. Blida 1	Invitée

BLIDA, 08/01/2024

RESUME

Au cours de ces dernières décennies, la sécurité des personnes, des biens et de l'information s'est imposée comme un enjeu majeur à l'échelle planétaire. Confrontées à des problèmes tels que la lutte au terrorisme, le renforcement de la sécurité intérieure et l'essor de la cybercriminalité, nos sociétés investissent de plus en plus pour assurer leur protection. Ce secteur offre donc de grandes opportunités, tant sur le plan du développement technologique que des services. En particulier, les technologies de l'information et des communications apportent des solutions nouvelles et sophistiquées pour la sécurité physique et informatique.

Les progrès récents dans le traitement du signal permettent le développement de systèmes de vidéosurveillance intelligents, notamment des systèmes capables d'adapter de manière flexible le taux de collecte de données vidéo permettant de fournir des informations plus riches pour une analyse précise et crédible.

En raison de son importance dans la sécurité et la sûreté des personnes, la vidéosurveillance intelligente (VSI) remplace aujourd'hui le personnel de vidéosurveillance traditionnel permettant ainsi de réaliser un traitement et une analyse rapides et en temps réel de données volumineuses.

La VSI, appelée aussi détection automatique d'anormalités mise sur des logiciels qui exploitent les capacités de calcul embarquées dans les caméras ou dans des serveurs pour analyser en temps réel les images qu'elles enregistrent.

L'objectif de cette thèse est de proposer une approche permettant de détecter tout événement anormal dans la vidéo et d'émettre des alertes. L'approche proposée repose sur plusieurs étapes dont : 1) l'acquisition d'une base de données vidéos réels dans un atelier sous différentes conditions d'éclairage ; 2) la mise en œuvre de méthodes de soustraction en arrière-plan pour la détection d'objets en mouvement à partir de scènes vidéo ; 3) l'expérimentation et la validation de la démarche adoptée ; une étude et une analyse approfondie est effectuée afin de tester la robustesse de la proposition. En exploitant des données réelles. Nous avons évalué de manière exhaustive les performances de tous les algorithmes de la bibliothèque BGS, en examinant les performances de hachage, le temps d'exécution, la mémoire et la consommation du processeur.

ABSTRACT

Over the past few decades, the security of people, property and information has become a major issue on a global scale. Faced with issues such as the fight against terrorism, the strengthening of internal security and the rise of cybercrime, our societies are investing more and more to ensure their protection. This sector therefore offers great opportunities, both in terms of technological development and services. In particular, information and communication technologies provide new and sophisticated solutions for physical and computer security.

Recent advances in signal processing enable the development of intelligent video surveillance systems, including systems capable of flexibly adapting the video data collection rate to provide richer information for accurate and credible analysis.

Due to its importance in the safety and security of people, intelligent video surveillance (VSI) today replaces traditional video surveillance personnel thus allowing real-time and fast processing and analysis of big data.

The VSI, also called automatic detection of abnormalities, relies on software that exploits the computing capacities embedded in the cameras or in the servers to analyze the images they record in real time.

The objective of this thesis is to propose an approach allowing to detect any abnormal event in the video and to emit alerts. The proposed approach is based on several steps including: 1) the acquisition of a base of real video clips in a workshop under different lighting conditions; 2) the implementation of background subtraction methods for the detection of moving objects from video scenes; 3) experimentation and validation of the approach adopted; an in-depth study and analysis is carried out in order to test the robustness of the proposal using real data. We comprehensively evaluated the performance of all algorithms in the BGS library, looking at hashing performance, execution time, memory, and CPU consumption.

ملخص

على مدى العقود القليلة الماضية، أصبح أمن الأشخاص والممتلكات والمعلومات قضية رئيسية على نطاق عالمي. في مواجهة قضايا مثل مكافحة الإرهاب وتعزيز الأمن الداخلي وتصاعد جرائم الإنترنت، تستثمر مجتمعاتنا أكثر فأكثر لضمان حمايتها. لذلك يوفر هذا القطاع فرصاً كبيرة، سواء من حيث التطور التكنولوجي أو الخدمات. على وجه الخصوص، توفر تقنيات المعلومات والاتصالات حلولاً جديدة ومتطورة للأمن المادي وأمن الكمبيوتر.

تتيح التطورات الحديثة في معالجة الإشارات تطوير أنظمة ذكية للمراقبة بالفيديو، بما في ذلك الأنظمة القادرة على التكيف بمرونة مع معدل جمع بيانات الفيديو لتوفير معلومات أكثر ثراءً لتحليل دقيق وموثوق.

نظرًا لأهميتها في سلامة وأمن الأشخاص، تحل المراقبة الذكية بالفيديو (VSI) اليوم محل موظفي المراقبة بالفيديو التقليديين مما يسمح بمعالجة وتحليل البيانات الضخمة في الوقت الفعلي والسريع.

يعتمد VSI ، المعروف أيضًا باسم الاكتشاف التلقائي للتشوهات ، على البرامج التي تستغل قدرات الحوسبة المضمنة في الكاميرات أو في الخوادم لتحليل الصور التي تسجلها في الوقت الفعلي.

الهدف من هذه الأطروحة هو اقتراح نهج يسمح باكتشاف أي حدث غير طبيعي في الفيديو وإصدار تنبيهات. النهج المقترح يقوم على عدة خطوات منها: (1) الحصول على قاعدة من مقاطع الفيديو الحقيقية في ورشة عمل تحت ظروف الإضاءة المختلفة. (2) تطبيق طرق الطرح الخلفية للكشف عن الأجسام المتحركة من مشاهد الفيديو ؛ (3) التجريب والتحقق من صحة النهج المعتمد ؛ يتم إجراء دراسة وتحليل متعمقين من أجل اختبار قوة الاقتراح. باستخدام بيانات حقيقية. قمنا بتقييم أداء جميع الخوارزميات في مكتبة BGS بشكل شامل، بالنظر إلى أداء التجزئة ووقت التنفيذ والذاكرة واستهلاك وحدة المعالجة المركزية

REMERCIEMENTS

Je tiens à remercier Mme Nadjia Benblidia, Professeur à l'Université de Blida1, qui m'a encadré tout au long de cette thèse et qui m'a soutenu moralement. Mme Benblidia est mon modèle depuis toujours. J'espère qu'elle sera également remerciée pour sa gentillesse et sa présence constante et pour les nombreux encouragements qu'elle m'a prodigués. Avec elle, j'ai compris le sens de la rigueur et de la précision.

Je remercie très sincèrement les membres de mon jury d'avoir accepté de participer à mon évaluation. J'exprime ma reconnaissance à Madame Saliha Oukid, Professeur à l'université de Blida 1 ainsi qu'à Monsieur Abdelkrim Meziane, Directeur de Recherche au CERIST. Merci à Monsieur Abderrezak Guessoum, Professeur à l'université de Blida 1, de m'avoir fait l'honneur de présider mon jury de thèse. Je ne terminerai pas sans remercier Madame F.Z. Reguieg, Maître de Conférences à l'université de Blida 1 de m'avoir fait l'honneur de participer au jury.

Je tiens également à exprimer ma reconnaissance envers ma famille et mes collègues qui m'ont soutenu durant ce parcours.

LISTE DES TABLEAUX

Tableau 1 : Liste des algorithmes de la bibliothèque que nous avons utilisés	69
Tableau 1 : les parametres des algorithmes de soustraction d'arrière-plan	69
Tableau 2: Masques de premier plan obtenus à partir de la Vidéo n°1.....	72
Tableau 3 : Masques de premier plan obtenus à partir de la Vidéo n°2.....	73
Tableau 4 : Les masques fournis ont été obtenus à partir de la vidéo BMC	74
Tableau 5 : Résultats des métriques d'évaluation utilisées pour analyser les	76
Tableau 6 : Résultats des métriques d'évaluation utilisées pour analyser les	77
Tableau 7 : Résultats des métriques d'évaluation utilisées pour analyser les	78
Tableau 8 : Caractéristiques de chaque algorithme	80

LISTE DES FIGURES

Figure 1. Architecture générale d'un système automatique d'analyse des	12
Figure 2. Détection et suivi d'objets	18
Figure 3. Soustraction de l'arrière-plan et extraction de l'avant plan.	19
Figure 4. La différence temporelle : a) une scène avec deux objets en	20
Figure 5. Détection d'un objet en mouvement selon la méthode du Flux optique	20
Figure 6. Le suivi de trajectoires d'objets sur une séquence vidéo	21
Figure 7. Les représentations de la forme d'un objet. (a) centroïde.	21
Figure 8. Classification d'objets (humains vs véhicules)	22
Figure 9. Un aperçu des approches de reconnaissance d'action et d'activité	23
Figure 10. Flux d'informations dans un réseau multi-caméras	24
Figure 11. Les informations circulent entre différentes unités sur des réseaux	25
Figure 12. Taxonomie possible des méthodes de détection d'objets en	30
Figure 13. Algorithme de codebook qui a été créé par Kim et al	38
Figure 14. Taxonomie des techniques utilisées pour l'extraction des	43
Figure 15. Deux matrices de cooccurrence calculées à partir des régions	47
Figure 16. Taxonomie possible des techniques utilisées jusqu'à présent pour	49
Figure 17. Taxonomie possible des techniques proposées pour le suivi	51
Figure 18. Résultats de suivi d'objet mobile utilisant l'estimateur à noyau	54
Figure 19. Schéma graphique de la dépendance en suivi visuel sous l'hypothès.	55
Figure 20. Étapes cycliques pour l'estimation de la distribution de probabilité	56
Figure 21. Représentation graphique des 3 étapes d'estimation numérique	58
Figure 22. Exemple de recalage avec la DTW	61
Figure 23. Exemple de structure d'un HMM	61
Figure 24. Structure d'un réseau de neurones	62
Figure 25. Réseau de neurone non récurrent	64
Figure 26. Réseau de neurones récurrent	64
Figure 27 : Une collection de séquences d'images de la Vidéo n°1	66
Figure 28 : Une collection de séquences d'images de la Vidéo n°2	67
Figure 29: Vérité terrain de l'image n° 253 de la Vidéo n°1	68
Figure 30 : Vérité terrain de l'image n° 272 de la Vidéo n°2	68
Figure 31: Vérité terrain pour l'image n°597 de la scène vidéo BMC.	68
Figure 32 . Calcul des métriques de classification	75
Figure 33 : La mesure F et la précision des différentes méthodes pour la video1	79
Figure 34 : La mesure F et la précision des différentes méthodes pour la video2.	80
Figure 35 : Valeur moyenne du temps, de la mémoire et de la consommation	82
Figure 36: Détecter le mouvement des personnes dans la séquence de la vidéo	85
Figure 37: Détecter le mouvement des personnes dans la séquence	85
Figure 38 : Détecter le mouvement des personnes dans la séquence	85

Figure 39 : Détecter le mouvement des personnes dans la séquence de	86
Figure 40 : Séquence 1 de la vidéo 1.....	86
Figure 41 : Séquence 2 de la vidéo 1.....	86
Figure 42 : Séquence 1 de la vidéo 2.....	86
Figure 43 : Séquence 2 de la vidéo 2.....	87

LISTE DES ABREVIATIONS

AC :	Accuracy
ART :	Adaptive Resonance Theory
BGSLibrary :	Background Subtraction Library
BMC :	Background Modeling Challenge
CCTV :	Closed Circuit TeleVision
CNN :	Convolutional neural network
CPU :	Central Processing Unit
CSS :	Curvature Scale Space
DTW :	Dynamic time warping
FM :	F-measure
FN :	False negative
FP :	False positive
FNR:	False negative rate
<i>FPR:</i>	False positive rate
GLT :	Goal-line technology)
GMM :	Gaussian mixture model
HCI :	Hue-Chroma-Intensity
HMM :	Hidden Markov model
HSI	Hue (H), saturation (S), intensity (I). The Hue com
HSV :	Hue, Saturation, Value
KDE :	Kernel density estimation
MoG :	Mixture of Gaussians
PNN :	Les réseaux de neurones probabilistes
PTZ :	(Pan-Tilt-Zoom)
RAM :	Random Access Memory

RVB :	Rouge-vert-bleu
SOBS :	Self-Organizing Background Subtraction
TN :	True negative
TP :	True positive
<i>TPR</i> :	True positive rate
TSL :	Teinte, saturation, lumière
VAR :	Video Assistant Referees
VSAM :	Video Surveillance and Monitoring

TABLE DES MATIERES

RESUME.....	b
REMERCIEMENTS.....	e
LISTE DES TABLEAUX.....	f
LISTE DES FIGURES.....	g
TABLE DES MATIERES.....	3
INTRODUCTION.....	7
1 .Contexte	7
2. Problématique et Objectifs.....	8
3. Contribution	9
4. Organisation de la thèse :	9
GENERALITES SUR LA VIDEOSURVEILLANCE INTELLIGENTE	11
1.1 Introduction	11
1.2 Définition de la Surveillance Intelligente	11
1.3 Domaines d'Application	12
1.3.1 Lieux Gouvernementaux.....	12
1.3.2. Éducation.....	13
1.3.3. Commerce.....	13
1.3.4. Transport.....	14
1.3.5. Banques	15
1.4 Technologie de la Vidéosurveillance	15
1.4.1. Fixe.....	15
1.4.2. PTZ (Pan-Tilt-Zoom) Motorisée.....	15
1.4.3.Dôme.....	16
1.4.4. Mégapixel	16
1.4.5. Infrarouge et Thermique	16
1.4.6. Panoramique	16
1.5 Étapes de l'Analyse Vidéo dans la Surveillance Intelligente	17
1.5.1 Détection des Changements dans les Systèmes d'Analyse Vidéo	18
1.5.1.1. Méthode de Soustraction de l'Arrière-plan.....	19

1.5.1.2.Méthode de Différence Temporelle	19
1.5.1.3.Méthode du Flux Optique	20
1.5.2 Suivi des Objets.....	20
1.5.2.1. Représentation des Objets	21
1.5.3 Classification et Identification des Objets	22
1.5.4 Classification des Activités et Comportements	22
1.6 Vidéosurveillance Intelligente dans les Systèmes Multi-Caméras	23
1.6.1. Calibration	24
1.6.2. Topologie du Réseau (Network typologie).....	24
1.6.2.1. Reconnaissance(Recognition).....	24
1.6.2.2. Analyse d'activité (Activity analysis)	25
1.7. Conclusion	25
DETECTION D'OBJETS EN MOUVEMENT	26
2.1 Introduction	26
2.2 Les termes utilisés	26
2.2.1 Détection de Mouvement (Motion detection)	26
2.2.2 Estimation du Mouvement (Motion estimation).....	26
2.2.3 Modélisation de l'Arrière-plan	26
2.2.4 Soustraction de l'Arrière-plan (Background subtraction).....	27
2.2.5 Segmentation de Mouvement.....	27
2.3 Travaux de la littérature	27
2.4 Différentes Taxonomies	28
2.5 Méthodes de Détection de Mouvement	30
2.5.1 Détection avec Modélisation de l'Arrière-plan.....	30
2.5.1.1 Modélisation de l'Arrière-plan	30
2.5.1.2 Simple Gaussienne.....	31
2.5.1.3 Mélange de Gaussiennes (Mixture of Gaussians-MoG)	32
2.5.1.4 Modèles statistiques fondés sur des distributions gaussiennes :.....	33
2.5.1.5 Les modèles flous.....	37
2.5.1.6 CodeBook.....	37
2.5.1.7 Méthodes Neurales et Neuro-floues	38
2.5.1.8 Détection de l'Avant-plan.....	39
2.5.2. Détection sans Modélisation de l'Arrière-plan.....	39

2.5.2.1 Dérivée Temporelle	40
2.5.2.1. Flot Optique	40
2.6. Conclusion	41
SUIVI D'OBJETS EN MOUVEMENT	42
3.1 Introduction	42
3.2. Extraction des Caractéristiques et Représentation des Cibles	42
3.2.1. Caractéristiques de Bas Niveau	43
3.2.1.1. Couleur	43
3.2.1.2 Moments Statistiques	44
3.2.1.3. Texture et Matrice de Cooccurrence	45
3.2.1.3.1 Matrice de Cooccurrence.....	45
3.2.2. Caractéristiques de Niveau Intermédiaire.....	47
3.2.2.1. Les contours	47
3.2.2.2. Les points d'intérêts.....	48
3.2.3. Caractéristiques de Haut Niveau	48
3.2.3.1. Modélisation d'objet	48
3.2.4 Représentation des Cibles.....	48
3.2.4.1 Forme	49
3.2.4.2. Apparence	49
3.2.4.2.1 Histogramme	50
3.3. Suivi d'Objets Mobiles.....	50
3.3.1. Méthodes Déterministes d'Estimation des Positions	51
3.3.1.1. Kanade-Lucas-Tomasi.....	51
3.3.1.2. Estimateur à Noyau "MeanShift"	52
3.3.2. Méthodes Stochastiques	54
3.3.2.1. Filtre Bayésien Récursif.....	54
3.3.2.2. Filtre de Kalman	56
3.3.2.3. Filtre Particulaire.....	57
3.4. Conclusion	58
RECONNAISSANCE DE COMPORTEMENTS	59
4.1. Introduction	59
4.2. Reconnaissance Statique	59
4.2.1 Approche Région.....	59

4.2.2 Approche Contour	59
4.3. Reconnaissance Dynamique	60
4.3.1 Approche de Recalage Temporel	60
4.3.2 Modèles de Markov Cachés	61
4.3.2.1. Structure d'un HMM et Apprentissage	61
4.3.3 Méthodes connexionnistes	62
4.3.3.1 - Structure d'un réseau de neurones	62
4.3.3.2. Apprentissage.....	62
4 Règles d'Apprentissage par Compétition.....	63
4.3.3.3 Différents Types de Réseaux de Neurones	63
4.5 Conclusion	64
IMPLEMENTATION PRATIQUE ET RESULTATS EXPERIMENTAUX.....	65
5.1 Introduction	65
5.2 Base de Données	65
5.2.1 Vidéo du groupe d'ouvriers (vidéo 1)	66
5.2.2 Vidéo de Deux Employés de la Société (Vidéo 2)	66
5.2.3 BMC(Background Modeling Challenge)	67
5.2.4 La vérité terrain.....	67
5.2.5 Bibliothèque de Soustraction d'Arrière-plan (BGSLibrary).....	68
5.3 Évaluation en Temps Réel des Performances des Algorithmes de la BGS Library	70
5.3.1 Évaluation des Performances de la Segmentation	71
5.3.2 Évaluation des Performances des Algorithmes en Fonction de Leurs Caractéristiques.....	80
5.4 Détection et Suivi Automatiques d'Objets en Mouvement	82
5.4.1 Algorithme de Détection	83
5.4.2 Algorithme de Suivi.....	83
5.4.3 Résultats.....	85
5.5 Conclusion	87
CONCLUSION	89
LISTE DES ABREVIATIONS	1
REFERENCES	91

INTRODUCTION

1. Contexte

Aujourd'hui, nos sociétés sont exposées à un ensemble de menaces et de vulnérabilités sans précédent. Cette recrudescence est due aux avancées considérables enregistrées par les technologies de l'information et des communications mais aussi au rythme auquel les économies modernes se sont liées.

Avec la propagation de la criminalité, du terrorisme, des vols et des délits, la sécurité des personnes, des biens et des lieux est devenue un enjeu majeur, posé à l'échelle mondiale. Cet intérêt pour le secteur de la sécurité s'est traduit par un développement technologique et des services dans le domaine des technologies de l'information et de la communication. La sécurité est représentée sous deux aspects, la sécurité physique et la sécurité logique (sécurité informatique).

La sécurité physique, est l'ensemble des mesures physiques prises par les individus pour protéger les personnes, les lieux et les biens. Quant à la sécurité logique ou sécurité informatique, elle vise à mettre en œuvre un ensemble de mesures de sécurité pour protéger les technologies de l'information telles que les programmes, les réseaux et les données contre la cybercriminalité.

La vidéosurveillance à distance de lieux publics et privés est connue par l'utilisation de caméras qui captent des flux d'images et les transfèrent à des dispositifs de contrôle pour les stocker, les archiver et les reproduire sur des écrans afin de prévenir les fraudes, vols et agressions et pour la gestion des accidents, La surveillance a été utilisée dans les années cinquante du siècle dernier ; elle était connue sous le nom de technologie de télévision en circuit fermé (CCTV) et l'émergence de l'enregistrement vidéo numérique a commencé dans les années soixante-dix, ce qui a révolutionné le domaine de la vidéosurveillance [1].

Le nombre de caméras installées dans les lieux publics, les places et autres lieux importants a connu une augmentation significative, atteignant des millions dans un pays, en particulier le Royaume-Uni et les États-Unis d'Amérique [1]. Dans un système de surveillance traditionnel, un agent de sécurité est affecté à l'opérateur pour regarder les vidéos capturées par les caméras. Cette opération tient compte d'une énorme quantité de vidéos enregistrés et du nombre d'heures d'enregistrement avec la présence de milliers de caméras. Ceci rendait la

surveillance plus difficile pour les personnes car la concentration d'une personne reste élevée en regardant plusieurs vidéos pendant de longues heures.

L'étude a démontré que la durée d'attention maximale pour toute personne surveillant une tâche ne dépasse pas 20 minutes. De plus, la surveillance de clips vidéo par des personnes est sujette à des erreurs ce qui engendre de nombreux événements importants manqués [2]. La solution fondamentale à ces problèmes réside dans les systèmes d'analyse vidéo intelligents, afin de fournir une alerte en temps réel dans les situations d'intérêt et une intervention immédiate avant qu'elle ne se produise. Après les événements du 11 décembre 2001, le secteur de la sécurité a été témoin d'un saut qualitatif dans la lutte contre les activités terroristes. Les études et les efforts se sont concentrés sur le développement de programmes intelligents pour analyser la vidéo, détecter et suivre les objets et signaler les événements suspects et prévenir les accidents avant qu'ils ne surviennent [1]. Le système d'analyse vidéo est apparu dans plusieurs secteurs tels que le sport, le commerce, le transport, la sécurité, le divertissement et le trafic. Par exemple, dans le domaine du sport lors de la Coupe du monde 2018, la surveillance des matchs à l'aide de systèmes intelligents a été utilisée pour améliorer la capacité de surveillance de l'arbitrage mais aussi pour détecter des incidents pouvant se produire dans les tribunes [1].

Le système appelé assistance vidéo (VAR) est une technologie permettant de surveiller les joueurs et d'analyser leurs performances en temps réel. De plus, un nouveau système de surveillance a été utilisé dans FIFA 2018, c'est la technologie GLT (Goal-Line Technology) qui permet de surveiller la ligne de but. Les systèmes de surveillance intelligents fonctionnent à l'aide d'algorithmes pour analyser des clips vidéo qui sont transformés en un ensemble d'images fixes. Une caméra vidéo produit environ 25 à 30 images par seconde parfois plus. Avec l'utilisation de caméras vidéo 4D et 3D, le nombre peut correspondre à des milliers ou des millions d'images dans un échantillon. Les systèmes de surveillance intelligents les plus importants sont les systèmes VSAM (Virtual Storage Access Method). VSAM est un système de vidéosurveillance automatisé qui estime l'image d'arrière-plan et la met à jour pour refléter les changements dans l'environnement d'arrière-plan [1] [2].

2. Problématique et Objectifs

Les caméras se trouvent de plus en plus dans des environnements où l'on s'intéresse à savoir à ce qui va se passer. Elles sont installées dans les lieux publics (routes, stations de métro, hôpitaux, supermarchés), ainsi que dans les lieux avec accès restreint (campus universitaires, bâtiment des entreprises, banques) ou dans les lieux personnels (à la maison). L'objectif de la vidéosurveillance est à la fois de détecter des événements prédéfinis (par ex : une voiture roule en sens interdit) ou d'analyser à long terme des comportements humains (par ex : l'analyse de la visite des clients dans les rayons dans un supermarché) dans les vidéos enregistrées. De nombreuses caméras et une

capacité énorme de stockage permettent d'avoir de grandes bases de vidéos de surveillance. La valeur de ces bases de données dépend de la capacité de retrouver des informations voulues dans ces bases. De nombreuses approches proposées cherchent à répondre au besoin d'avoir des outils de recherche. La fusion d'informations est apparue afin de gérer des quantités très importantes de données multi-sources. Elles consistent à combiner des informations issues de plusieurs sources afin d'aider à la prise de décision. Les méthodes de fusion ont été adaptées et développées pour des applications en traitement du signal et plus particulièrement pour la classification. L'objectif principal de cette thèse est d'étudier et d'analyser les nouvelles technologies de l'information pour la vidéosurveillance intelligente.

3. Contribution

La surveillance intelligente nécessite la détection d'objets en mouvement dans la séquence d'images de la vidéo. Ces dernières années, plusieurs techniques ont été proposées pour identifier des objets en mouvement. Cependant, la littérature manque d'une évaluation complète de ces méthodes, ce qui rend difficile la sélection des méthodes les plus performantes. Nos contributions sont listées comme suite :

- ✓ l'acquisition d'une base de clips vidéo réels dans un atelier sous différentes conditions d'éclairage ;
- ✓ l'étude des principales méthodes qui détectent les objets en mouvement dans les scènes vidéo ;
- ✓ la mise en œuvre de méthodes de soustraction en arrière-plan pour la détection d'objets en mouvement à partir de scènes vidéo ;
- ✓ l'évaluation de manière exhaustive des performances de 43 algorithmes de la bibliothèque BGS, en examinant les performances de hachage, le temps d'exécution, la mémoire et la consommation du processeur ;
- ✓ l'extraction des caractéristiques et facteurs affectant l'efficacité de chaque technique afin d'offrir aux chercheurs des moyens efficaces et simples pouvant être utilisés dans le domaine de la surveillance intelligente.

4. Organisation de la thèse :

La thèse se compose de cinq chapitres

Chapitre Un : Il présente un aperçu de la surveillance intelligente, de ses utilisations et des technologies utilisées.

Chapitre Deux : il décrit les techniques utilisées pour détecter les objets en mouvement dans les clips vidéo.

Chapitre Trois : il présente les méthodes d'extraction de caractéristiques nécessaires pour l'estimation des emplacements cibles ainsi que celles utilisées pour le suivi visuel.

Chapitre Quatre : il expose les méthodes de reconnaissance de comportements.

Chapitre cinq : il illustre le protocole expérimental exploitant les méthodes utilisées dans la surveillance intelligente ainsi que les résultats obtenus.

CHAPITRE 1

GENERALITES SUR LA VIDEOSURVEILLANCE INTELLIGENTE

1.1 Introduction

La vidéosurveillance intelligente constitue un système de surveillance automatisé développé pour détecter de nombreuses situations dangereuses dans différentes localités afin d'assurer la sécurité dans les espaces privés et publics. La surveillance intelligente est devenue un sujet de recherche d'une importance capitale, particulièrement en raison des avancées technologiques constatées dans divers domaines que nous aborderons dans ce chapitre. Fondamentalement, du point de vue scientifique, la surveillance intelligente repose sur l'analyse vidéo au moyen d'un ensemble d'algorithmes permettant d'examiner en temps réel le flux d'images généré par les caméras vidéo au sein de réseaux de surveillance. Nombreuses sont les études qui ont émergé en faveur de la surveillance intelligente.

1.2 Définition de la Surveillance Intelligente

La surveillance intelligente englobe une technologie sophistiquée conçue pour observer et identifier les comportements de personnes, d'objets et de situations potentiellement suspectes au sein de séquences vidéo. Ce processus est orchestré par des programmes informatiques et des algorithmes qui sont traités par un ordinateur. L'essence de cette technologie réside dans la capacité de convertir les éléments vidéo en données, ces dernières étant ultérieurement transmises et archivées en vue d'une analyse minutieuse et d'une classification pertinente en tenant compte des données disponibles (cf. figure 1). À titre d'exemple, elle peut engendrer une alerte destinée au personnel de surveillance, ce qui leur permet d'intervenir promptement.

Les systèmes de surveillance intelligente sont conçus en s'appuyant sur des modèles mathématiques visant à détecter les objets en mouvement, à différencier les mouvements naturels des anomalies, et à suivre les variations temporelles et spatiales. Ce processus aboutit à la création d'une base de données consignnant les caractéristiques de mouvement détectées, ainsi que leurs spécificités.

L'élaboration de la surveillance intelligente repose sur l'utilisation de technologies de pointe provenant de divers domaines, dont la vision par ordinateur, le traitement du signal, l'intelligence artificielle et l'exploration de données [2].

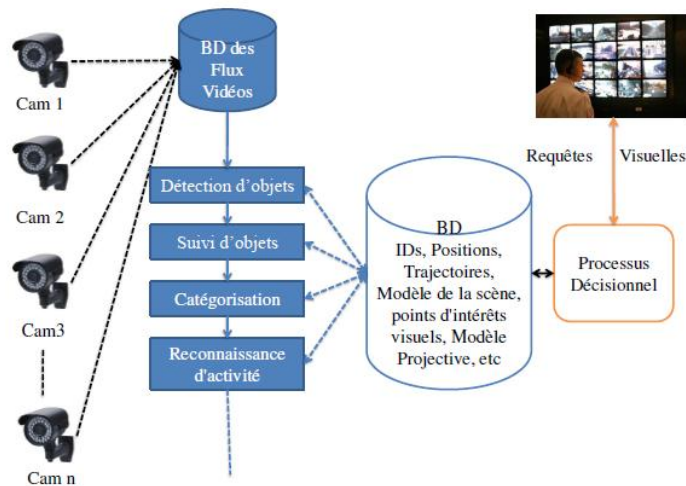


Figure 1. Architecture générale d'un système automatique d'analyse des vidéos distribuées en vidéosurveillance [6]

1.3 Domaines d'Application

La vidéosurveillance est largement déployée dans divers contextes urbains tels que les villes, les quartiers, les réseaux de transport, les campus, les festivals, les compétitions sportives, ainsi que lors de sommets économiques et autres événements majeurs. En vue d'assurer une surveillance optimale et de garantir les avantages escomptés, il est courant de déployer plusieurs centaines de caméras lors de tels événements, ces dispositifs étant gérés par un nombre substantiel d'intervenants opérant au sein des services de sécurité et d'urgence.

Il convient de noter que la vidéosurveillance est souvent intégrée à une panoplie de systèmes de sécurité et de contrôle, incluant notamment la maîtrise des incendies, les systèmes de téléphonie, les communications sans fil, et autres infrastructures connexes. Étant donné la primauté d'une réponse d'urgence prompt ainsi que le volume considérable de caméras déployées, les applications d'analyse de vidéos intelligentes se révèlent idéales pour déclencher des alertes en cas de situations suspectes ou d'événements potentiellement préjudiciables

1.3.1 Lieux Gouvernementaux

La vidéosurveillance revêt un rôle essentiel au sein des sphères sensibles du gouvernement, visant à consolider la protection contre les actes terroristes, les agressions extérieures et les activités d'espionnage à l'échelle internationale. En effet, divers objectifs peuvent être atteints grâce à la mise en place de systèmes de vidéosurveillance dans les espaces publics, tels que [2].

- les points frontaliers,
- les édifices gouvernementaux et les sites institutionnels,
- les laboratoires de recherche,
- les installations militaires,
- les rassemblements politiques.

Les impératifs de la surveillance intelligente sont nombreux au sein du secteur gouvernemental, englobant :

- l'identification des individus et des véhicules,
- le comptage des flux de personnes,
- la surveillance des rassemblements de foules,
- la détection de comportements potentiellement suspects ou agressifs,
- le repérage des intrusions et des situations dangereuses,
- la surveillance des voies de circulation.

1.3.2. Éducation

L'intégration de la vidéosurveillance au sein des établissements éducatifs, tels que les écoles et les universités, revêt une importance cruciale en vue d'assurer la sécurité des enseignants, des professeurs et des étudiants, ainsi que de préserver les biens, les équipements et les infrastructures des risques inhérents au vandalisme et au vol.

Par ailleurs, le contexte actuel est marqué par une recrudescence des actes criminels au sein des campus scolaires et universitaires, mettant en lumière des cas de meurtres et de conflits débouchant sur des tragédies dans plusieurs pays. Dans ce contexte, une surveillance adéquate et ciblée au sein des établissements éducatifs s'avère primordiale. Elle vise à identifier et à surveiller les incidents de vandalisme, les vols, les comportements répréhensibles, les conduites inappropriées, la gestion des entrées, ainsi que l'accès de personnes étrangères ou de véhicules non identifiables [2].

Par conséquent, il est impératif de mettre en place des systèmes de vidéosurveillance intelligents au sein du secteur éducatif. Ces dispositifs jouent un rôle essentiel dans la prévention et la réduction des risques liés aux actes criminels et aux comportements indésirables, contribuant ainsi à instaurer un environnement sécurisé et propice à l'apprentissage au sein des écoles et des universités.

1.3.3. Commerce

La vidéosurveillance constitue une composante essentielle du secteur commercial, visant à garantir à la fois la sécurité interne – notamment la protection des magasins et entrepôts – et la sécurité externe – englobant les parkings, les camions et le contrôle des marchandises. Cette utilisation trouve son origine dans les risques substantiels de vols et d'attaques auxquels les entreprises commerciales sont exposées, engendrant des pertes significatives. Les applications de la vidéosurveillance s'étendent tant aux grandes entreprises qu'aux petites structures, incluant les magasins et les marchés. Les objectifs inhérents à ces initiatives se déclinent comme suit [2].

- la surveillance des transactions en espèces en vue de prévenir le vol et la fraude ;

- la préservation des actifs physiques et des infrastructures ;
- la gestion des biens et des marchandises, notamment le suivi des stocks et des livraisons ;
- la sécurité des employés et des clients ;
- la préparation à des scénarios d'urgence, tels que les incendies ou les alarmes ;
- le contrôle des accès, y compris la surveillance des entrées, des sorties, des parkings et des véhicules.

Par ailleurs, le secteur du commerce exprime clairement des besoins prépondérants en matière d'analyse vidéo et de surveillance intelligente, en vue de sécuriser le personnel et les biens, de gérer les opérations et de favoriser les initiatives de marketing. À titre d'illustration, la société canadienne ORGANIX IT basée au Québec a mis en œuvre un système de surveillance intelligent au sein d'une chaîne de magasins [2].

L'adoption de ces solutions technologiques témoigne de l'importance accordée à l'optimisation de la sécurité, à la gestion efficace des risques et à la préservation des actifs au sein du secteur commercial, en alignement avec les impératifs de protection des intérêts de l'entreprise et de ses parties prenantes.

1.3.4. Transport

En réponse à l'essor des menaces et des risques au sein des infrastructures de transport, tels que les aéroports, les gares, les ports et les réseaux de transports publics à l'échelle mondiale, la mise en œuvre d'une surveillance et d'une analyse vidéo adéquate et intelligente représente un défi majeur visant à garantir la sécurité et la sûreté des individus ainsi que des biens matériels. Après les événements du 11 septembre 2001, une intensification des mesures de sécurité a été observée dans ces contextes, étant donné que ce secteur entretient une relation directe avec l'économie nationale. En raison du vaste nombre de voyageurs transitant par ces diverses installations de transport, et compte tenu des rassemblements d'envergure s'y tenant et de l'ampleur de leurs infrastructures, tout incident lié à la sécurité ou au terrorisme pourrait avoir des répercussions majeures, engendrant des pertes substantielles tant en termes de vies humaines que de biens matériels.

La mise en œuvre de systèmes de surveillance intelligente et d'analyse vidéo au sein du secteur des transports vise à accomplir plusieurs objectifs primordiaux, notamment surveiller les flux d'entrées et de sorties, repérer les individus suspects, superviser les bagages et identifier les bagages abandonnés en l'absence de passagers, reconnaître les visages, comptabiliser les voyageurs, relever les numéros d'immatriculation des véhicules entrant et sortant des installations de transport, tout en identifiant des comportements [2].

Les défis auxquels font face les techniques d'analyse vidéo dans ce contexte incluent la variété des comportements des voyageurs, la congestion des espaces et les variations climatiques (telles que le brouillard, la neige, la pluie, ainsi que les reflets d'eau à la surface de la mer et les mouvements des vagues).

Dans ce domaine, on observe également le recours à des méthodes d'identification biométrique, comme les empreintes digitales et l'iris, en plus de l'utilisation de cartes à puce et de numéros d'identification personnels, contribuant ainsi à renforcer l'arsenal de mesures déployées en faveur de la sécurité et de la gestion des risques.

1.3.5. Banques

La présence de dispositifs de vidéosurveillance au sein des établissements bancaires revêt une importance significative en ce qu'elle concourt à garantir la sécurité desdites institutions, tout en exerçant un effet dissuasif à l'encontre des vols qualifiés et des actes de vol à main armée. Cette efficacité découle de l'utilisation de séquences vidéo soigneusement archivées, permettant d'engager des enquêtes ultérieures en vue d'identifier et de poursuivre les individus impliqués dans des actes criminels.

Par ailleurs, l'intégration de la surveillance intelligente agit en faveur d'une intensification de l'efficacité des opérations de surveillance. L'analyse vidéo s'avère particulièrement apte à identifier des schémas de comportements récurrents, notamment au travers de multiples réseaux de caméras, à travers l'application de technologies de reconnaissance faciale.

1.4 Technologie de la Vidéosurveillance

La vidéosurveillance repose essentiellement sur le rôle prépondérant des dispositifs de caméras qui constituent le premier élément de l'infrastructure de vidéosurveillance intelligente. La diversité des besoins de surveillance induit une variation dans les caractéristiques de ces caméras. Une panoplie de modèles de caméras est disponible, englobant des options fixes, mobiles, numériques et analogiques. À cet égard, nous nous attardons sur quelques exemples significatifs:

1.4.1. Fixe

La surveillance au moyen d'une caméra fixe se caractérise par sa capacité à focaliser son champ visuel dans une direction précise, englobant ainsi une zone déterminée [2]. Cette approche, qualifiée de traditionnelle, jouit d'une utilisation répandue et constitue un choix de prédilection. Elle s'avère particulièrement adéquate lorsque l'objectif est de maintenir une visibilité permanente sur une direction spécifique.

1.4.2. PTZ (Pan-Tilt-Zoom) Motorisée

Cette variante de caméra, quant à elle, offre la possibilité d'ajuster sa position soit de manière automatisée, soit manuellement. Cette adaptation permet

de suivre les déplacements de personnes ou d'objets au sein de la scène surveillée, ainsi que d'élargir la portée des zones d'intérêt [2]. Un exemple concret de son utilisation se retrouve dans le traçage des plaques d'immatriculation des véhicules.

1.4.3. Dôme

Le dôme de surveillance adopte une configuration hémisphérique, pouvant être soit fixe soit mobile. Dans le cas des modèles mobiles, ces derniers ont la capacité de se déplacer et d'effectuer des rotations horizontales sur un angle de 360 degrés, ainsi que des rotations verticales sur un degré de 180 degrés. Cette conception permet au dispositif de couvrir une vaste étendue en termes de surveillance. La caractéristique de sa forme hémisphérique offre une certaine discrétion tout en renforçant sa résistance face aux actes de vandalisme et aux variations météorologiques [2].

1.4.4. Mégapixel

Il présente une résolution élevée, variant de 1 à 16 mégapixels, offrant ainsi des images plus minutieuses qui peuvent être agrandies à l'aide de logiciels informatiques. En outre, sa large plage de vision permet de réduire le nombre de caméras requises pour la surveillance d'une zone donnée, couvrant des angles allant de 140 à 360 degrés. Cette haute précision confère aux algorithmes la capacité d'améliorer la performance en matière de détection et de reconnaissance, notamment dans les domaines tels que la lecture des plaques d'immatriculation et la reconnaissance faciale [2].

1.4.5. Infrarouge et Thermique

Ce modèle de caméra est spécifiquement conçu pour des applications de surveillance nocturne. En effet, il se caractérise par sa capacité à capturer des images de haute qualité même dans des conditions d'obscurité totale. Lors des périodes nocturnes, cette caméra est en mesure de produire des images monochromes en noir et blanc. Cependant, pendant la journée, elle est capable de capturer des images en couleurs. Certains modèles de ce type de caméra sont équipés d'une source de lumière infrarouge intégrée, laquelle entre en action lorsque le niveau de luminosité décroît en deçà d'un seuil prédéterminé durant la nuit.

Par ailleurs, il convient de mentionner, les caméras thermiques enregistrent le rayonnement thermique émis par les objets, ne dépendant pas d'une source lumineuse externe [2].

1.4.6. Panoramique

La caméra de vidéosurveillance offre un champ de vision unidirectionnel à 360 degrés, ce qui permet de réajuster l'image par défaut à l'aide d'un logiciel dédié. Cependant, sa précision ne répond pas aux exigences des analyses nécessitant un niveau de détail élevé, ce qui est essentiel pour assurer une surveillance de qualité des visages et des iris [2].

1.5 Étapes de l'Analyse Vidéo dans la Surveillance Intelligente

La vidéosurveillance intelligente et l'analyse vidéo ont connu une forte activité ces dernières années. De nombreuses études se sont concentrées sur l'amélioration et la mise en œuvre d'algorithmes efficaces. Cette dynamique a conduit à l'émergence de produits commerciaux qui ont enrichi ce secteur et ont favorisé le développement des techniques de détection et de suivi d'éléments au sein de séquences vidéo.

Le fonctionnement de la vidéosurveillance intelligente repose principalement sur des programmes et des algorithmes mathématiques. Ces derniers sont analysés par des ordinateurs. Le travail d'analyse s'établit généralement sur deux niveaux : la détection du mouvement et la reconnaissance des séquences. Les programmes d'analyse ont pour tâche de traiter les séquences d'images générées par les flux vidéo. Leur objectif premier est de détecter automatiquement les objets et de les différencier en fonction de leur nature. De plus, ces programmes ont la capacité d'extraire les événements jugés pertinents. Une fois cette étape de détection effectuée, ils peuvent également procéder à l'identification et au suivi des objets, comme cela est illustré dans la figure 1. Il est essentiel que ces processus se déroulent en temps réel afin de permettre la génération d'alertes préventives avant la survenue d'accidents [2].

L'analyse vidéo dans le contexte de la surveillance intelligente implique plusieurs étapes distinctes, chacune ayant un rôle fondamental et une fonction spécifique au sein de ce processus. En effet, la surveillance intelligente peut être segmentée en cinq étapes essentielles [3] :

- La détection de changement.
- La segmentation des objets en mouvement.
- Le suivi des objets.
- La classification et l'identification des objets.
- La classification des tâches de niveau supérieur pour les activités et les comportements.

Chacune de ces étapes est caractérisée par l'application de méthodes spécifiques. Ces méthodes permettent de décrypter les comportements résultant des interactions entre les personnes ou les objets en mouvement. A chaque étape sont associés plusieurs algorithmes qui servent de base à la recherche en surveillance intelligente. Étant donné que ces méthodes peuvent être adaptées à diverses applications, il est impératif de les analyser minutieusement et de les intégrer à chaque étape en fonction des besoins spécifiques.

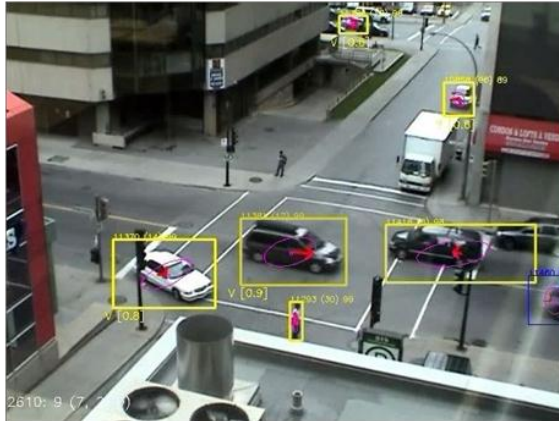


Figure 2. Détection et suivi d'objets [1]

1.5.1 Détection des Changements dans les Systèmes d'Analyse Vidéo

La détection des changements revêt une importance cruciale au sein des systèmes d'analyse vidéo, constituant ainsi une étape fondamentale dans ce domaine. Elle consiste en l'extraction des mouvements présents au sein d'une séquence vidéo ainsi que la détection de toute altération dans la scène observée. Cette étape automatisée permet de repérer toute activité au sein de la zone sous surveillance et peut également contribuer à identifier l'apparition ou la disparition d'éléments dans cette même zone [2]. En effet, cette phase préliminaire joue un rôle clé dans les systèmes de vision informatique et d'analyse vidéo, en supportant l'identification de cibles, l'extraction de caractéristiques, la configuration automatique des algorithmes de suivi d'objets, et la présélection des éléments au sein des réseaux multi caméras.

Le processus de détection des changements implique la classification des pixels contenus dans une image extraite de la séquence vidéo. Les objets en mouvement sont attribués à la catégorie des éléments en premier plan, tandis que les parties stables de la scène sont regroupées en tant qu'éléments d'arrière-plan. Cette segmentation pixel par pixel permet de distinguer nettement les éléments d'avant-plan de ceux d'arrière-plan, ce qui facilite la mise à jour de la représentation de l'arrière-plan tout en extrayant les éléments saillants d'avant-plan. Cependant, il convient de souligner que la modélisation de l'arrière-plan est confrontée à d'importants défis du fait des variations climatiques et des influences extérieures agissant sur la scène observée. Parmi ces facteurs figurent les ombres, les mouvements du soleil, les nuages, les reflets sur les surfaces d'eau et de verre, ainsi que les variations soudaines de luminosité.

Diverses études ont émergé dans le but de concevoir des modèles de fond capables de s'adapter aux changements environnementaux, qu'ils soient naturels ou brusques. Une gamme de méthodes différentes a été proposée, allant de l'estimation de mouvement et des dérivées temporelles à la création de modèles pour l'arrière-plan. Parmi ces méthodes, celle qui prédomine dans le domaine de l'analyse vidéo est la détection basée sur la modélisation de l'arrière-plan. Dans

cette approche, l'image courante de la séquence vidéo est soustraite de l'image estimée de l'arrière-plan, ce qui permet de le modéliser. De nombreuses études ont été dédiées à la recherche d'une modélisation appropriée de l'arrière-plan, ce qui a conduit à la proposition de différentes méthodes de modélisation [4].

1.5.1.1. Méthode de Soustraction de l'Arrière-plan

La détection des changements au sein d'une séquence vidéo peut être conceptualisée comme la quête de différences de couleur ou de texture entre les images successives. Cette démarche implique une comparaison pixel par pixel au sein de la même séquence. La technique de soustraction de l'arrière-plan repose sur une approche intelligente qui se concentre sur la confrontation de chaque image au sein d'une séquence vidéo avec une image de référence, reconnue comme représentant l'arrière-plan.

Le fond, ou arrière-plan, est établi à partir d'une image de référence, dépeignant la scène dépourvue de tout objet en mouvement. Les régions en décalage, dénommées ainsi en raison de la différence d'intensité entre les pixels, se forment lorsque cette différence dépasse un seuil défini T . Les pixels sont alors catégorisés comme appartenant au premier plan, tandis que l'ensemble des pixels présentant une différence d'intensité inférieure au seuil T constituent les éléments constitutifs de l'arrière-plan [5].



Figure 3. Soustraction de l'arrière-plan et extraction de l'avant plan

1.5.1.2. Méthode de Différence Temporelle

La méthode de différence temporelle émerge comme l'une des approches les plus prévalentes pour extraire les variations entre deux ou trois images successives d'une séquence vidéo. Cette technique fonctionne en comparant, pixel par pixel, le contraste temporel de l'intensité lumineuse de chaque pixel. Néanmoins, cette méthode n'a pas pour vocation de capturer l'ensemble des pixels en mouvement au sein d'une scène. Elle a tendance à négliger certaines des variations liées aux mouvements ralentis. De ce fait, cette limitation engendre une diminution de l'efficacité de cette méthode [2], [3], [5]. Un exemple de détection inexacte d'objet en mouvement est montré dans la figure 4.



Figure 4. La différence temporelle : a) une scène avec deux objets en mouvement b) La différenciation échoue à détecter tous les pixels de l'objet à gauche (couleur uniforme). Les régions en mouvement sont marquées en rouge [6]

1.5.1.3.Méthode du Flux Optique

Le flux optique est, par définition, le motif du mouvement apparent des objets, des surfaces et des contours dans une scène visuelle, causé par le mouvement relatif entre un observateur (œil ou caméra) et la scène. Le flux optique a pour but d'estimer le vecteur de vitesse de chaque pixel (flux optique dense) ou bloc de pixels, en utilisant l'information spatio-temporelle fournie par une séquence vidéo.

Ceci est accompli en évaluant les vitesses apparentes de déplacement des motifs de luminosité des pixels au sein de l'image. Cependant, la complexité des calculs associés à cette méthode limite son efficacité pour le traitement en temps réel [2], [7]. Un exemple de ce type de méthode est donné dans la figure suivante.

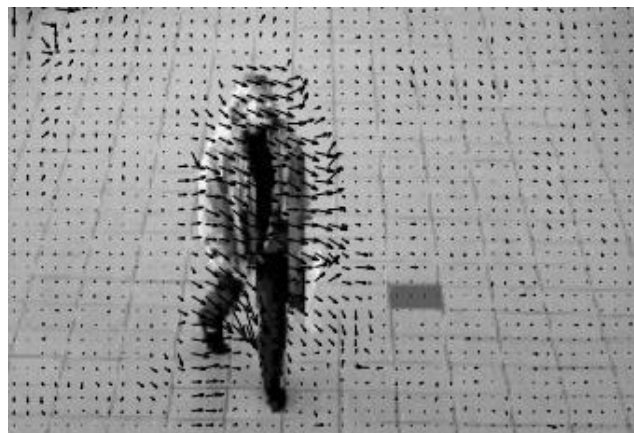


Figure 5. Détection d'un objet en mouvement selon la méthode du Flux optique [8]

1.5.2 Suivi des Objets

Le suivi d'objets constitue une étape cruciale dans le processus des systèmes de surveillance intelligents, visant à restreindre le suivi des individus ou des objets en mouvement. Son objectif est de suivre la trajectoire tracée par l'objet en mouvement, tout en déterminant sa position et son horaire au sein de la scène.

Le processus de détection et de suivi d'objet s'effectue en reliant les objets détectés dans l'image actuelle avec ceux détectés dans l'image précédente au fil

de la séquence d'images. Diverses techniques de suivi sont disponibles ; elles diffèrent en fonction du modèle d'apparence choisi pour les éléments mobiles. Par ailleurs, la représentation de l'objet varie en fonction du domaine d'application spécifique [2], [3], [5].

Le suivi d'objets présente divers défis, parmi lesquels on peut citer [4]:

- les similitudes entre les objets en mouvement : tels que la couleur, la forme et la vitesse ;
- les altérations de la forme de l'objet en raison de changements dans son mode de capture, comme la rotation ;
- les variations de l'éclairage ambiant ;
- les perturbations sonores.

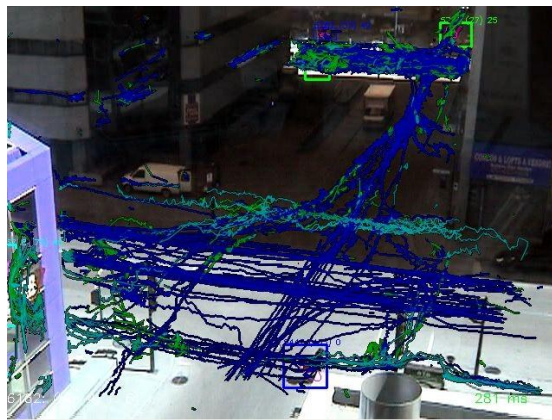


Figure 6. Le suivi de trajectoires d'objets sur une séquence vidéo [2]

1.5.2.1. Représentation des Objets

De nombreuses méthodes de suivi ont été proposées dans la littérature. Elles peuvent être divisées, en trois principales catégories qui sont le suivi par point d'intérêt, le suivi de noyaux et le suivi de silhouette, comme on peut le constater sur la Figure 6 [9].

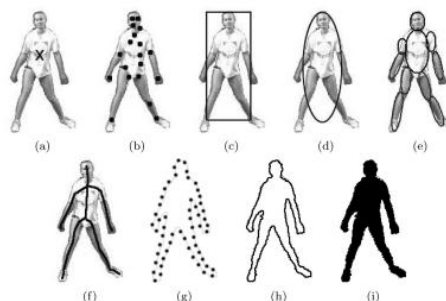


Figure 7. Les représentations de la forme d'un objet. (a) centroïde, (b) plusieurs points, (c) fenêtre englobant rectangulaire, (d) fenêtre englobant elliptique, (e) modèle a base de blobs, (f) squelette de l'objet, (g) les points de contrôle sur les contours de l'objet, (h) contour de l'objet complet, (i) silhouette de l'objet [9]

1.5.3 Classification et Identification des Objets

L'objectif de la caméra capte une scène comprenant diverses entités en mouvement telles que les piétons, les véhicules, les humains, les animaux, etc. Il est essentiel de déterminer le type d'objets enregistrés pour assurer un suivi et une analyse précis. Deux méthodes principales sont utilisées pour caractériser les objets en mouvement, à savoir les approches basées sur le mouvement et celles basées sur la segmentation [3], [10].

À titre d'exemple, comme illustré dans la figure 8, une personne prend une forme rectangulaire où la longueur excède la largeur, tandis qu'un véhicule adopte une forme rectangulaire avec une largeur supérieure à la longueur. Cette distinction découle des informations fournies par les algorithmes mathématiques aux systèmes relevant des méthodes de classification des objets selon leur forme. En ce qui concerne les approches liées au mouvement, l'activité humaine possède des caractéristiques spécifiques.

Il existe plusieurs mesures principales utilisées dans la surveillance intelligente, y compris la reconnaissance faciale et la reconnaissance de la marche.

Une fois que la catégorie à laquelle appartient les entités en mouvement est déterminée, le processus d'identification est initié, nécessitant des systèmes d'analyse puissants [2]. Diverses mesures clés sont employées dans le domaine de la surveillance intelligente, incluant la reconnaissance faciale et la reconnaissance de la démarche.

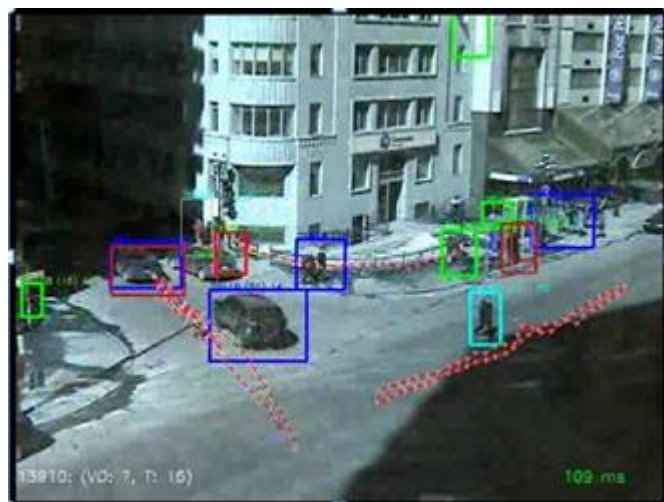


Figure 8. Classification d'objets (humains vs véhicules) [2]

1.5.4 Classification des Activités et Comportements

L'un des objectifs primordiaux de la surveillance intelligente réside dans l'interprétation des comportements individuels des objets en mouvement capturés

dans la scène. Cette interprétation implique la définition de modèles de comportement et de mouvement émergents dans l'image, ainsi que l'extraction de descriptions d'événements à un niveau d'analyse élevé. Cette tâche s'avère être un défi complexe auquel l'analyse vidéo est confrontée [2].

Dans le domaine du contrôle, diverses techniques telles que les modèles markoviens, les réseaux de neurones et les réseaux bayésiens sont employées pour modéliser la trajectoire et détecter des comportements spécifiques. Le principal objectif de la plupart de ces méthodes est de déclencher une alerte en cas d'incohérence avec le modèle comportemental déduit de la scène [2].

L'évaluation des comportements humains dans la vidéo demeure l'une des approches les plus prometteuses dans le champ de la vision industrielle, constituée en tant que dernière étape de traitement. Cette étape offre une explication sémantique des événements survenant dans la scène, en combinant les informations obtenues après un traitement rigoureux des séquences vidéo, en respectant les étapes précédentes (détection de mouvement, suivi d'objets et classification). Diverses fonctionnalités ont été proposées pour décrypter les comportements, englobant le suivi d'objets [11], la forme et les caractéristiques visuelles des objets [12], ainsi que les indications de mouvement dans l'image [13].

Les termes 'activité' et 'action' ont été diffusés dans de nombreuses littératures, avec 'action' dénotant de simples mouvements ordinaires d'une personne tels que la marche ou la nage. En revanche, 'activité' renvoie à une séquence d'actions complexes impliquant un groupe de personnes interagissant entre elles. Cette distinction entre les deux termes a été largement abordée dans la littérature [3], [5].

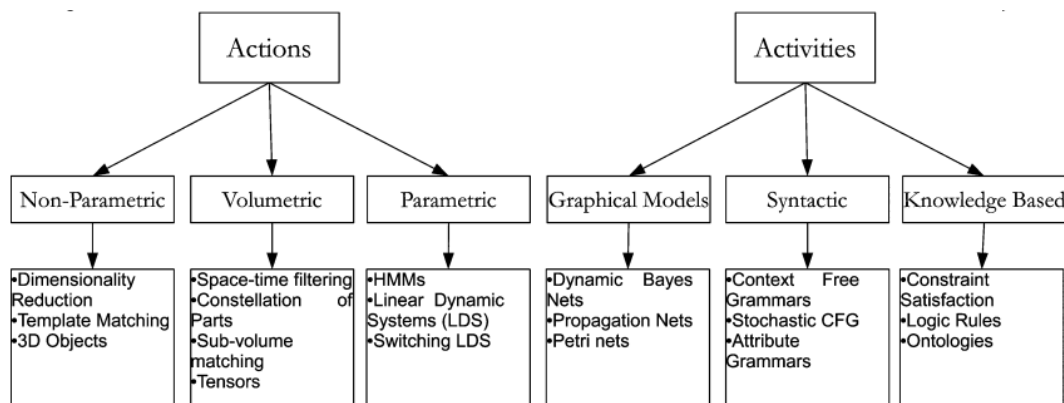


Figure 9. Un aperçu des approches de reconnaissance d'action et d'activité [14]

1.6 Vidéosurveillance Intelligente dans les Systèmes Multi-Caméras

Avec l'expansion croissante des réseaux de surveillance intelligents et la priorité accordée à leur développement par la recherche, il arrive fréquemment qu'un groupe de caméras soit déployé pour couvrir une zone donnée, permettant

ainsi la surveillance d'objets en mouvement sous divers angles. Ce groupe de caméras peut comporter une combinaison de dispositifs fixes et mobiles, offrant une flexibilité accrue pour la surveillance de l'environnement.

Au sein de ces configurations, les caméras interagissent entre elles en partageant les données capturées, lesquelles sont simultanément soumises à une analyse. Dans le cas où les caméras sont dotées de capacités intelligentes, un système central unique peut coordonner le groupe, centralisant ainsi l'échange d'informations et leur traitement via une unité centrale intermédiaire. Cette approche facilite particulièrement le suivi d'objets sur des champs de vision étendus [2]. Les réseaux de surveillance multi-caméras ont recours à diverses technologies pour leur mise en œuvre [15].

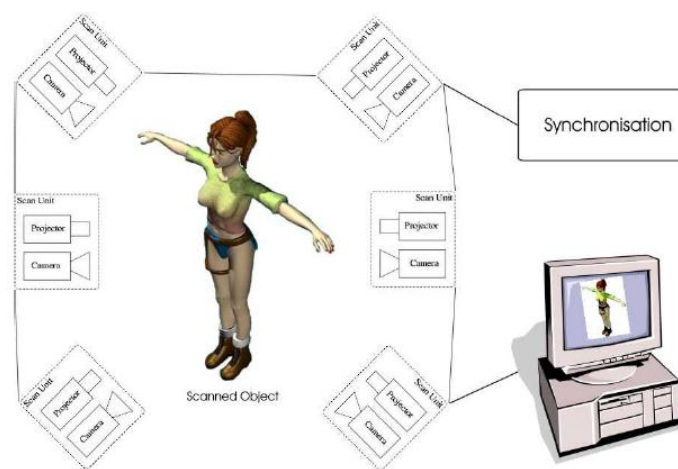


Figure 10. Flux d'informations dans un réseau multi-caméras[16]

1.6.1. Calibration

La calibration consiste à enregistrer diverses vues de caméras dans un système de coordonnées commun en définissant à la fois les paramètres internes (comme la géométrie et les propriétés optiques de la caméra et les paramètres externes (représentant la relation entre le cadre en 3D de la caméra et la position de la scène dans un système de coordonnées global arbitraire).

1.6.2. Topologie du Réseau (Network topology)

L'étude de la topologie des réseaux de caméras concerne les transitions temporelles, les champs de vision (qu'ils se chevauchent ou non) et la convergence spatiale des caméras. Cette approche permet d'aménager une vaste zone d'observation pour analyser des objets sous divers angles à partir de différents points de vue dans la scène [16].

1.6.2.1. Reconnaissance(Recognition)

L'analyse de la topologie du réseau permet d'explorer la détection et le suivi d'objets à travers différentes caméras [18].

1.6.2.2. Analyse d'activité (Activity analysis)

Cette perspective de recherche se focalise sur la reconnaissance et l'analyse de diverses activités, visant à caractériser les comportements humains spécifiques [17].

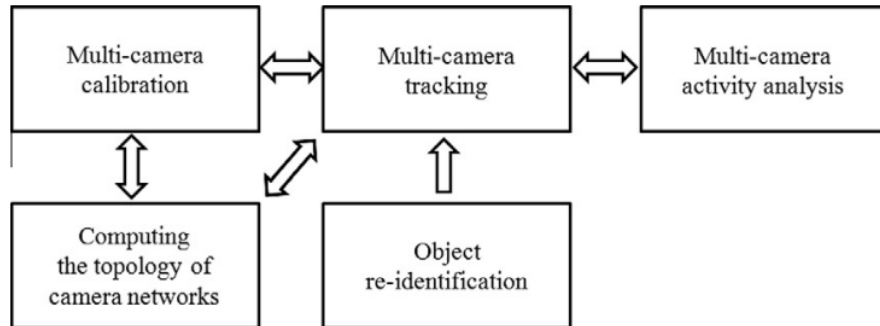


Figure 11. Les informations circulent entre différentes unités sur des réseaux multi-caméras [6]

1.7. Conclusion

Ce chapitre a offert une vue d'ensemble complète de la vidéosurveillance intelligente, en englobant un large éventail de ses domaines d'application. De plus, une analyse approfondie des types de caméras les plus couramment employés dans ce contexte a été entreprise. En outre, une exploration détaillée des étapes fondamentales de l'analyse vidéo, qui s'avèrent cruciales pour établir une surveillance intelligente efficace, a été réalisée. En conclusion, ce chapitre s'est achevé par une investigation approfondie des réseaux multi-caméras.

Le prochain chapitre sera dédié à la détection des objets en mouvement, à travers une étude analytique minutieuse de la littérature existante, tout en examinant attentivement les approches à la fois populaires et récentes qui sont pertinentes dans ce domaine.

CHAPITRE 2

DETECTION D'OBJETS EN MOUVEMENT

2.1 Introduction

La détection d'objets mobiles revêt une importance significative en tant que tâche de vision de bas niveau dans la compréhension des événements de haut niveau, notamment le comportement des objets au sein d'une scène, dans les systèmes de vision modernes pour la surveillance intelligente. L'automatisation de la détection d'objets en mouvement est devenue une étape incontournable pour la localisation spatiale des cibles, l'extraction des caractéristiques et l'initialisation automatisée des algorithmes de suivi d'objets. Cette détection contribue à classifier les pixels dans une séquence vidéo en tant qu'éléments appartenant à l'avant-plan (objets en mouvement) ou à l'arrière-plan. Dans la littérature, plusieurs approches ont été proposées pour parvenir à cette classification.

2.2 Les termes utilisés

Les algorithmes de détection de mouvement engendrent une variété de termes qui ont circulé dans la littérature. Dans de nombreux cas, différents termes ont été employés par divers auteurs pour désigner des algorithmes et des méthodes poursuivant le même objectif. Ainsi, l'organisation et la classification de ces algorithmes se révèlent être une tâche complexe.

2.2.1 Détection de Mouvement (Motion detection)

Ce terme englobe une notion générale visant à repérer les points de mouvement au sein d'une scène. Cet objectif incite l'algorithme à traiter deux variables : une variable quantitative (la quantité de mouvement) et une variable qualitative (une valeur booléenne) [18].

2.2.2 Estimation du Mouvement (Motion estimation)

Ce terme désigne quasiment le même objectif que le terme précédent, qui est de localiser les points de mouvement dans la scène. Cependant, l'algorithme ne considère qu'une variable quantitative, et le résultat est également quantitatif (la quantité de mouvement). Ce terme est souvent utilisé dans le contexte de l'estimation du flot optique, générant un champ de vecteurs à deux dimensions projeté sur l'image pour représenter un mouvement tridimensionnel [18].

2.2.3 Modélisation de l'Arrière-plan

Le terme englobe toutes les méthodes de détection de mouvement qui impliquent la création d'un modèle de fond. Ce modèle peut prendre la forme d'une image sans mouvement, comme illustré dans [19], un modèle statistique décrivant la

fonction de distribution des niveaux de gris ou des couleurs en tout point (c'est de loin l'approche la plus fréquente depuis 1997 [20]. ou encore une base d'images caractéristiques qui constitue un sous-espace vectoriel dans lequel on considère que les pixels représentant l'arrière-plan vont se trouver [18] [21].

2.2.4 Soustraction de l'Arrière-plan (Background subtraction).

Ce processus de détection de mouvement opère en soustrayant la valeur absolue du modèle d'arrière-plan de l'image actuelle, étant donné que le modèle d'arrière-plan constitue une image. Si le modèle adopté est statistique, la probabilité est calculée pour chaque pixel [18].

2.2.5 Segmentation de Mouvement

Cette méthode se concentre sur la division de l'image en régions caractérisées par une homogénéité de mouvement. Cette division joue un rôle fondamental dans l'unité de suivi et de détection du mouvement, et s'appuie sur des approches telles que l'estimation du flot optique [22] ainsi que les dérivées spatio-temporelles de l'intensité lumineuse [23].

2.3 Travaux de la littérature

La plupart des publications dans ce domaine sont orientées vers la détection de mouvement, qui est impliquée dans de nombreuses applications d'analyse d'images et de vidéos. La recherche contemporaine s'est focalisée sur le suivi et l'estimation des postures dans des contextes naturels, démontrant des progrès significatifs vers la compréhension automatique des comportements humains et des actions dans les vidéos. Au début des années 2000, la détection de mouvement était perçue comme un secteur de recherche en pleine effervescence. Dans cette recherche [24], les tendances récentes dans la détection et l'analyse du mouvement dans les vidéos ont été exposées, avec un aperçu des évolutions intervenues dans ce domaine entre 2000 et 2006.

En 2003, une étude a été présentée dans le contexte de la détection de mouvement [10]. Cet examen a abordé diverses méthodes liées au mouvement humain et à la segmentation de régions en mouvement.

En 2004, un état de l'art sur les différents systèmes de vidéosurveillance est présenté [25]. La détection de mouvement y est présentée comme la première étape de tout système de vidéosurveillance.

En 2005, l'attention s'est portée sur les méthodes de modélisation et de soustraction du fond appliquées aux scènes routières [26]. L'auteur a mené une comparaison entre des algorithmes de soustraction du fond pour la détection de véhicules en mouvement dans un environnement urbain. Piccardi a passé en revue et classifié plusieurs techniques de soustraction du fond, les comparant en termes de vitesse, de besoins en mémoire et de précision [27].

K. Toyama et al. [28] ont proposé un système de vidéosurveillance composé de trois unités : détection de mouvement, extraction de région d'intérêt et suivi [28]. Ils ont comparé leur algorithme avec huit autres méthodes de modélisation d'arrière-plan antérieures à 1999. Leur système, appelé Wallflower, s'attelle à résoudre divers problèmes courants relatifs à la maintenance de l'arrière-plan, en proposant cinq opérations fondamentales.

1. La différenciation sémantique des objets ne doit pas être prise en charge par le module de maintenance de l'arrière-plan.
2. La soustraction de l'arrière-plan doit segmenter les objets d'intérêt dès leur première apparition (ou réapparition) dans une scène.
3. La définition d'un critère approprié de stationnarité au niveau du pixel ; les pixels répondant à ce critère sont déclarés comme arrière-plan et ignorés.
4. L'adaptation du modèle d'arrière-plan aux changements soudains et progressifs de l'arrière-plan.
5. La prise en compte des changements à différentes échelles spatiales par les modèles d'arrière-plan.

M. Pless a élaboré un modèle statistique pour une apparence vidéo dynamique typique, et a comparé différents modèles statistiques caractérisant la dynamique locale des vidéos [29]. Par la suite, les travaux ont porté sur la détection du changement entre deux vues d'une même scène ; une étude est présentée sur les algorithmes de détection du changement et une analyse des principes d'évaluation de leurs performances [21].

2.4 Différentes Taxonomies

Au cours des dernières années, de nombreuses méthodes et algorithmes d'extraction de mouvement ont été proposés dans le domaine de l'analyse vidéo. Des travaux ont été publiés afin de proposer des méthodes et des techniques performantes pour répondre à différents défis tels que les objets solides, le fond dynamique, l'éclairage variable, etc... La classification de ces méthodes n'est pas une tâche facile. La plupart des travaux actuels sont orientés vers des applications précises telles que la réalité virtuelle, les systèmes de surveillance intelligents et autres des systèmes modernes.

Toyama et al. [28] ont choisi neuf algorithmes considérés comme classiques. qu'ils ont implémentés et testés sur un jeu de séquences vidéo illustrant les difficultés les plus fréquemment rencontrées. Il s'agit d'une classification par l'exemple sans pour autant que soit précisé ce qui rend chaque exemple représentatif d'une classe bien septique d'algorithmes.

Dans leurs travaux, Moeslund et al. mettent en évidence des spécificités qui caractérisent les algorithmes [24]. Tout d'abord, les auteurs incluent l'espace colorimétrique dans l'implémentation en arrière-plan, y compris les modèles statistiques et un modèle binaire basé sur des chaînes pour le codage de voisinage [30]. La seconde caractéristique notée par les auteurs est la méthode

de validation des pixels étiquetés comme n'appartenant pas à l'arrière-plan, Il s'agit en fait d'un post-traitement à la détection de mouvement. Le troisième élément distinctif est la méthode de mise à jour du modèle de l'arrière-plan, que ce soit par addition pondérée des nouvelles observations, ou par l'ajout de nouveaux modes dans un modèle statistique multimodal.

La classification proposée par Piccardi [27] distingue les méthodes de base et les méthodes plus élaborées. Parmi les méthodes de base, on trouve la modélisation du fond par différence absolue entre images successives [19]. La deuxième catégorie de méthodes concerne : la modélisation du fond par Gaussien simple [20], le Mélange gaussien [31], par estimation non paramétrique des densités ponctuelles[32], la modélisation par algorithme sous-espace vectoriel [21] ainsi que la Modélisation par l'algorithme Mean-Shift [33].

D'autres travaux présentent des méthodes de segmentation de fond, y compris des algorithmes statistiques, des algorithmes basés sur la différence entre deux ou trois images consécutives et des algorithmes de flux optique [10], [25].

Cheung et Kamath ont proposé une classification binaire qui divise les algorithmes en deux catégories [26] : les algorithmes récursifs et les algorithmes non récursifs. Les méthodes non récursives sont celles qui utilisent un stock des dernières images observées et la détection de mouvement est le résultat des statistiques calculées sur l'ensemble des images du stock. Leur défaut majeur est l'espace mémoire nécessaire au stockage du buffer. Parmi ces méthodes, on compte la différence absolue entre images consécutives, la modélisation de l'arrière plan par la médiane des dernières images observées, le filtrage prédictif linéaire [28] et l'estimation non paramétrique des densités de probabilité d'appartenance à l'arrière-plan. L'algorithme récursif est celui dont le modèle de l'arrière-plan n'est pas un buffer des dernières observations. Ils ont le mérite de nécessiter moins d'espace de stockage, mais font courir le risque de voir des erreurs demeurer dans le modèle pendant plus longtemps. Les modèles statistiques paramétriques font partie de cette famille.

Yilmaz et al. ont utilisé une classification qui distingue les méthodes de soustraction du fond en les divisant en deux catégories [9]: la première concerne les méthodes locales, qui créent un modèle statistique en tout point de l'image sans recourir à la proximité spatiale. Ce groupe comprend des méthodes qui dépendent de la différence entre des images successives, y compris des méthodes qui construisent un modèle gaussien ou un multi-gaussien en tout point de l'image. La deuxième catégorie concerne les méthodes semi-locales, qui sont des méthodes qui prennent en compte le voisinage des pixels lors de la création d'un modèle d'arrière-plan, et incluent un ensemble de méthodes qui s'appuient sur un modèle de texture pour distinguer les images [30].

En 2015, Bousetouane a élaboré une classification des méthodes de détection de mouvement basée sur la modélisation du fond [4]. L'intérêt de deux grandes familles de méthodes est illustré dans la figure suivante.

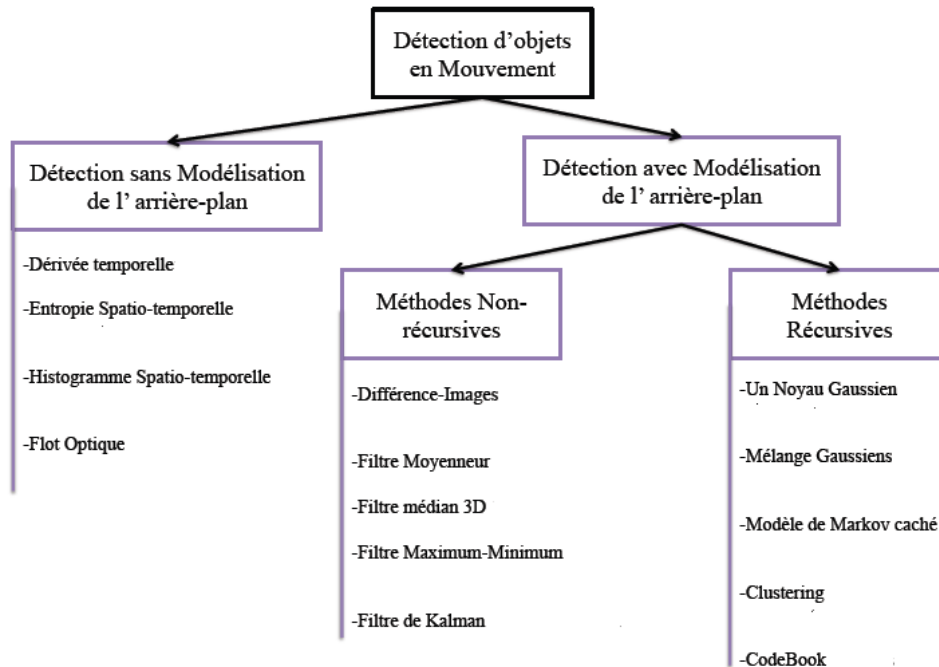


Figure 12. Taxonomie possible des méthodes de détection d'objets en mouvement [4]

2.5 Méthodes de Détection de Mouvement

2.5.1 Détection avec Modélisation de l'Arrière-plan

2.5.1.1 Modélisation de l'Arrière-plan

La soustraction du fond (Background subtraction) représente une catégorie de techniques couramment employées pour segmenter des objets en mouvement dans une scène, notamment dans des applications de surveillance [34]. Cette méthode consiste à détecter les régions en mouvement en effectuant une soustraction pixel par pixel entre l'image actuelle et une image de référence d'arrière-plan. Les pixels présentant une différence supérieure à un seuil sont considérés comme faisant partie de l'avant-plan. Une fois un masque d'avant-plan créé, des opérations morphologiques de post-traitement, telles que l'érosion et la dilatation, sont appliquées pour réduire les effets du bruit et renforcer les régions détectées. La mise à jour régulière de l'image de référence permet d'adapter le modèle d'arrière-plan aux changements dynamiques de la scène au fil du temps [3].

2.5.1.2 Simple Gaussienne

Parmi les approches utilisées, la méthode de la simple gaussienne (Simple Gaussian) vise à modéliser la distribution temporelle de l'intensité de chaque pixel dans l'image courante de la séquence vidéo à l'aide d'une distribution gaussienne simple à l'instant t . Pour cela, la valeur M est utilisée pour déterminer le nombre de pixels précédents servant d'échantillon pour construire une fonction de distribution de probabilité. Cette fonction est caractérisée par un écart-type moyen. Cependant, cette méthode requiert le stockage des intensités des M pixels précédents pour chaque image de la séquence, ce qui engendre une utilisation conséquente de mémoire et ralentit l'exécution en temps réel du processus. Pour pallier à cet inconvénient, des chercheurs ont proposé une méthode plus performante, basée sur la moyenne glissante $I_t = I(X; Y; t)$ où l'intensité du pixel en question est notée X à l'instant t , et la moyenne glissante est déterminée par Bt . Cette approche adaptative permet la mise à jour du fond avec les informations de chaque image de la séquence, contribuant ainsi à former le fond à partir des pixels qui ne changent pas par rapport à l'image précédente. Cette méthode prend en considération les variations du bruit du capteur, qui altère l'uniformité de l'arrière-plan, ainsi que les changements progressifs de la luminosité de la scène. Le bruit du capteur est modélisé comme un bruit gaussien pour les paramètres estimés $\sigma_{cam}^2 = (\sigma_{rcam}^2, \sigma_{gam}^2, \sigma_{bam}^2)$ en prenant en compte les trois canaux RGB. De plus, chaque pixel du modèle d'arrière-plan est modélisé sous la forme d'une distribution gaussienne pour prendre en compte le bruit. L'image de référence est régulièrement mise à jour afin de refléter les changements progressifs de luminosité. Les différentes valeurs prises par le pixel (x, y) au fil du temps constituent le processus à analyser [20], [35].

$$x1: xT = \{x1, \dots, xT\} = \{I(x, y, t): 1 \leq t \leq T\} \quad (1)$$

Chaque pixel est caractérisé par une distribution gaussienne de moyenne $\mu t = (\mu r t, \mu g t, \mu b t)$ et une variance $\sigma^2 t = (\sigma^2 r t, \sigma^2 g t, \sigma^2 b t)$ pour les 3 canaux RGB:

$$\eta(x_t, \mu, \Sigma) = \frac{1}{2\pi^{n/2} |\Sigma|^{1/2}} e^{-\frac{1}{2}(x_t - \mu)^T \Sigma^{-1} (x_t - \mu)} \quad (2)$$

Où n : nombre de canaux (généralement $n=3$ pour RGB) et $\Sigma = \sigma I$ (où I est la matrice identité). La mise à jour des paramètres que représentent les équations 3 et 4 :

$$\mu_{t+1} = \alpha\mu_t + (1 - \alpha)X_{t+1} \quad (3)$$

$$\sigma_{t+1}^2 = \alpha(\sigma_t^2 + (\mu_{t+1} - \mu_t)^2) + (1 - \alpha)(X_{t+1} - \mu_{t+1})^2 \quad (4)$$

Où : X_{t+1} est la valeur du pixel $I(x, y, t+1)$ et α est un paramètre de contrôle du taux d'adaptation ($0 < \alpha < 1$).

Un pixel est classé en mouvement si :

$$|X_{rt} - \mu_{rt}| > 3 \max(\sigma_{rt}, \sigma_{rcam}) \quad (5)$$

$$|X_{gt} - \mu_{gt}| > 3 \max(\sigma_{gt}, \sigma_{gcam}) \quad (6)$$

$$|X_{bt} - \mu_{bt}| > 3 \max(\sigma_{bt}, \sigma_{bcam}) \quad (7)$$

La valeur de σ_{cam}^2 est une très petite valeur correspondant au faible bruit [36].

2.5.1.3 Mélange de Gaussiennes (Mixture of Gaussians-MoG)

L'arrière-plan de la scène vidéo contient des objets non stationnaires tels que des feuilles et des branches d'arbres dont le mouvement dépend du vent qui affecte l'intensité des pixels. De ce fait, une simple modélisation gaussienne de l'arrière-plan s'avère dans ce cas non suffisante. Alternativement, une méthode de modélisation de fond a été proposée par un mélange de gaussiennes par Grimson et al. [37]. Cette méthode a été notamment utilisée dans plusieurs recherches pour la modélisation d'arrière-plan complexe et statique [31], [38], [39]. La distribution temporelle de l'intensité de chaque pixel dans le temps t est modélisée par un mélange gaussien k . Cette modélisation permet de représenter l'image de fond par une distribution multimodale dans un mélange gaussien. Le mélange gaussien est caractérisé par une moyenne gaussienne, une matrice de covariance, et probabilité a priori pour chaque nombre k . Ces informations sont mises à jour dans chaque image de la séquence Vidéo par pixel I_t . Chaque gaussien k de la distribution correspond à la probabilité d'observer une certaine intensité, faisant du mélange gaussien une représentation plus générale par rapport à une simple gaussienne. Stauffer et al. suggèrent la probabilité d'observer l'intensité des pixels de l'image [39]. Chaque pixel est comparé à l'ensemble actuel de modèles pour trouver une correspondance. Les paramètres du modèle correspondant sont mis à jour en fonction du facteur d'apprentissage. Si la correspondance n'existe pas, le modèle le moins probable est supprimé et remplacé par le nouveau Gaussien avec initialisation à la valeur du pixel actuel comme montré ci-dessous :

$$I(x, y, t) = I_t = \sum_{i=1}^k w_{i,t} * \eta(I(x, y, t), \mu_{i,t}, \Sigma_{i,t}) \quad (8)$$

$$P_x(Z_t) = \sum_{i=1}^k w_{i,t} \eta(z_t, \mu_{i,t}, \Sigma_{i,t}) \quad (9)$$

Où : K est le numéro de la distribution gaussienne,

$w_{i,t}$ est le poids donné à la $i^{\text{ème}}$ gaussienne au temps t,

η est la fonction de densité de probabilité gaussienne,

$\mu_{i,t}$ et $\Sigma_{i,t}$ sont respectivement la moyenne et la matrice de variance-covariance de la $i^{\text{ème}}$ gaussienne au temps t.

Chaque pixel $P_x(Z_t)$ de l'arrière-plan est modélisé par un mélange de k gaussiennes.

Stauffer et al. [31] ont préconisé une valeur de k située dans l'intervalle entre 3 et 5. Ils ont considéré une composante de couleur indépendante avec une variance uniforme, soit $\Sigma_{i,t} = \sigma_{i,t}^2 I$. Pour chaque nouvelle observation z_t , la distribution gaussienne qui présente la probabilité la plus élevée d'expliquer z_t est actualisée. On considère que la gaussienne i explique z_t si : $\|z_t, \mu_{i,t}\| \leq 2.5\sigma_{i,t}$. La mise à jour de la gaussienne i se déroule de la manière suivante [4] [18]:

$$w_{i,t} = (1 - \alpha)w_{i,t} + \alpha \quad (10)$$

$$\mu_{i,t} = (1 - \rho)\mu_{i,t} + \rho z_t \quad (11)$$

$$\sigma_{i,t}^2 = (1 - \rho)\sigma_{i,t-1}^2 + \rho(z_t - \mu_{i,t})^T (z_t - \mu_{i,t}) \quad (12)$$

Avec $\rho = \alpha \eta(z_t, \mu_{i,t}, \Sigma_i)$, la constante de temps ; $1/\alpha$ déterminant la vitesse à laquelle les paramètres sont mis à jour.

2.5.1.4 Modèles statistiques fondés sur des distributions gaussiennes :

Parmi les méthodes basées sur la distribution gaussienne, nous distinguons : Pfinder [40] GrimsonGMM [31], ZivkovicGMM [41];[42] Lopez-RubioAE [43] et ElgammalKDE [32] .

Modèle Wren (Pfinder)

Cette méthode vise à modéliser la couleur de chaque pixel de l'image en utilisant une distribution gaussienne pour représenter l'arrière-plan. En revanche, pour la scène de premier plan, une approche différente est adoptée, avec l'utilisation d'une classe A attribuée à chaque point appartenant au premier plan. La probabilité d'appartenance à la classe A pour un pixel donné est quantifiée par

la fonction $P_a(x_t)$ donnée ci-dessous. Ici, x désigne le pixel considéré, et la classe correspondant à ce pixel à l'instant t est définie par la fonction $a_t(x)$.

$$P_a(x_t) = \eta(x_t, \mu_{a,t}(x), \Sigma_{a,t}(x)) \quad (13)$$

$$\eta(X, \mu, \Sigma) = (2\pi)^{-D/2} \det(\Sigma)^{-1/2} \quad (14)$$

$$Pr(X) = \exp\left(-\frac{1}{2}(X - \mu)^T \Sigma^{-1} (X - \mu)\right) / (2\pi^{m/2} |\Sigma|^{1/2}) \quad (15)$$

$$\mu_{a,t}(X) = E(X_t(x) | a_t(x)) \quad (16)$$

$$\Sigma_{a,t}(x) = E[(X_t(x) - \mu_{a,t}(x))(X_t(x) - \mu_{a,t}(x))^T | a_t(x)] \quad (17)$$

Un pixel est considéré comme faisant partie de l'arrière-plan si la probabilité associée à la classe d'arrière-plan dépasse un seuil minimum (T), soit $P_b(x_t) > T$.

Modèle de Stauffer et Grimson

Le modèle élaboré par Stauffer et Grimson [31] repose sur une approche paramétrique pour la modélisation de l'arrière-plan. Il fait appel à une combinaison de trois à cinq distributions gaussiennes K pour représenter l'arrière-plan, et la probabilité de la couleur d'un pixel est évaluée selon l'expression suivante :

$$P(x_t) = \sum_{i=1}^K w_{i,t} \eta(x_t, \mu_{i,t}(x), \Sigma_{i,t}(x)) \quad (18)$$

Où : x_t est le pixel, $w_{i,t}$ est la k probabilité gaussienne au temps t , $\Sigma_{i,t}$ est la matrice de covariance, $\mu_{i,t}$ est la moyenne, η est une distribution de probabilité k gaussienne, la distribution numérique gaussienne est K .

$$\Sigma_k = \begin{pmatrix} 1\sigma_k^2 & 0 & 0 \\ 0 & 2\sigma_k^2 & 0 \\ 0 & 0 & 3\sigma_k^2 \end{pmatrix} \quad (19)$$

$$\Sigma_{i,t} = \sigma_k^2 x_t \quad (20)$$

Les composantes de couleur (R, G, B) de chaque pixel sont supposées indépendantes du même contraste, et la gaussienne est mise à jour à chaque nouvelle observation, Cette condition est remplie avec $\|X_t - \mu_{(i,t)}\| \leq 2.5\sigma_{(i,t)}$, la mise à jour gaussienne se déroule comme suit :

$$w_{i,t} = (1 - \alpha)w_{i,t} + \alpha \quad (21)$$

$$\mu_{i,t} = (1 - \rho)\mu_{i,t} + \rho x_t \quad (22)$$

$$\sigma_{i,t}^2 = (1 - \rho)\sigma_{i,t-1}^2 + \rho(x_t - \mu_{i,t})^T(x - \mu_{i,t}) \quad (23)$$

Où : $\rho = \alpha\eta$ est la constante de temps,

L'un des défis auxquels le modèle de Stauffer et Grimson fait face est la détermination du nombre de composantes de distribution caractéristiques [44], [45].

Modèle KDE (Kernel density estimation)

Dans ce modèle, afin d'estimer la fonction de densité de probabilité $P(x_t)$, les chercheurs ont utilisé l'estimateur à noyau K en adoptant une approche non paramétrique [32] :

$$P(x_t) = \frac{1}{N} \sum_{i=1}^N K(x_t - x_i) \quad (24)$$

La dernière densité de pixels observée par B est l'ensemble de N éléments, en supposant que les canaux de couleur sont indépendants et K est une fonction de noyau gaussien comme suit :

$$P(x_t) = \frac{1}{N} \sum_{i=1}^N \prod_{j=1}^M \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{1}{2} \frac{(x_{t_j} - x_{i_j})^2}{\sigma_j^2}} \quad (25)$$

Où : m_j est le médian et M est le nombre de canaux, chaque binaire successif $(x_i, x_{i+1}) \in A$ est calculé indépendamment pour chaque canal $|x_{i_j} - x_{i+1_j}|$. Avec l'équation obtenue, $P(x_t) \geq T$, le pixel fait partie de l'arrière-plan où T est la constante de seuil. Le pixel dans cette méthode appartient au premier plan si la condition suivante est constante $((P_L(x_t) \geq T) \wedge (P_S(x_t) \geq T))$ où P_L et P_S sont respectivement les modèles à long terme et à court terme, d'une fonction de densité de probabilité.

Dans le modèle longue portée, les N derniers pixels d'intensité sont pris en compte sans condition supplémentaire. Pour le modèle à court terme, les N intensités marquées comme arrière-plan dans les itérations précédentes sont prises en compte. L'espace colorimétrique de trois unités (r, g, s) dépendant de (R, G, B) est alors utilisé :

$$r = \frac{R}{R + G + B} \quad g = \frac{G}{R + G + B} \quad s = R + G + B \quad (26)$$

Le pixel est représenté par $x_t = [r_t, g_t, s_t]$ où S est l'échelle de luminance.

$$C = \left\{ x_i | x_i \in A, B_1 \leq \frac{S_t}{S_i} \leq B_2 \right\} \quad (27)$$

Où C est la densité de couleur du pixel S_t , afin de réduire les faux positifs B_1 et B_2 A et B sont déterminés par des valeurs fixes [44].

Modèle GMM de Zivkovic

Zivkovic et Van Der Heijden ont apporté des améliorations au modèle de Stauffer et Grimson en introduisant une distribution de mélange avec des composantes adaptables [41] [42]. Le nombre de composantes est défini comme K_t^x pour un pixel x_t $1 \leq K_t^x \leq K$. Seul le nombre de pixels requis est utilisé à chaque instant, c'est pourquoi cette méthode permet d'obtenir un temps de traitement plus efficace par rapport à [9] et de meilleurs résultats. La fonction de mise à jour du poids est ensuite définie par :

$$w_{i,t} = w_{i,t-1} + \alpha(M_{i,t} - w_{i,t-1}) - \alpha cT \quad (28)$$

où cT est une constante. Initialement, chaque distribution de mélange est formée d'une seule composante centrée sur le premier échantillon d'entrée. Dans les images suivantes, s'il n'y a aucun composant qui correspond à l'entrée, c'est-à-dire $\forall i \leq K_t^p, M_{i,t} = 0$, un nouveau composant est créé, centré sur la valeur actuelle du pixel avec $w_{i,t} = \alpha$ et un haut variance $\sigma_{i,t}^2$. Si aucune correspondance n'est trouvée et que $K_t^p = K$, alors la distribution la moins probable est remplacée par une nouvelle distribution comme l'équation (28).

Modèle de Lopez-Rubio

Lopez-Rubio et al. ont proposé un modèle qui représente la scène en utilisant une seule distribution gaussienne pour l'arrière-plan et une distribution uniforme pour l'avant-plan [44].

$$P(x_t) = \pi_B G(P_t | \mu_B, \Sigma_B + \Phi) + \pi_B U(x_t) \quad (29)$$

$$U(x_t) = \begin{cases} 1/Vol(H) & \text{if } P_t \in H \\ 0 & \text{Otherwise} \end{cases} \quad (30)$$

où H est le support de la pdf uniforme, $Vol(H)$ est le volume des dimensions d de H , Φ est une matrice diagonale constante, et Σ_B n'est pas limité à être une matrice diagonale donc la composante gaussienne $G(P_t | \mu_B, \Sigma_B + \Phi)$ est capable de

capturer une plus large gamme d'arrière-plans dynamiques que les propositions précédentes [44].

2.5.1.5 Les modèles flous

Récemment, certains auteurs ont introduit des concepts flous dans les différentes étapes du processus de soustraction de fond [50]. Dans [46], les auteurs effectuent la soustraction de l'arrière-plan par la mesure de similarité des caractéristiques de couleur et de texture de l'image d'entrée et du modèle d'arrière-plan en utilisant Sugeno Integral [47]. Plus tard, El Baf et Boumans [48] ont obtenu de meilleurs résultats avec Choquet Integral [49]. M. Azab et al. [50] ont utilisé le Choquet Integral avec des caractéristiques de couleur, de bord et de texture. Dans [51], les auteurs ont proposé une fonction floue pour calculer l'extraction du premier plan et mettre à jour le modèle d'arrière-plan. Pour mettre à jour le modèle de fond avec Fuzzy Running Average, M. H. Sigari [51] suggère l'utilisation d'une fonction de saturation linéaire au lieu de la fonction de limiteur nette.

2.5.1.6 CodeBook

Cette méthode proposée par [52] pour la modélisation d'arrière-plan, est basée sur l'historique d'observations de chaque pixel dans une séquence vidéo est utilisée pour construire un codebook pour chaque pixel. Ce codebook est caractérisé par un ou plusieurs codewords. L'algorithme fonctionne en deux étapes : l'étape d'apprentissage et l'étape de mise à jour. L'apprentissage consiste à définir le modèle de fond utilisé pour comparer les images à traiter. Ce modèle est ensuite mis à jour avec de nouvelles images pour rendre compte des changements qui s'opèrent dans le paysage. Cette méthode est puissante et efficace dans un grand nombre de cas d'utilisation, y compris les arrière-plans dynamiques (par exemple, les feuilles, les fontaines, les bords de mer, les drapeaux...) et les changements subtils d'éclairage. Ce modèle est représenté par une liste de livres de codes (1 par pixel) contenant chacun N mots de passe. Un mot de code (codeword) est généré (ou mis à jour) si le pixel observé est similaire à un mot de code existant à chaque itération d'apprentissage [53].

Chaque codeword $C_L, C_L = 1, \dots, x$ est représenté par un vecteur RVB , $V_L = (R, V, B)$.et un Hexuplet $AxL = \langle I_L, I_L, F_L, \lambda_L, P_L, Q_L \rangle$, où $I_L = \min(I_L, I_L)$, $I_L = \max(I; I_L)$ représente respectivement le minimum et le maximum de la densité attribuée au "codeword". F_L est la fréquence de correspondance du codeword. λ_L la durée pendant laquelle ce mot de code n'a pas été mis à jour. P_L Et Q_L x et y représentent respectivement le premier et le dernier temps d'accès au mot de code. Le changement de couleur λ_L entre un pixel et un mot de code est calculé comme suit : Il est défini par deux vecteurs qui contiennent des valeurs RGB pour le codeword et de nombreuses autres données telles que les valeurs de luminosité minimale et maximale, les informations temporelles et la fréquence de surveillance du codeword.

$$ALT\ Couleur(X_L, V_L) = \delta = \sqrt{\|X_t\|^2 - C_p^2} \quad (31)$$

$\|X_t\|^2$ et C_p^2 sont calculés comme suit :

$$\|X_t\|^2 = R^2 + V^2 + B^2 \quad (32)$$

C_p^2 est l'auto-corrélation entre les composants colorés d'un pixel R; V;B et le codeword, qui est normalisé par la brillance:

$$C_p^2 = \|X_t\|^2 \cos^2 \theta = \frac{(R_i R + V_i V + B_i B)^2}{R_i^2 + V_i^2 + B_i^2} \quad (33)$$

Initialement, le codebook de chaque pixel est vide et ne contient pas des codeword. Quand un échantillon de pixel est observé alors cet échantillon sera classé comme un codeword dans le codebook, et les valeurs seront utilisées pour estimer la mesure qui représente le codeword. Si un codeword existe déjà dans le codebook alors l'échantillon de ce nouveau pixel est comparé avec chaque codeword dans le codebook utilisant la mesure d'altération de la couleur .[4], [52]. D'autre chercheurs[54] prouvent l'insuffisance de cette méthode dans certaines conditions réelles d'application comme : la présence de mouvement lent dans le fond, la présence d'objets avec des couleurs semblables, etc.

Algorithm for Codebook construction

- I. $L \leftarrow 0^1$, $\mathcal{C} \leftarrow \emptyset$ (empty set)
 - II. **for** $t = 1$ to N **do**
 - (i) $\mathbf{x}_t = (R, G, B)$, $I \leftarrow \sqrt{R^2 + G^2 + B^2}$
 - (ii) Find the codeword \mathbf{c}_m in $\mathcal{C} = \{\mathbf{c}_i | 1 \leq i \leq L\}$ matching to \mathbf{x}_t based on two conditions (a) and (b).
 - (a) $colordist(\mathbf{x}_t, \mathbf{v}_m) \leq \varepsilon_1$
 - (b) $brightness(I, \langle \hat{I}_m, \hat{I}_m \rangle) = \mathbf{true}$
 - (iii) If $\mathcal{C} = \emptyset$ or there is no match, then $L \leftarrow L + 1$. Create a new codeword \mathbf{c}_L by setting
 - $\mathbf{v}_L \leftarrow (R, G, B)$
 - $\mathbf{aux}_L \leftarrow \langle I, I, 1, t - 1, t, t \rangle$.
 - (iv) Otherwise, update the matched codeword \mathbf{c}_m , consisting of $\mathbf{v}_m = (\hat{R}_m, \hat{G}_m, \hat{B}_m)$ and $\mathbf{aux}_m = \langle \hat{I}_m, \hat{I}_m, f_m, \lambda_m, p_m, q_m \rangle$, by setting
 - $\mathbf{v}_m \leftarrow \left(\frac{f_m \hat{R}_m + R}{f_m + 1}, \frac{f_m \hat{G}_m + G}{f_m + 1}, \frac{f_m \hat{B}_m + B}{f_m + 1} \right)$
 - $\mathbf{aux}_m \leftarrow \langle \min\{I, \hat{I}_m\}, \max\{I, \hat{I}_m\}, f_m + 1, \max\{\lambda_m, t - q_m\}, p_m, t \rangle$.
 - III. **end for**
 - III. For each codeword \mathbf{c}_i , $i = 1, \dots, L$, wrap around λ_i by setting $\lambda_i \leftarrow \max\{\lambda_i, (N - q_i + p_i - 1)\}$.
-

Figure 13 Algorithme de codebook qui a été créé par Kim et al [52]

2.5.1.7 Méthodes Neurales et Neuro-floues

Fondamentalement, le réseau neuronal apprend à classer chaque pixel de l'image. Pour chaque pixel, le réseau de neurones détermine si un pixel appartient au premier plan ou à l'arrière-plan [47]. Dans [55], les auteurs ont utilisé un réseau neuronal multicouche à feed-forward avec 124 neurones. L'approche de

segmentation en arrière-plan proposée par [55] s'appuie sur un PNN (Probabilistic Neural Networks) adapté. Le modèle d'arrière-plan est appris par le réseau neuronal tandis qu'un classificateur bayésien identifie si un pixel appartient au premier plan ou à l'arrière-plan. [56] ont utilisé un réseau SOM (Self-Organizing Map) pour réaliser des BS. Chaque pixel possède une carte neuronale donnée par des vecteurs de poids. Plus tard, [57] a amélioré les travaux précédents en ajoutant une fonction floue dans l'étape d'apprentissage en arrière-plan. Les auteurs précédents appliquent également une analyse de cohérence spatiale sur le réseau SOM de chaque pixel pour améliorer la robustesse contre les fausses détections et pour traiter les problèmes de décision survenant généralement lorsque des paramètres précis sont impliqués.

2.5.1.8 Détection de l'Avant-plan

La carte de soustraction est générée par le processus de comparaison entre l'image courante de la séquence vidéo et l'arrière-plan estimé, permettant ainsi la détection du premier plan. Ce processus intervient après avoir modélisé l'arrière-plan à l'aide des méthodes précédentes. Les pixels du premier plan sont différenciés des pixels de l'arrière-plan pour obtenir le masque de premier plan (objets en mouvement). Cette distinction est réalisée au moyen d'un processus de seuillage binaire appliqué à la carte de soustraction. Dans ce contexte, il convient de mentionner la recherche menée par Garcia-Martin et Martinez [58] ; elle a permis de calculer la valeur absolue de la différence entre les pixels correspondants de l'arrière-plan estimé et de l'image courante pour extraire la carte de soustraction. Par la suite, Cucchiara et al [59] ont proposé une méthode reposant sur la différence entre les images d'arrière-plan et l'image courante pour obtention de la carte de soustraction [59].

$$F(I(x, y, t), B) = \begin{cases} 255 & \text{si } |I(x, y, t) - B| > T^\circ \\ 0 & \text{Sinon} \end{cases} \quad (34)$$

T° est un seuil prédéterminé dépend de l'application, la caméra et les données utilisées. B est l'arrière-plan estimé utilisant l'une des méthodes présentées précédemment.

2.5.2. Détection sans Modélisation de l'Arrière-plan

Les méthodes adoptées dans ce contexte se basent sur des mesures simples visant à estimer la variation temporelle de couleur entre deux images successives, telles que l'histogramme spatio-temporel. Dans la plupart des cas, ces méthodes sont qualifiées d'estimateurs de mouvement plutôt que de détection de mouvement.

2.5.2.1 Dérivée Temporelle

Cette approche est également connue sous le nom de "frame différence" [60]. Elle vise à détecter le mouvement et à mesurer les changements de luminosité des pixels entre deux images successives.

Soient I_{t-1} une image capturée à l'instant $t-1$ et I_t une image capturée à l'instant t , la dérivée temporelle est calculée dans le temps selon l'équation suivante :

$$\forall (x, y) \in \mathbb{R}^2 \forall (t) > 0$$

$$I_t(x, y, t) = |I(x, y, t - 1) - I(x, y, t)| \quad (35)$$

Malgré la simplicité de cette méthode, elle ne peut pas relever les véritables défis tels que le vent et le ralenti.

2.5.2.1. Flot Optique

Du point de vue physique, le flot optique représente la scène visuelle résultant du mouvement relatif entre l'observateur et la scène observée [60]. En analyse vidéo, le flot optique désigne un champ vectoriel bidimensionnel entre les pixels qui se déplacent d'une image à une autre au sein d'une séquence vidéo. Plusieurs études, ont présenté des méthodes pour estimer le flot optique, notamment des approches basées sur la corrélation, comme les méthodes différentielles proposées par Lucas et Kanade [61]. Cette méthode est réputée pour sa robustesse et sa rapidité, ce qui en fait l'approche la plus couramment employée [18], [4], [21].

$$\forall X \in \mathbb{R}^2 \forall (t) > 0 \quad I(X + d, t) - I(X, t - 1) = 0 \quad (36)$$

Le flux optique est estimé au point X et au vecteur d . La fonction d'erreur quadratique est calculée par le voisin N .

$$\varepsilon(d) = \sum_{x' \in N(x)} [I(x' + d, t) - I(x', t - 1)]^2 \quad (37)$$

L'hypothèse de conservation spatio-temporelle de l'éclairage de la scène est comme l'équation (38). En supposant que $I(X, t)$ est la luminance du point $X = (x, y)^T$ dans l'image au temps t .

$$\frac{dI(x, t)}{dt} = 0 \quad (38)$$

Cette conservation spatio-temporelle de l'intensité permet de calculer le vecteur de mouvement $\mathbf{d} = [d_x, d_y]^T$ du point \mathbf{x} , à partir des techniques basées sur le gradient. Dans le cas où le mouvement d est purement translationnel entre les

images consécutives, l'équation qui décrit ce mouvement de la scène est donné par.

$$I(x, t) = I(x + (d), t + 1) \quad (39)$$

$$\nabla I(x, t) + dI_t(x, t) = 0 \quad (40)$$

En développant $I(x + (d), t + 1)$ au premier ordre et en prenant en compte l'invariance spatiotemporelle, le flot optique d des points est donné par l'équation (40).

Où : $\nabla I(x, t) = (I_x(x, t), I_y(x, t))^T$ est la dérivée spatiale de l'image I dans la direction x et y respectivement, et $I_t(x, t)$ représente la dérivée partielle temporelle de l'image. Ainsi, les composantes du vecteur d sont obtenues en utilisant les hypothèses de cohérence spatiale et temporelle locale [62], [63].

2.6. Conclusion

La détection des objets en mouvement à travers l'analyse vidéo revêt une importance cruciale pour une variété d'applications utilisées dans divers domaines. Les efforts de recherche se sont concentrés sur le développement de la technologie d'analyse vidéo, notamment la surveillance intelligente et l'analyse du comportement humain. Les méthodes de modélisation de l'arrière-plan permettent de segmenter les pixels de la séquence vidéo en deux groupes distincts : ceux du premier plan (en mouvement) et ceux de l'arrière-plan. C'est ainsi que les méthodes de soustraction de l'arrière-plan se démarquent et s'imposent comme les plus largement utilisées, se distinguant ainsi des autres approches.

Dans ce chapitre, nous avons minutieusement examiné les méthodes de détection de mouvement, en faisant mention des classifications les plus répandues évoquées dans la littérature. Nous avons fourni des explications détaillées sur plusieurs méthodes couramment utilisées, mettant en avant les aspects les plus significatifs de chaque approche. Dans le prochain chapitre, nous explorerons un domaine tout aussi essentiel dans l'analyse vidéo, qui vient compléter la détection : le suivi des objets en mouvement.

CHAPITRE 3

SUIVI D'OBJETS EN MOUVEMENT

3.1 Introduction

Le suivi d'objets représente un pilier fondamental de la vidéosurveillance intelligente. Il incarne une problématique complexe au cœur de l'analyse vidéo, confrontée à une myriade de défis. Parmi ceux-ci, la variation dans l'apparence des objets joue un rôle prédominant. La similitude entre différents objets en mouvement, qu'elle soit fondée sur la couleur, la forme, la vitesse ou d'autres caractéristiques, engendre des perturbations lors de l'identification de la cible. Des éléments tels que la modification de la couleur d'un objet en fonction des variations d'éclairage, que ce soit en environnement clos ou en extérieur, ainsi que les situations d'occultation où la cible est masquée par un autre élément, doivent être pris en considération. Dans ce contexte, le suivi visuel des objets en mouvement dans les vidéos continue d'exercer un fort attrait pour la recherche scientifique. De multiples algorithmes ont été conçus pour réaliser ce suivi, démontrant ainsi la persistance de la pertinence de cette problématique. Ce chapitre se propose donc de présenter un état de l'art des techniques avancées pour le suivi optique des objets en mouvement au sein de séquences d'images.

3.2. Extraction des Caractéristiques et Représentation des Cibles

Diverses techniques ont été proposées pour l'extraction de repères visuels, englobant des aspects tels que la couleur, le mouvement et la texture, afin de générer des caractéristiques appliquées aux objets, également connues sous le nom de descripteurs. La littérature recense plusieurs approches pour l'extraction de ces caractéristiques, les classer et les hiérarchiser étant une tâche complexe. Ces techniques se regroupent généralement en trois catégories principales : les techniques de bas niveau, qui impliquent des transformations mathématiques fondamentales sur les images ou vidéos ; les techniques de niveau intermédiaire, basées sur des points et des zones d'intérêt ; et enfin les techniques de haut niveau, qui reposent sur la modélisation d'objets d'intérêt. Souvent, ces techniques sont assimilées aux méthodes de représentation de cibles au sein des systèmes de suivi. Dans ce contexte, Bousetouane [4] a fait une distinction entre le processus d'extraction de caractéristiques et le processus de représentation de cibles, proposant une classification illustrée dans la figure suivante.

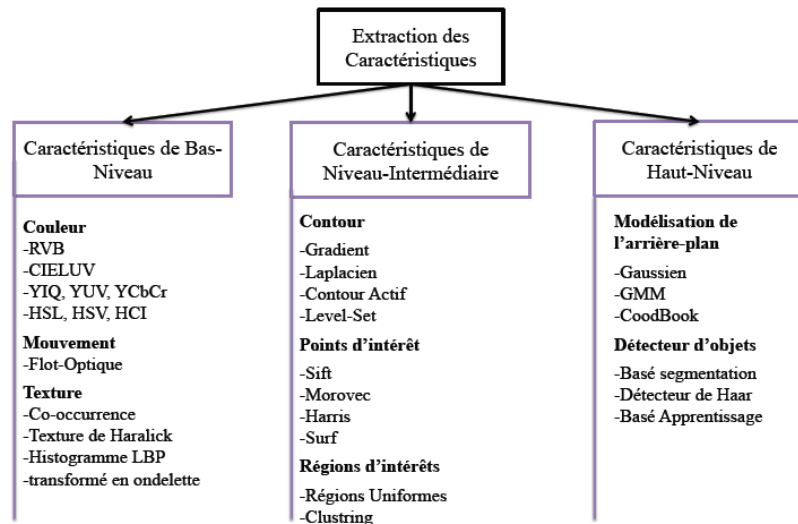


Figure 14. Taxonomie des techniques utilisées pour l'extraction des caractéristiques dans les systèmes de suivi visuel [4]

3.2.1. Caractéristiques de Bas Niveau

Les caractéristiques de bas niveau englobent des transformations mathématiques calculées au niveau du pixel et sont considérées comme parmi les plus largement utilisées dans le suivi d'objets en mouvement.

3.2.1.1. Couleur

Les caractéristiques liées à la couleur se positionnent parmi les éléments les plus fiables des systèmes de suivi. Elles permettent de décrire les aspects visuels des objets et comprennent plusieurs exemples, notamment le calibrage des couleurs [64], la représentation cible de l'algorithme de suivi [65], et le filtre à particules [66].

1. Modèle RVB

Le modèle RVB est un espace colorimétrique non linéaire englobant les composantes de couleur rouge, vert et bleu, représentées par des coordonnées cartésiennes. Un inconvénient majeur de cet espace colorimétrique réside dans sa sensibilité aux changements de couleur ainsi qu'aux mouvements rapides, ce qui peut entraîner une réduction du niveau d'exactitude lors de l'identification de cibles cruciales, notamment dans des conditions d'application réelles.

2. Modèle TSL

TSL (Teinte, Saturation, Luminosité). Cet ensemble englobe plusieurs espaces colorimétriques, notamment l'intensité *HSI*, la valeur *HSV*, la teinte et l'intensité *HCI*. Dans le contexte du suivi optique, l'espace colorimétrique *HSV* est fréquemment utilisé pour décrire les cibles. Au sein de l'algorithme de suivi, diverses techniques ont été proposées pour réaliser la conversion de l'espace *HSV* vers l'espace *RVB* [67].

3.2.1.2 Moments Statistiques

Les moments statistiques constituent des mesures appliquées aux pixels dans le but d'extraire des informations relatives au voisinage, notamment la relation entre les pixels, l'orientation centrale et l'orientation. Loza définit les moments statistiques comme des indicateurs statistiques prédictifs pour une variable aléatoire dans le traitement d'images [68]. Les auteurs démontrent que les méthodes basées sur les moments statistiques offrent des résultats plus rapides et de meilleure qualité que les méthodes graphiques pour la description et le suivi de cibles. Par exemple, ils permettent le suivi d'éléments tels que les yeux, les visages ou tout point d'intérêt en utilisant les moments statistiques pour décrire la cible.

Soit $I(x, y, t)$ une région d'intérêt dans une image centrée en c au temps t , $H(i)$ est l'histogramme de couleur calculé à partir de l'image $I(x, y, t)$, $i = 0, 1, 2, \dots, G$ est le nombre des niveaux d'intensités dans l'image, G est le maximum du nombre des niveaux d'intensités dans l'image. Soit $f(i)$ la fonction de densité de probabilité de l'occurrence des niveaux de l'intensité où $f(i) = H(i) / NM$, NM est le nombre total des niveaux d'intensités. Dans [69], les moments statistiques sont calculés à partir de l'histogramme $H(i)$ de l'image $I(x, y, t)$, comme suit.

1. Moyenne

C'est le premier moment $m1$. Elle permet de caractériser la tendance centrale de l'apparence à l'intérieur des régions d'intérêts (cibles) et fournit une bonne indication sur la nature de la distribution locale du niveau de l'intensité, est donnée par :

$$m1 = E|f(i)| = \sum_{i=0}^{G-1} if(i) \quad (41)$$

2. Variance

C'est le deuxième moment $m2$. Ce moment permet de caractériser la propagation de l'intensité à l'intérieur des régions d'intérêt (cibles). Ce moment d'ordre 1 donne une bonne indication sur le changement dans le voisinage du centre de la région d'intérêt.

$$m2 = \sigma^2 = \sum_{i=0}^{G-1} (i - m1)^2 f(i) \quad (42)$$

3. Coefficient de Dissymétrie (Skewness)

Au troisième moment ($m3$), l'asymétrie dans la distribution de la densité locale autour de la zone d'intérêt est caractérisée. Ce coefficient de dissymétrie (Skewness) est fréquemment employé dans des contextes médicaux [70].

$$m3 = \sigma^{-3} = \sum_{i=0}^{G-1} (i - m1)^3 f(i) \quad (43)$$

4. Coefficient d'aplatissement (Kurtosis)

Ce moment est combiné avec le moment d'aberration pour permettre une analyse discriminative des images. Ainsi, ce moment tend à converger vers une valeur élevée lorsque le nombre de pixels se rapproche de la valeur moyenne. La formule mathématique du kurtosis est la suivante [4] :

$$m_4 = \sigma^{-4} = \sum_{i=0}^{G-1} (i - m_1)^4 f(i) - 3 \quad (44)$$

3.2.1.3. Texture et Matrice de Cooccurrence

L'analyse de texture est l'une des techniques les plus importantes pour la description spatiale des zones d'intérêt dans une image. Elle étudie les relations entre les pixels à différents niveaux en fonction de la variation spatiale de l'intensité des pixels. Dans la littérature, des techniques telles que le Local Binary Pattern (LBP), les statistiques de premier ordre, la transformée d'ondelettes et la corrélation automatique sont utilisées pour l'analyse de la texture. Une autre approche courante est l'utilisation des matrices de cooccurrence. Ces matrices capturent la fréquence d'apparition des paires de niveaux de gris dans une certaine relation spatiale. Les paramètres de texture de Haralick sont largement utilisés dans cette approche. L'analyse de texture par ces paramètres a été appliquée dans divers domaines, tels que la classification d'images satellites, le catalogage d'images, et bien d'autres [71], [72].

3.2.1.3.1 Matrice de Cooccurrence

La matrice de cooccurrence en niveaux de gris, proposée par Haralick [73], est une matrice carrée dans laquelle chaque élément est représenté par un point M . Cette matrice examine les relations spatiales entre les valeurs de gris dans une certaine direction et une certaine distance. Elle permet d'étudier les variations d'amplitude et les motifs de distribution. La matrice de cooccurrence est mathématiquement représentée par l'équation suivante :

$$M(d_x, d_y)(i, j) = \sum_{p=1}^n \sum_{q=1}^m \begin{cases} 1 & \text{si } I(p, q) = i \wedge I(p + d_x, q + d_y) \\ 0 & \text{sinon} \end{cases} \quad (45)$$

Où : (i, j) est l'échelle de niveau de gris de la cible.

$(n; m)$ la dimension de la région d'intérêt.

I_t est l'image de la cible en niveau de gris capturée au temps t .

Cette matrice est définie sur une paire (D, O°) .

O° est l'orientation du vecteur construit par deux pixels. $O^\circ = \{0^\circ 45^\circ 90^\circ 135^\circ\}$.

D est la distance entre deux pixels.

Haralick a proposé un ensemble de paramètres qui correspondent à la description globale de la texture [73]. Parmi ces paramètres, nous citons :

1. Homogénéité

Plus on retrouve le même couple de pixels, plus la valeur de ce paramètre sera élevée. Par exemple : image uniforme ou texture périodique dans le sens de la translation.

$$HOM = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{p_{i,j}}{(1 + (i - j))} \quad (46)$$

Où : N_g est la taille du tableau, $(P_i; j)$ est le contenu de la matrice de cooccurrence, $(i; j)$ sont les indices du tableau.

2. Contraste

Une forte valeur traduit de grandes variations locales (frontières).

$$CONT = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{p_{i,j}}{(i - j)^2} \quad (47)$$

3. Entropie

Ce paramètre mesure le degré d'organisation de la texture.

$$ENT = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{ij} * \log(P_{i,j}) \quad (48)$$

4. Corrélacion

Coefficient de corrélation de la distribution du niveau de gris. Lorsque les valeurs de la matrice de cooccurrence sont uniformément distribuées ce coefficient de corrélation converge vers la plus grande valeur, dans le cas contraire le coefficient converge vers la plus petite valeur.

$$CORR = \frac{-\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - \mu) * (j - \mu) * P_j}{\sigma_x \sigma_y} \quad (49)$$

5. Dissimilarité

Mesure d'homogénéité, cette mesure converge vers une faible valeur si la texture de la région d'intérêt est homogène.

$$DISS = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{ij} * (i - j) \quad (50)$$

6. Seconde moment angulaire

Ce indice converge vers une faible valeur lorsque les $p(i; j)$ auront des valeurs très proches et il converge vers une grande valeur lorsqu'il y aura un mélange de grandes et petites valeurs.

$$ASM = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{ij}^2 \quad (51)$$

N_g est le niveau de gris de l'image I_t et P_{ij} sont les valeurs de la matrice de cooccurrence.

Les paramètres de Haralick ont été utilisés dans le domaine du suivi d'objets en mouvement [74]. Ils permettent aussi de déduire une information constante entre les images, et d'extraire la nature de la relation spatiale entre les pixels dans la région d'intérêt [65].



Figure 15. Deux matrices de cooccurrence calculées à partir des régions d'intérêts (objets en mouvement) après la génération des cartes de soustraction binarisées et étiquetées [4]

3.2.2. Caractéristiques de Niveau Intermédiaire

Les systèmes de suivi rencontrent des limitations dans l'extraction des caractéristiques de bas niveau. Pour résoudre ce problème, des sous-ensembles de pixels sont analysés. Dans cette section, nous explorons en détail les caractéristiques de niveau intermédiaire.

3.2.2.1. Les contours

Les détecteurs de contours, qui produisent des images contenant les contours des points d'intérêt de l'image sont utilisés. Cette opération consiste à calculer le seuil du gradient ; la carte des contours est calculée de la manière suivante :

$$E(x, y) = \begin{cases} 1 & \text{si } |\nabla I(x, y)| > T \\ 0 & \text{Sinon} \end{cases} \quad (52)$$

T est un seuil prédéterminé et $I(x; y)$ est l'image.

Les contours sont largement utilisés dans les systèmes de suivi, notamment pour coder une cible humaine entre les caméras, comme illustré dans [75]. Cependant, la représentation des cibles humaines par des contours (descripteurs de formes) pour la reconnaissance reste insuffisante, vu que les êtres humains ont la même forme.

3.2.2.2. Les points d'intérêts

Ce type d'extraction de caractéristiques est répandu dans les systèmes de suivi. Parmi les détecteurs de points d'intérêt les plus utilisés par la communauté scientifique, on trouve le détecteur d'angle Moravian [76], le détecteur d'angle Harris [77] et SIFT [78]. Ces détecteurs sont relativement coûteux en termes de calculs et sont souvent utilisés dans des applications en temps réel.

3.2.3. Caractéristiques de Haut Niveau

Ce type d'extraction de caractéristiques regroupe un ensemble de techniques appliquées à une image pour identifier des objets d'intérêt. Cet ensemble de techniques peut inclure des applications de niveau intermédiaire, telles que des points, des régions d'intérêt ou des détecteurs. Cela peut se faire par modélisation de fond dans certains cas, ou par apprentissage. Des méthodes d'extraction d'objets d'une image ou de séquences d'images de haut niveau ont été proposées [4].

3.2.3.1. Modélisation d'objet

Différents systèmes de détection d'objets en mouvement ont été proposés [79], basés sur l'apprentissage et la mise à jour de l'algorithme d'apprentissage dans le temps. Ces systèmes ont connu une large adoption.

La modélisation de l'arrière-plan est l'une des techniques les plus populaires et les plus utilisées par les chercheurs. Cette technique permet d'extraire les objets en mouvement dans l'image [80], [81].

3.2.4 Représentation des Cibles

La détermination des objets d'intérêt dépend de la représentation des modèles qui recueillent les caractéristiques des objets en mouvement ou des objets d'intérêt. Ces propriétés sont traduites dans l'apparence du mouvement ou dans les signaux temporels et spatiaux. La qualité de ces modèles repose sur leur capacité à être robustes face aux variations d'éclairage, au bruit et à d'autres facteurs environnementaux. Plusieurs études ont abordé ce problème en proposant différents modèles de représentation de cible. Parmi eux, on trouve des modèles basés sur des histogrammes de couleurs simples, des modèles

d'articulation déformables, etc. La figure ci-dessous illustre un schéma qui regroupe la plupart des modèles de représentation proposés.

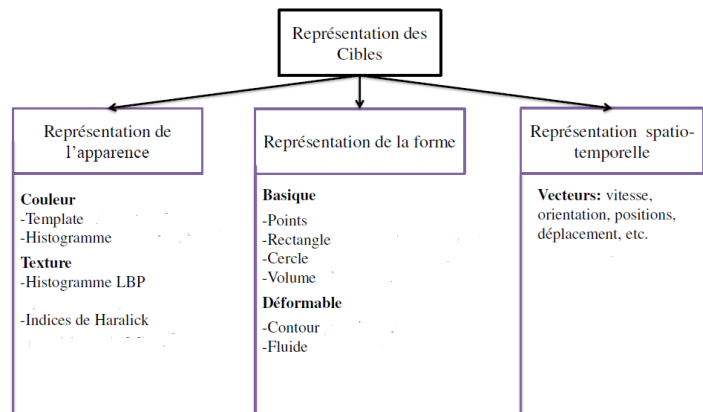


Figure 16. Taxonomie possible des techniques utilisées jusqu'à présent pour la représentation des cibles [4]

3.2.4.1 Forme

En ce qui concerne la forme des objets en mouvement ou des éléments d'intérêt, plusieurs méthodes sont employées pour représenter la forme. Ces méthodes peuvent être divisées en trois catégories : la représentation basique, la représentation articulée et la représentation déformée [4] .

1. Modèles Basiques

Les modèles basiques se fondent sur des formes géométriques simples telles qu'un rectangle, un carré, un cercle ou un point. La forme rectangulaire est la plus couramment utilisée dans le suivi visuel, car elle est simple et pratique.

2. Modèles Articulatoires

Les modèles articulés sélectionnent ou approximent la forme des objets en mouvement en se basant sur des points, des surfaces ou des dimensions. La plupart de ces modèles sont employés dans l'analyse des comportements humains.

3 .Modèles Déformables

La rigidité des formes ou les hypothèses cinématique autour d'objet d'intérêt (et qui sont largement utilisés dans les modèles articulatoires) ne peuvent être appliquées pour toutes les classes d'objets comme les objets à rigidité complexe. Pour cette raison des modèles de contours et fluides sont utilisés. Ces modèles se basent essentiellement sur des détecteurs de contours dans les images.

3.2.4.2. Apparence

La représentation des cibles basée sur l'apparence est définie comme un modèle de projection de la signature de l'apparence d'objet d'intérêt sur le plan de

l'image. Contrairement au modèle de représentation de forme, les modèles d'apparence d'objets d'intérêts doivent être spécifiques à chaque objet d'intérêt (invariants par rapport aux conditions d'éclairage et discriminants inter-objet). Dans le suivi visuel les deux modèles d'apparence les plus utilisés se basent sur les caractéristiques de la couleur et qui sont : l'histogramme et les patrons (Histogram, Template).

3.2.4.2.1 Histogramme

L'histogramme de la couleur encode la distribution statistique des valeurs des pixels dans une région d'intérêt. L'histogramme a été prouvé dans [82] comme étant un modèle de représentation des cibles invariant à la rotation et robuste à l'occultation partielle. Soit $R(c; t)$ une région d'intérêt centrée au point c à l'instant t d'une image $I(x, y, t)$, où l'histogramme de couleur $H_{rgb}(c_t)$ est calculé, la densité de couleur $q_u = u = 1 :: n$ qui estime la distribution de la couleur au temps t est approximée par :

$$q_u = w \sum_{d \in H_{rgb}} K_{xt}(x_d) \delta[b_t(x_d) - \mu] \quad (53)$$

δ est la fonction delta de Kronecker, x_d est la position de chaque pixel dans $H_{rgb}(c_t)$, w est une constante, K_{xt} est un noyau d'Epanechnikov [4][82].

3.3. Suivi d'Objets Mobiles

Le suivi des trajectoires des objets en mouvement représente une étape cruciale dans le domaine de la vidéosurveillance intelligente. Un grand nombre d'études et de méthodes ont été développées, offrant ainsi une diversification des approches. Les travaux de recherche suivants ont examiné ces différentes méthodes, qui peuvent être regroupées en deux catégories : les méthodes déterministes et les méthodes aléatoires [9][22].

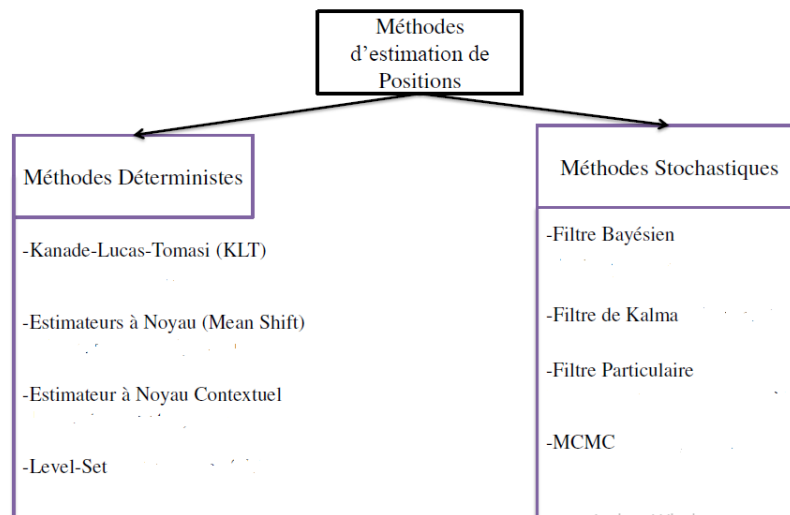


Figure 17. Taxonomie possible des techniques proposées pour le suivi et l'estimation des positions d'objets mobiles

3.3.1. Méthodes Déterministes d'Estimation des Positions

Les méthodes déterministes d'estimation des positions se focalisent sur les caractéristiques extraites de la séquence d'images pour identifier les positions cibles optimales dans le temps. Parmi les algorithmes les plus performants, nous pouvons citer le suivi de points caractéristiques (KLT), l'estimateur à noyau (Mean-Shift) et la méthode de Level-Set.

3.3.1.1. Kanade-Lucas-Tomasi

Cette méthode (KLT) [83], se base essentiellement sur un Template (patron) comme un modèle de présentation de l'apparence d'une cible pour l'estimation de ses positions au fil du temps : KLT utilise le concept de régression linéaire multiple, calculée par le flot optique. La cible est représentée par une zone carrée de taille $N = (2w - 1) \times (2w - 1)$, où w est la largeur. Le problème de suivi peut être défini comme l'estimation du mouvement lors de la transition d'un objet d'un état à un autre, où l'état est défini comme : $x_t = (u_t, v_t)$ avec (u_t, v_t) représentant le vecteur de déplacement. (u_t, v_t) est le vecteur de déplacement. Supposant que les coordonnées du modèle $I_t(\cdot)$ dans le plan image soient alignées avec les coordonnées de l'image I_t , quel que soit t . Donnant l'état initial de la cible $x(0)_t$ au temps t avec le déplacement estimé x_{t-1} au temps $t-1$, l'état de la cible x_t au temps t peut être défini comme suit :

$$X_t = x(0)_t + \Delta x_t \quad (54)$$

Δx_t est le petit déplacement de la cible ajouté au déplacement précédent, en outre imposant la contrainte de l'éclairage constante [83], la différence de l'apparence entre le Template (patron) et la fenêtre centrée en x_k est donnée comme suit :

$$I_K(w) = I_t(w - x_k) + n_k(w) = I_t(w - x(0)_t - \Delta x_t) + n_k(w) \quad (55)$$

$n_k(w)$ est le motif de bruit supplémentaire dans les valeurs de pixel.

w est la position du pixel dans l'image.

x est un petit décalage qui minimise l'erreur entre la région de la meilleure estimation x_k et le modèle.

$$\varepsilon(\Delta x_t) = \sum_{(w-x_t) < w} [(I_t(w - x_t^\circ) - \Delta x_t) - I_t(w)]^2 \quad (56)$$

le modèle $I_t(w - x_t^\circ)$ peut être approché avec une série de Taylor centrée autour de x_{t-1} pour un petit décalage de la cible δx_t .

$$(I_t(w - x_k^\circ) - \Delta x_t) \approx I_t(w - x_t^\circ) + b' \Delta x_t \quad (57)$$

b' est le gradient du gabarit, donné par :

$$b = \frac{\partial I_T(w - x_k^\circ)}{\partial w} \quad (58)$$

$$b = \frac{\partial I_T(w - x_k^\circ)}{\partial w} \quad (59)$$

3.3.1.2. Estimateur à Noyau "MeanShift"

L'estimateur à noyau "MeanShift" est une méthode qui utilise un noyau qui se déplace dans la séquence d'images. Le noyau d'intérêt peut prendre différentes formes géométriques telles que rectangulaires, ovales ou articulées. Il a été utilisé dans l'algorithme de suivi de transformation moyenne [84]. Cette méthode se déroule en deux étapes importantes : la description des cibles à l'aide d'un modèle de couleur non paramétrique (histogramme) et la convergence de la transformation moyenne. La recherche d'une nouvelle position mobile est itérative dans le temps t par rapport au temps $t-1$. L'algorithme de suivi "MeanShift", présenté dans [65], est composé de six étapes :

1. Déterminer la position initiale de la cible et initialiser l'algorithme.
2. Calculer l'histogramme (distribution des couleurs) dans la région d'intérêt en utilisant l'équation suivante :

$$q_u = w \sum_{d \in H_{rgb}} K_{xt}(x_d) \delta[b_t(x_d) - \mu] \quad (60)$$

δ est la fonction de Kronecker delta,

x_d est la position de chaque pixel dans $H_{rgb}(ct)$,

w est une constante, K est un noyau d'Epanechnikov [85].

3- Calculer la nouvelle position temporaire de la cible dans l'image courante

a. Le poids est calculé pour chaque pixel par l'équation suivante :

$$w_i = \sum_{u=1}^M \frac{q_x}{\sqrt{p_x(y_0)}} g[b(x_i) - u] \quad (61)$$

$$g(x) = -k \cdot (x) \quad (62)$$

M est le nombre des niveaux d'intensité, q_u est l'histogramme de couleur de l'objet candidat, p_u est l'histogramme de couleur de l'objet courant. $g(x)$ est une fonction de noyau. $b(x)$ est l'occurrence d'un niveau d'intensité de la distribution de couleur.

b. La nouvelle position temporaire de la cible dans l'image actuelle est calculée par l'équation suivante :

$$y_1 = \frac{\sum_{i=1}^N x_i w_i g\left(\left\|\frac{y_0 - w_i}{h}\right\|^2\right)}{\sum_{i=1}^N w_i g\left(\left\|\frac{y_0 - w_i}{h}\right\|^2\right)} \quad (63)$$

Où x_i est la position temporaire de la cible. h est la taille de la fenêtre.

c. Dans la nouvelle position temporaire, l'histogramme de couleur est calculé à l'aide de l'équation (59).

4. L'histogramme de chromaticité candidat est comparé à l'histogramme de couleur de référence, et leurs similarités sont extraites à l'aide de la distance de Bhattacharya de la manière suivante :

$$d(y) = \sqrt{1 - \rho[p(y), q]} \quad (64)$$

$$\rho[p(y), q] = \sum_{u=1}^N \sqrt{p_u(y), q_u} \quad (65)$$

$q_u, u = 1, \dots, n$: est l'histogramme de couleur calculé au temps $t - 1$ dans une région d'intérêt centrée au point x .

$p_u; u = 1, \dots, n$: est l'histogramme de couleur calculé au temps t dans une région d'intérêt centrée au point y .

La similarité entre ces deux histogrammes calculée par la distance de Bhattacharya est donnée par les équations 64,65.

La série de Taylor est employée pour réaliser une approximation de cette distance autour l'histogramme de couleur candidat , selon l'équation suivante :

$$\rho[p(y), q] \approx \frac{1}{2} \sum_{u=1}^N \sqrt{p_u(y_0), q_u} + \frac{1}{2} \sum_{u=1}^N p_u(y) \sqrt{\frac{q_u}{p_u(y_0)}} \quad (66)$$

5. La procédure de convergence du Mean-Shift : Maximisation du Coefficient de Bhattacharya (Minimisation de la distance de Bhattacharya)

6. Retourner à l'étape 3 si la maximisation du coefficient de Bhattacharya est vérifiée, sinon la cible est dans un état "non-mobile".

Malgré la large utilisation de l'algorithme de suivi (estimateur de densité à noyau, Mean-Shift) pour le suivi des cibles, cet algorithme reste insuffisant dans des conditions réelles d'application, comme : la variation de la couleur, occultation complexe, déformation locale, etc. La cause essentielle de ces insuffisances reste dans la description des cibles d'intérêt par des pixels isolés comme l'histogramme de la couleur. Pour améliorer cet algorithme, plusieurs chercheurs [64][65], [86], [87] ont montré que la combinaison de plusieurs indices visuels peut améliorer la convergence de cet algorithme dans des conditions complexes, mais le choix des indices et la manière de les combiner restent des problèmes au cœur de la recherche.



Figure 18. Résultats de suivi d'objet mobile utilisant l'estimateur à noyau contextuel Mean-shift [4]

3.3.2. Méthodes Stochastiques

Les méthodes stochastiques ou de filtrage d'état récursif se démarquent par leur rapidité et leur précision dans la caractérisation des objectifs, grâce au transfert de l'estimation de la position actuelle de la cible à l'instant t vers l'estimation de la probabilité future des positions de la cible x à l'instant t1. Ces méthodes reposent sur l'inférence bayésienne.

3.3.2.1. Filtre Bayésien Récursif

Le filtre bayésien récursif [4] [88], est un algorithme de filtrage d'état pour l'estimation de la dynamique des systèmes. Il permet d'adresser le problème de l'incertitude à travers la modélisation de l'espace d'état x_t et l'observation z_t

comme étant deux processus stochastiques. Dans un filtre Bayésien, deux modèles sont essentiels : (1) Le modèle de la dynamique du système où l'évolution d'états (équation 67) ; (2) L'observation courante de l'objet extraite à partir des séquences d'image et qui pourrait être des indices visuels comme : l'histogramme de la couleur où la forme de l'objet d'intérêt (équation 68). Sous l'hypothèse markovienne : étant donnée l'observation $z_1 : t - 1$, l'état courant x_t du système dépend que de l'état précédent x_{t-1} et l'observation courante z_t ne dépend que de l'état courant x_t . La récursivité est entièrement déterminée par l'équation d'observation h_t et la dynamique de l'état x_t définie dans l'équation f_t [4].

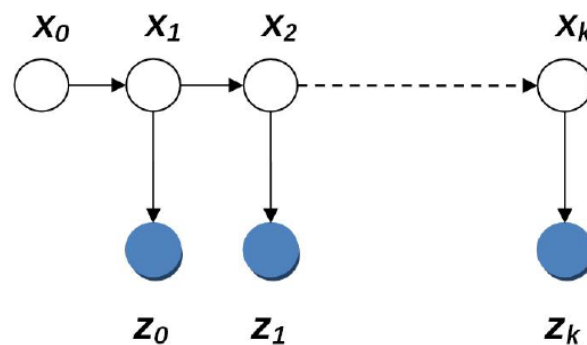


Figure 19 . Schéma graphique de la dépendance en suivi visuel sous l'hypothèse Markovienne. Chaque cercle représente un vecteur aléatoire ; Les flèches montrent la dépendance entre les vecteurs aléatoires (vecteur d'états x_t et le vecteur d'observation z_t) ; L'observation courante z_t dépend uniquement de l'état courant. L'état courant x_t dépend uniquement de l'état précédent x_{t-1} [4]

$$x_t = f_t(x_{t-1}, v_{t-1}) \quad (67)$$

$$z_t = h_t(x_t, \varphi_t) \quad (68)$$

Où : φ_t et v_{t-1} sont des bruits. Dans le suivi visuel, le but principal du filtre bayésien est l'estimation de la fonction de distribution de probabilité "PDF" $P(x_t|z_t)$, que l'objet situant dans l'état x_t .connaissant toutes les observations précédentes. L'estimation de la "PDF" est essentiellement basée sur deux étapes : La prédiction et la correction (mise à jour) [4][88].

1- Prédiction

Dans la première étape, la dynamique des objets est modélisée à l'aide d'un modèle de prédiction d'état, tel qu'un modèle autorégressif basé sur des séries temporelles [89] ou le modèle Chapman [90]

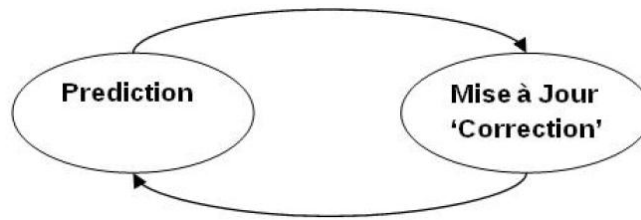


Figure 20. Étapes cycliques pour l'estimation de la distribution de probabilité a postérieure utilisant le filtre de Kalman [4]

2- Correction (Mise à jour)

Après le processus de prédiction d'état temporaire au temps t , l'observation z_t est disponible. La dernière probabilité est estimée en utilisant la loi bayésienne [89]. $P(z_t|x_t)$ représente le modèle de probabilité et est également une mesure de similarité entre deux observations consécutives. L'inférence bayésienne est scientifiquement complexe, et d'autres algorithmes tels que le filtre de Kalman et le filtre à particules ont été proposés pour faciliter l'estimation ultérieure des probabilités.

$$P(x_t|z_t) = \frac{P(z_t|x_t)P(x_t|z_{t-1})}{\int P(z_t|x_t)P(x_t|z_{t-1})dx_t} \quad (69)$$

3.3.2.2. Filtre de Kalman

Le filtre de Kalman [91] est un algorithme récursif linéaire qui permet d'estimer la probabilité a postérieure $P(x_t|z_t)$ en se basant sur les deux hypothèses suivantes : (1) la linéarité de l'équation d'évolution (équation 67). (2) la gaussianité de la distribution préalable et les deux processus de bruits. Le filtre de Kalman se base sur les deux étapes du filtre bayésien récursif : la prédiction et la mise à jour. Pour chaque nouvelle position à estimer, le filtre de Kalman se déroule en deux étapes cycliques (Figure 21) :

Ce processus consiste à prédire la nouvelle position temporaire de l'objet d'intérêt (Phase de prédiction), et l'ajustement ou la correction de cette position (Phase de Correction). L'algorithme de Kalman est très utilisé dans le suivi des cibles dans des vidéos tel que : le suivi des personnes [92], des voitures, le suivi distribué [93], etc. Malgré la popularité du filtre de Kalman, cet algorithme reste insuffisant dans des conditions réelles d'application, comme la présence de la multi-modalité et le comportement non-linéaire des cibles. Récemment, plusieurs efforts ont été mis pour l'amélioration de cet algorithme, citons a titre d'exemple : l'utilisation de la mixture gaussienne pour la modélisation de la distribution d'observation afin de contrôler le problème de la multi-modalité, mais cette solution reste insuffisante surtout pour des applications intégrant des aspects temps réels [4].

3.3.2.3. Filtre Particulaire

Contrairement au filtre de Kalman où les équations (équation.67 et équation.68) sont supposées linéaires et gaussiennes, le filtre particulaire [94] est utilisé pour des systèmes non-linéaire avec des modèles d'observation non-gaussiens. Ce filtre se base principalement sur l'algorithme de simulation de Monte-Carlo séquentielle [88]. Le filtre particulaire est une forme d'approximation numérique de la probabilité a-posteriori utilisant un ensemble de variables aléatoires. Ainsi, il est bien adapté pour suivre une trajectoire avec des changements brusques de direction, et pour le suivi multi-cible. La rapidité et l'extensibilité de cet algorithme ont récemment attiré l'attention de plusieurs chercheurs. Ce filtre a été adapté à plusieurs sous domaines du suivi visuel, comme : le suivi des visages, des mains, des personnes, suivi distribué multi objet, etc.

La densité de probabilité a posteriori $P(x_t|z_t)$ dans le filtre particulaire est approximée par une somme de N_t fonctions de (delta Dirac), chaque particule est modélisée par une variable aléatoire composée d'un poids \hat{w}_t^i et un centre $x_t : \{x_{1:k}^i, \hat{w}_k^i\}_{i=1}^N$. La probabilité à posteriori [94]. est approximée par :

$$P(x_t|z_t) \approx \sum_{i=1}^N \hat{w}_t^i \delta(x_t - x_t^i) \quad (70)$$

Le poids des particules est présenté comme suit :

$$\hat{w}_t^i = \frac{P(z|x_t^i)P(x_t^i|x_{t-1}^i)}{q(x_t^i|x_{t-1}^i, z_t)} \quad (71)$$

$P(z|x_t^i)$, est le modèle de vraisemblance [94] . Il se base sur une mesure de similarité entre deux espaces de représentation homogène, calculée dans une région d'intérêt en deux temps consécutifs. En pratique, dans le cadre du suivi visuel, la couleur est l'indice visuel le plus utilisé pour le calcul de la vraisemblance entre deux régions d'intérêt. Basant sur la distance de Bhattacharya (équation 64), la vraisemblance est estimée par l'équation suivante :

$$L_{H_{RGB}}((z_{H_{RGB},t}|x_t)) \propto e^{-\frac{d_c^2(H_{t-1}, H_t)}{2\sigma^2}} \quad (72)$$

L'estimation des trajectoires cibles dans les séquences d'images se décompose en 3 étapes principales : (1) prédiction, (2) mesure et (3) ré-échantillonnage [4], [94] :

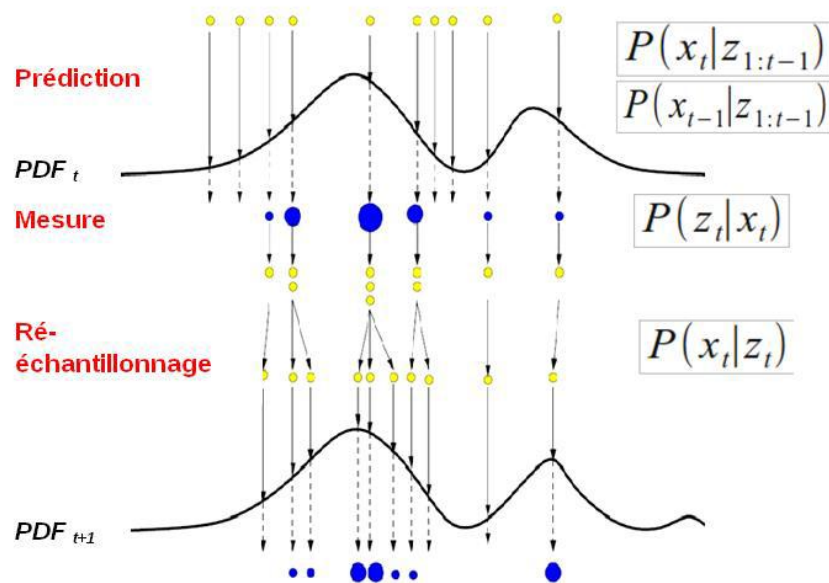


Figure 21. Représentation graphique des 3 étapes d'estimation numérique de la probabilité a postérieure utilisant l'échantillonnage de Monte-Carlo et l'équation récursive bayésienne (Filtre Particulaire) [4]

3.4. Conclusion

Dans ce chapitre, nous avons exposé les techniques actuellement employées pour le suivi d'objets en mouvement. Chaque technique a été examinée en détail, mettant en évidence ses caractéristiques spécifiques ainsi que ses avantages et ses limitations. Dans le prochain chapitre, nous explorerons un domaine très important de l'analyse vidéo qui vient compléter le processus de la détection et le suivi d'objets en mouvement: la reconnaissance de comportement.

CHAPITRE 4

RECONNAISSANCE DE COMPORTEMENTS

4.1. Introduction

La reconnaissance des comportements constitue un domaine de recherche en constante évolution, où la détection d'événements anormaux suscite un vif intérêt dans le domaine de la vision par ordinateur. Elle est reconnue comme une technologie clé de la surveillance intelligente. Les événements anormaux sont identifiés en temps réel à partir de séquences vidéo, caractérisant un nombre restreint d'occurrences qui se démarquent des événements normaux. Toutefois, la diversité des scénarios d'événements anormaux présente un défi, car il est complexe de définir une interface couvrant la variété des événements anormaux possibles. Une solution couramment adoptée consiste à considérer un événement anormal comme étant de faible probabilité par rapport aux événements normaux, permettant ainsi une approche statistique dans la détection d'événements anormaux [95], [96]. Dans cette section, nous explorerons deux catégories de méthodes de reconnaissance : l'identification statique, basée sur des critères obtenus à partir d'une seule image, et l'identification dynamique, basée sur l'analyse temporelle des séquences d'images [96].

4.2. Reconnaissance Statique

Les méthodes pour la reconnaissance statique sont basées sur des comparaisons entre l'image courante et des informations préalablement enregistrées. Une des applications type est la reconnaissance de posture où l'on cherche à reconnaître la posture d'une personne à partir d'éléments connus ou d'informations extraites des images précédentes.

4.2.1 Approche Région

Cette méthode identifie les situations suspectes en utilisant un descripteur de région. Cette technique offre des informations pertinentes sur la posture, la taille et la forme d'une personne. Dans cette optique, les auteurs ont employé la vitesse et la forme de la silhouette d'une personne [97]. De même, d'autres travaux ont utilisé un filtre à particules pour approximer la silhouette d'une personne avec une ellipse [98].

4.2.2 Approche Contour

Cette approche se distingue par son analyse précise, qui consiste à extraire la silhouette d'une personne sous forme de schéma, puis à identifier son statut ou son identité en fournissant une description détaillée de ce schéma.

Mokhtarian a utilisé le descripteur CSS (Curvature Scale Space) pour décrire les contours du corps [99]. Il a comparé les silhouettes en alignant les courbes CSS en fonction de la courbure maximale.

Une autre approche a utilisé la transformée de Fourier discrète pour représenter la silhouette humaine à l'aide des descripteurs de Fourier [100]. Plus il y a de descripteurs, plus d'informations détaillées sur la silhouette sont disponibles. Pour comparer les modèles, les descripteurs doivent être ordonnés de manière croissante. Dans ce contexte, les auteurs ont créé 40 descripteurs standards pour décrire la silhouette humaine.

Veeraraghavan et al. [101] se sont basés sur la théorie mathématique des formes de Kendall pour analyser les déformations de forme et identifier les démarches. Ils ont utilisé ces données pour décrire la silhouette d'une personne.

Bauckhage et al. [102] ont utilisé une grille 2D pour extraire des points caractéristiques de la silhouette. Cette grille permet de déterminer si une personne a une démarche normale ou anormale à l'aide d'une SVM10.

4.3. Reconnaissance Dynamique

L'identification dynamique est une étape cruciale pour identifier les trajectoires des objets en mouvement et les classer en trajectoires normales ou suspectes. Plusieurs classifications ont été mentionnées dont: les méthodes de connectivité, les méthodes basées sur le changement temporel et les modèles de Markov cachés [103]–[105].

4.3.1 Approche de Recalage Temporel

Plusieurs méthodes sont utilisées pour le recalage temporel, dont la plus connue est la déformation temporelle dynamique, également connue sous le nom de DTW11. À l'origine, cette méthode a été développée pour la reconnaissance vocale, mais elle a ensuite été adaptée à d'autres domaines tels que le suivi de chemin [106], la reconnaissance de caractères [107] et la reconnaissance de gestes [108]. Cette méthode emploie la programmation dynamique (comme illustré dans la figure ci-dessous) pour trouver le meilleur alignement entre deux trajectoires. Rao et al. proposent d'utiliser la courbure spatio-temporelle d'une trajectoire 2D pour le recalage temporel en détectant les moments dynamiques [106]. Jiao et al. suggèrent quant à eux de reconstruire une trajectoire 3D à partir de ses projections 2D extraites de plusieurs caméras [109].

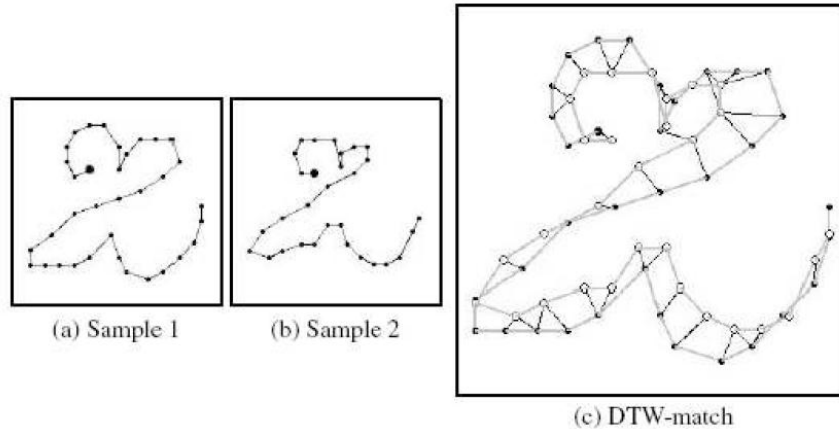


Figure 22. Exemple de recalage avec la DTW [96], [107]

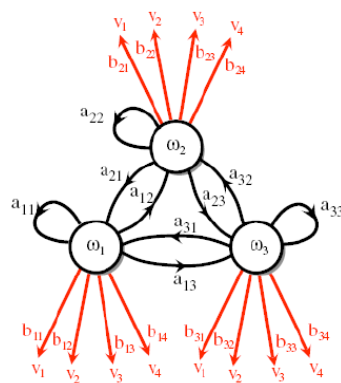
4.3.2 Modèles de Markov Cachés

Les modèles de Markov cachés (MMC, plus connus sous le nom de HMM) sont des automates probabilistes, particulièrement bien adaptés pour classifier des séquences d'actions. Ils sont par exemple utilisés pour la reconnaissance de la parole, l'analyse d'écriture ou pour la reconnaissance de gestes. Ils se basent sur l'hypothèse de Markov qui énonce que l'état futur ne dépend que de l'état présent.

4.3.2.1. Structure d'un HMM et Apprentissage

Le but d'un HMM est d'identifier une séquence d'événements étant donné un ensemble d'observations V (états visibles). Les paramètres d'un modèle HMM de 1er ordre (ne dépendant que de l'état précédent) sont montrés sur la figure 23.

Les paramètres du HMM sont déterminés par un algorithme d'apprentissage. Cet algorithme sert à entraîner le HMM à partir d'une séquence d'événements. Les algorithmes d'apprentissage les plus connus sont l'algorithme de Viterbi et l'algorithme de Baum Welch (algorithme de "retour arrière"). Chaque action à reconnaître est modélisée par un HMM, donc il faudra autant de HMMs qu'il y a d'actions à reconnaître.



- États visibles : $V^T = \{v(1), v(2), \dots, v(T)\}$
- États cachés : $\omega^T = \{\omega(1), \omega(2), \dots, \omega(T)\}$
- $a_{ij} = P(\omega_j(t+1) | \omega_i(t))$: probabilité de transition pour passer de l'état $\omega_i(t)$ à l'état $\omega_j(t+1)$
- $b_{jk} = P(v_k(t) | \omega_j(t))$: probabilité d'émission c'est à dire probabilité d'observer un état $v_k(t)$ étant donné un état $\omega_j(t)$

Figure 23. Exemple de structure d'un HMM [96], [103]

4.3.3 Méthodes connexionnistes

L'approche connexionniste consiste à utiliser un réseau de neurones pour classifier les données en une ou plusieurs classes. la majorité des applications ont pour but de différencier les comportements normaux des comportements anormaux. Pour cela, il est nécessaire d'avoir un nombre suffisant et bien représentatif de données d'entraînement. Il existe plusieurs types de réseaux de neurones, chacun avec ses particularités.

4.3.3.1 - Structure d'un réseau de neurones

Un neurone possède plusieurs entrées x_i , chacune étant affectée d'un poids w_i , et une sortie y . Le passage des entrées vers la sortie se fait par une fonction d'activation f de type sigmoïde, radiale ou autres (figure 24). Un réseau de neurones est composé d'un ensemble de neurones connectés entre eux par des liaisons affectées d'un poids. En modifiant les poids de certaines de ces connexions, on peut adapter le réseau pour qu'il donne en sortie des réponses différentes. La modification des poids est faite à l'aide d'un algorithme d'apprentissage

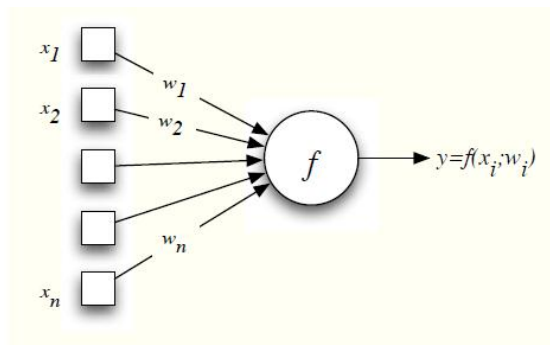


Figure 24. Structure d'un réseau de neurones [96]

4.3.3.2. Apprentissage

L'apprentissage du réseau neuronal consiste à ajuster les poids des neurones du réseau, ce qui permet au réseau de mieux s'adapter à la tâche à accomplir. Nous distinguons deux types d'apprentissage :

a- Apprentissage Supervisé

L'apprentissage supervisé implique l'adaptation d'un réseau de neurones à un ensemble de données fournies au réseau via des entrées, dans le but de produire une sortie souhaitée.

b- Apprentissage Non-supervisé

Dans ce type d'apprentissage, un réseau de neurones est construit en se basant sur des critères de similarité entre ses entrées.

Voici quelques règles de mise à jour des poids des connexions neuronales qui permettent aux réseaux de neurones de s'adapter aux tâches de reconnaissance requises.

1. Règle de Hebb

Cette règle s'appuie sur des concepts neurobiologiques. Selon cette règle, lorsque deux neurones connectés sont activés simultanément, le poids de la connexion entre eux est augmenté.

2. Règles de Correction d'Erreur

Cette méthode s'applique dans le cas de l'apprentissage supervisé. Ici, (d) représente le résultat souhaité et (y) est le résultat calculé par le réseau. L'erreur $(d - y)$ ajuste les connexions en modifiant leurs poids pour réduire l'erreur.

3. Apprentissage de Boltzmann

C'est une règle de type stochastique qui consiste à mettre à jour les neurones de façon probabiliste.

4 Règles d'Apprentissage par Compétition

L'apprentissage par compétition est du type « Tout pour le vainqueur » (winner-take-all). C'est à dire qu'il favorise le neurone gagnant en le rapprochant du vecteur d'entrée à qui il doit sa victoire. Pour chaque entrée, il y a un seul neurone gagnant.

4.3.3.3 Différents Types de Réseaux de Neurones

Il existe de nombreux types de réseaux de neurones, nous allons aborder dans cette partie les réseaux les plus populaires : les réseaux de neurones non récurrents, les réseaux de neurones récurrents .

1. Réseaux de Neurones Non-récurrents

Les réseaux de neurones multicouches sont les réseaux de neurones non récurrents les plus simples. Ils comportent des couches de neurones, notamment une couche d'entrée, une couche de sortie et une ou plusieurs couches cachées, comme illustré dans la figure 25. Dans ce type de réseau, les informations circulent de l'entrée vers la sortie dans une seule direction. Les neurones cachés de ce type de réseau utilisent généralement une fonction d'activation de type sigmoïde.

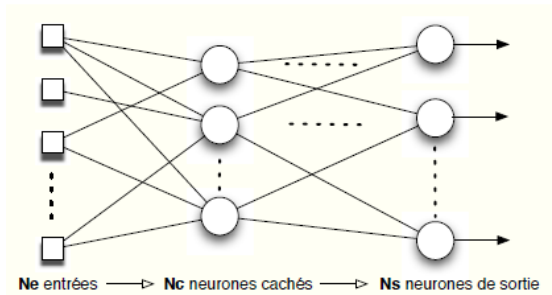


Figure 25. Réseau de neurone non récurrent [96]

Sixsmith et Johnson ont utilisé un réseau multicouche pour la détection des chutes à l'aide d'un capteur infrarouge [110]. Les auteurs ont fourni des scénarios de chutes de personnes pour entraîner le réseau de neurones sur ces scénarios. Le réseau neuronal multicouche a réussi à détecter un tiers des cas sujets aux chutes parmi ceux présentés.

2. Réseaux de Neurones Récurrents

Les réseaux récurrents sont des réseaux entièrement interconnectés dans lesquels les entrées et les sorties sont combinées. Il existe plusieurs types de réseaux qui s'appuient sur des structures récurrentes, tels que les réseaux de Hopfield, les réseaux d'apprentissage par compétition et les réseaux ART (Adaptive Resonance Theory) [96].

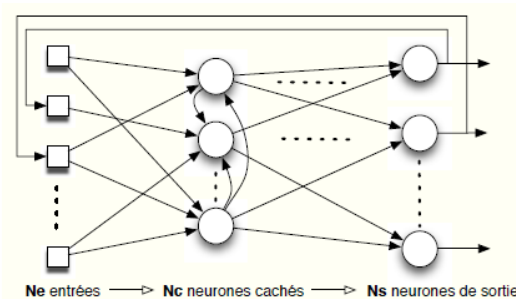


Figure 26. Réseau de neurones récurrent [96]

4.5 Conclusion

La reconnaissance des comportements et la détection des anomalies constituent un domaine de recherche en constante évolution. Ce chapitre a mis en lumière les méthodes actuellement employées pour la reconnaissance des comportements. Dans le prochain chapitre, nous allons présenter l'implémentation pratique et les résultats expérimentaux.

CHAPITRE 5

IMPLEMENTATION PRATIQUE ET RESULTATS EXPERIMENTAUX

5.1 Introduction

La vidéosurveillance intelligente repose fondamentalement sur la détection et le suivi d'objets en mouvement au sein de séquences vidéo. Dans cette thèse, nous introduisons une méthodologie d'évaluation des performances des algorithmes de soustraction de fond pour la détection d'objets en mouvement. Cette approche se base sur une évaluation approfondie des performances en termes de segmentation, de consommation de mémoire, de temps d'exécution et d'utilisation du processeur. L'objectif est de mettre en évidence les avantages et les limites de chaque méthode de soustraction de fond et de suivi. Ces résultats permettront aux chercheurs et professionnels de choisir et de concevoir les méthodes les mieux adaptées grâce à cette analyse comparative.

De plus, dans le cadre de cette étude, nous avons constitué une base de données comprenant des clips vidéo réels capturés dans diverses conditions d'éclairage. Nous mettons cette base de données à la disposition de la communauté scientifique, offrant ainsi une ressource précieuse pour les recherches futures dans ce domaine.

5.2 Base de Données

Nous avons constitué un ensemble de séquences vidéo authentiques dans un environnement sous contrôle, comprenant diverses conditions d'éclairage. Pour chaque vidéo, nous avons établi une vérité terrain en créant une image de référence, que nous avons ensuite comparée aux résultats de chaque méthode.

Ces clips vidéo authentiques ont été capturés dans un atelier de la Compagnie des Eaux de la ville de Bou Saada, mettant en scène un groupe d'ouvriers en pleine exécution de leurs tâches.

De plus, nous avons intégré des vidéos préexistantes issues de l'ensemble de données Background Modeling Challenge (BMC), qui englobe une gamme variée d'évaluations reposant sur des vidéos à la fois réelles et synthétiques. Cette fusion crée un ensemble de défis intéressants pour les chercheurs qui

souhaitent évaluer les méthodes de soustraction de fond dans le domaine de l'analyse vidéo [81].

5.2.1 Vidéo du groupe d'ouvriers (vidéo 1)

Cette vidéo a été enregistrée en plein jour pour capturer des ombres. Sa durée est de 10 secondes et elle est composée de 341 images. Nous la désignons sous le nom de « Vidéo n°1 ». Nous avons sélectionné la trame numéro 253 et extrait la vérité terrain correspondante. Ci-dessous, nous avons choisi quelques images à présenter dans la figure suivante.



Figure 27. Une collection de séquences d'images de la Vidéo n°1

5.2.2 Vidéo de Deux Employés de la Société (Vidéo 2)

Cette vidéo a été enregistrée tôt le matin afin d'éviter les ombres et met en scène le déplacement de deux ouvriers et d'une voiture. La vidéo a une durée de 32 secondes et comprend 581 images. Nous la désignons sous le nom de « Vidéo n°2 ».



Figure 28. Une collection de séquences d'images de la Vidéo n°2

5.2.3 BMC(Background Modeling Challenge)

Nous avons sélectionné la vidéo synthétique n°211 à des fins d'utilisation dans un environnement nuageux avec du bruit. Ci-dessous, nous présentons quelques captures d'écran de cette vidéo. La durée totale de la vidéo est de 59 secondes, comprenant 1498 images.

5.2.4 La vérité terrain

Nous avons extrait la vérité terrain pour une image de chaque séquence vidéo. Pour la vidéo n°1 (où les ombres sont claires), nous avons sélectionné l'image n°253. Pour la vidéo n°2 (où le manque d'ombres est évident), notre choix s'est porté sur l'image n°272. En ce qui concerne les données BMC, la vérité terrain était disponible pour toutes les images, et nous avons opté pour l'image n°597.

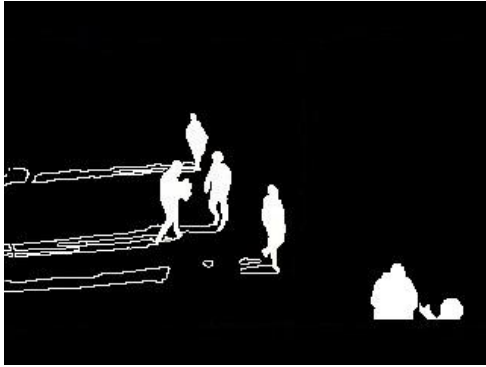


Figure 29. Vérité terrain de l'image n° 253 de la Vidéo n°1



Figure 30. Vérité terrain de l'image n° 272 de la Vidéo n°2



Figure 31. Vérité terrain pour l'image n°597 de la scène vidéo BMC

5.2.5 Bibliothèque de Soustraction d'Arrière-plan (BGSLibrary)

Au cours des dernières années, une variété de méthodes de modélisation de l'arrière-plan ont vu le jour pour la détection d'objets en premier plan dans les vidéos. Ces méthodes se distinguent par leurs avantages et leurs limitations en

termes de détection d'objets en mouvement. C'est dans ce contexte que la BGSLibrary a été créée par Andrews Sobral en 2012 [111]. Cette bibliothèque offre une collection initiale de 29 algorithmes de soustraction d'arrière-plan, programmée en langage C++. Au fil du temps, elle s'est enrichie de nombreuses autres méthodes, comptant actuellement environ 43 des algorithmes les plus couramment utilisés dans le domaine de l'analyse vidéo et de la vidéosurveillance.

Tableau 1 : Liste des algorithmes de la bibliothèque que nous avons utilisés

1= FrameDifference	17=DPEigenbackground	33=VuMeter
2=StaticFrameDifference	18=DPTexture	34=KDE
3=WeightedMoving Mean	19=T2FGMM_UM	35=IndependentMultimodal
4=WeightedMovingVariance	20=T2FGMM_UV	36=MultiCue
5=MixtureOfGaussianV1	21=T2FMRF_UM	37=SigmaDelta
6=MixtureOfGaussianV2	22=T2FMRF_UV	38=SuBSENSE
7=AdaptiveBackgroundLearning	23=FuzzySugenolIntegral	39=LOBSTER
8=AdaptiveSelectiveBackgroundLearning	24=FuzzyChoquetIntegral	40=PAWCS
9=GMG	25=MultiLayer	41=TwoPoints
10=KNN	26=PixelBasedAdaptiveSegmenter	42=ViBe
11=DPAdaptiveMedian	27=LBSimpleGaussian	43=CodeBook
12=DPGrimsonGMM	28=LBFuzzyGaussian	44=Ground truth
13=DPZivkovicAGMM	29=LBMixtureOfGaussians	45=Original frame
14=DPMean	30=LBAadaptiveSOM	
15=DPWrenGA	31=LBFuzzyAdaptiveSOM	
16=DPPPratiMediod	32=LBP_MRF	

Tableau 2 : Parametres des algorithmes de soustraction d'arrière-plan (BGS Library).

Propriétés	ID de méthode	Nom de la méthode	Paramètres	Auteurs
Basic methods, mean and variance over time	StaticFrameDifferenceBGS	Static Frame Difference	T = 15	-
	FrameDifferenceBGS	Frame Difference	T = 15	-
	WeightedMovingMeanBGS	Weighted Moving Mean	T = 10	-
	WeightedMovingVarianceBGS	Weighted Moving Variance	T = 15	-
	AdaptiveBackgroundLearning	Adaptive Background Learning	T =15, $\alpha=0.5$	-
	DPMeanBGS	Temporal Mean	T = 2700; $\alpha=10^{-7}$; LF=30	-
	DPAdaptiveMedianBGS	Adaptive Median	T = 20; LF = 30; SR =10	[112]
DPPPratiMediodBGS	Temporal Median	T =30; SR =5; HS=16; $\gamma= 5$	[113]	
Fuzzy based methods	FuzzySugenolIntegral	Fuzzy Sugeno Integral	T = 0:67; LF =10; $\alpha_{learn}= 0:5$; $\alpha_{update}= 0:05$; RGB+ LBP	[114]
	FuzzyChoquetIntegral	Fuzzy Choquet Integral	T = 0:67; LF =10; $\alpha_{learn}= 0:5$; $\alpha_{update}= 0:05$; RGB+ LBP	[115]
	LBFuzzyGaussian	Fuzzy Gaussian	T = 160; LR = 150; $\rho=100$; $\sigma =195$	[116]
Statistical methods using one Gaussian	DPWrenGABGS	Gaussian Average	T =12:15; LF =30; $\alpha=0:05$	[40]
	LBSimpleGaussian	Simple Gaussian	LR= 50; $\rho=255$; $\sigma =150$	[117]
	DPGrimsonGMMBGS	Gaussian Mixture Model	T = 9; $\alpha= 0.05$; n = 3	[118]
	MixtureOfGaussianV1BGS	Gaussian Mixture Model	T = 10; $\alpha= 0.01$	[119]

Statistical methods using multiple Gaussians	MixtureOfGaussianV2BGS	Gaussian Mixture Model	$T = 5; \alpha = 0.01$	[42]
	DPZivkovicAGMMBGS	Gaussian Mixture Model	$T = 20; \alpha = 0.01, n=3$	[41]
	LBMixtureOfGaussians	Gaussian Mixture Model	$T = 80; \alpha = 60; \rho=120; \sigma =210$	[120]
Type-2 Fuzzy based methods	T2FGMM_UM	Type-2 Fuzzy GMM-UM	$T = 1; K_m = 2.5; n =3; \alpha =0.01$	[48], [121], [122]
	T2FGMM_UV	Type-2 Fuzzy GMM-UV	$T = 1; K_v = 0.6; n =3; \alpha =0.01$	[48], [121], [122]
	T2FMRF_UM	Type-2 Fuzzy GMM-UM with MRF	$T = 1; K_m = 2.0; n =3; \alpha =0.01$	[123]
	T2FMRF_UV	Type-2 Fuzzy GMM-UV with MRF	$T = 1; K_v = 0.9; n =3; \alpha =0.01$	[123]
Statistical methods using color and texture features	MultiLayerBGS	Multi-Layer BGS	Paramètres de [124]	[124]
Non-parametric methods	PixelBasedAdaptiveSegmenter	Pixel-Based Adaptive Segmenter	Paramètres de [125]	[125]
	GMG	GMG	$T = 0.7; LF =20$	[126]
	VuMeter	VuMeter	$T = 0.03; \alpha =0.995; \text{binsize} = 8$	[127]
Methods based on eigenvalues and eigenvectors	DPEigenbackgroundBGS	Eigenbackground/ SL-PCA	$T = 255; HS = 10; ED=10$	[128]
Neural and neuro-fuzzy methods	LBAdaptiveSOM	Adaptive SOM	$LR = 180; LR_{\text{training}} = 255; \rho =100; \rho_{\text{training}} = 240; TS = 40$	[56]
	LBFuzzyAdaptiveSOM	Fuzzy Adaptive SOM	$LR = 180; LR_{\text{training}} =255; q = 100; q_{\text{training}} = 240; TS =40$	[57]

Où : T est seuil, LF (learningFrames), SR = samplingRate, HS = historySize, γ =weigh, α = (alpha or learningRate), ρ = sensitivity, σ = noiseVariance, n = gaussians, ED = embeddedDim, TS = trainingSteps.

5.3 Évaluation en Temps Réel des Performances des Algorithmes de la BGS Library

Dans cette étude, nous avons effectué une évaluation complète en temps réel de l'ensemble des algorithmes de la BGSLibrary. Pour cette évaluation, nous avons choisi de mesurer quatre métriques couramment utilisées.

Tout d'abord, nous avons analysé les performances de segmentation pour évaluer la capacité des algorithmes à extraire les objets en mouvement en utilisant un ensemble de métriques standard telles que les vrais positifs (TP), les vrais négatifs (TN), les faux positifs (FP) et les faux négatifs (FN).

Ensuite, nous avons examiné le temps d'exécution, qui varie d'un algorithme à l'autre. Certains offrent des réponses rapides tandis que d'autres demandent davantage de temps.

Troisièmement, nous avons évalué la consommation de mémoire. La plupart des algorithmes présentent une consommation mémoire constante, mais certains montrent une tendance à l'augmentation. La mémoire est utilisée pour stocker les paramètres et variables locales déclarées et utilisées par toutes les fonctions en cours d'appel. Plus il y a d'appels de fonction en cours, et plus la mémoire sera donc mobilisée.

Enfin, nous avons scruté les métriques de performance liées au processeur. Cette évaluation a été réalisée en temps réel en utilisant un processeur.

5.3.1 Évaluation des Performances de la Segmentation

Pour évaluer les performances de segmentation des algorithmes, nous avons recours à un ensemble de métriques largement utilisées dans le domaine de l'analyse et de l'évaluation vidéo. Parmi celles-ci, on retrouve la F – mesure (FM) ainsi que la précision (AC), le rappel (RC), le taux de faux négatifs (FNR), le taux de faux positifs (FPR), le taux de vrai positif (TPR), la précision (PC), dont les calculs s'appuient sur les valeurs de TP, TN, FP et FN. Ces échelles de mesure sont construites en se basant sur des algorithmes d'évaluation et de comparaison.

Plus précisément, TP (vrais positifs) représente l'ensemble des pixels identifiés correctement comme faisant partie du premier plan. FP (faux positifs) concerne les pixels d'arrière-plan incorrectement classés comme étant du premier plan. TN (True Negative) englobe les pixels correctement classés comme arrière-plan. Enfin, FN (faux négatifs) concerne les pixels de premier plan malencontreusement attribués à l'arrière-plan.

Toutefois, il est essentiel de souligner que l'implémentation de l'algorithme fournit également une représentation visuelle des images extraites.

$$Recall(RC) = \frac{TP}{TP + FN} \quad (73)$$

$$Precision(PC) = \frac{TP}{TP + FP} \quad (74)$$

$$TPR = \frac{TP}{TP + FN} \quad (75)$$

$$FPR = \frac{FP}{FP + TN} \quad (76)$$

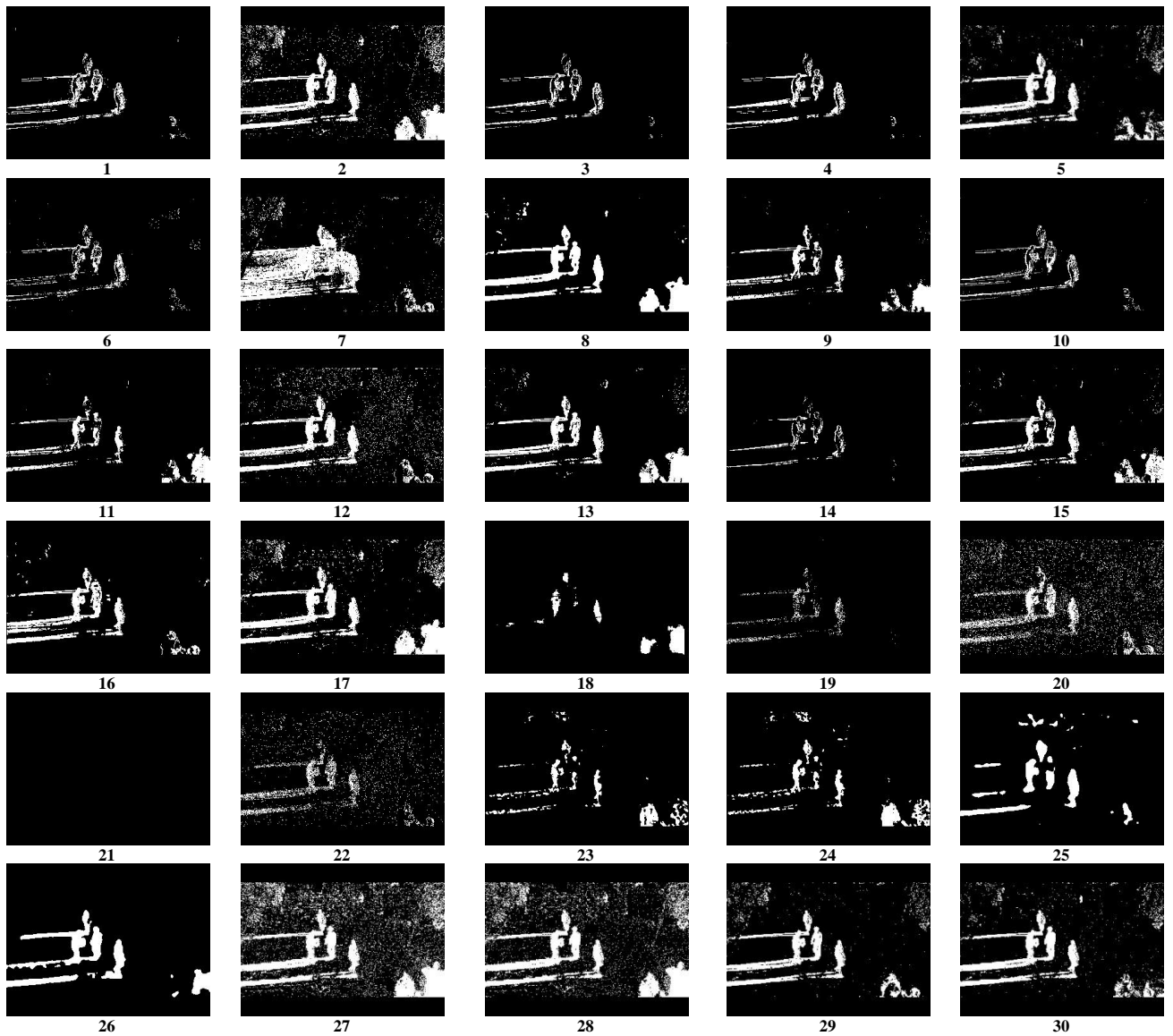
$$FNR = \frac{FN}{TP + FN} \quad (77)$$

$$F - \text{measure} = \frac{2 * Recall * Precision}{Recall + Precision} \quad (78)$$

$$Accuracy(AC) = \frac{TP}{TP + FN + FP} \quad (79)$$

L'analyse a été réalisée sur un processeur principal i5 cadencé à 2,40 GHz, avec 8 Go de RAM. La bibliothèque a été compilée en C++, basée sur OpenCV. De plus, nous avons utilisé Visual Studio 2017 pour exécuter ces algorithmes en simulation. Les comparaisons ont ensuite été effectuées à l'aide de MATLAB 2017. Les résultats, c'est-à-dire les masques, sont présentés dans les trois tableaux suivants. Les masques sont numérotés en fonction du tableau 1.

Tableau 3: Masques de premier plan obtenus à partir de la Vidéo n°1.



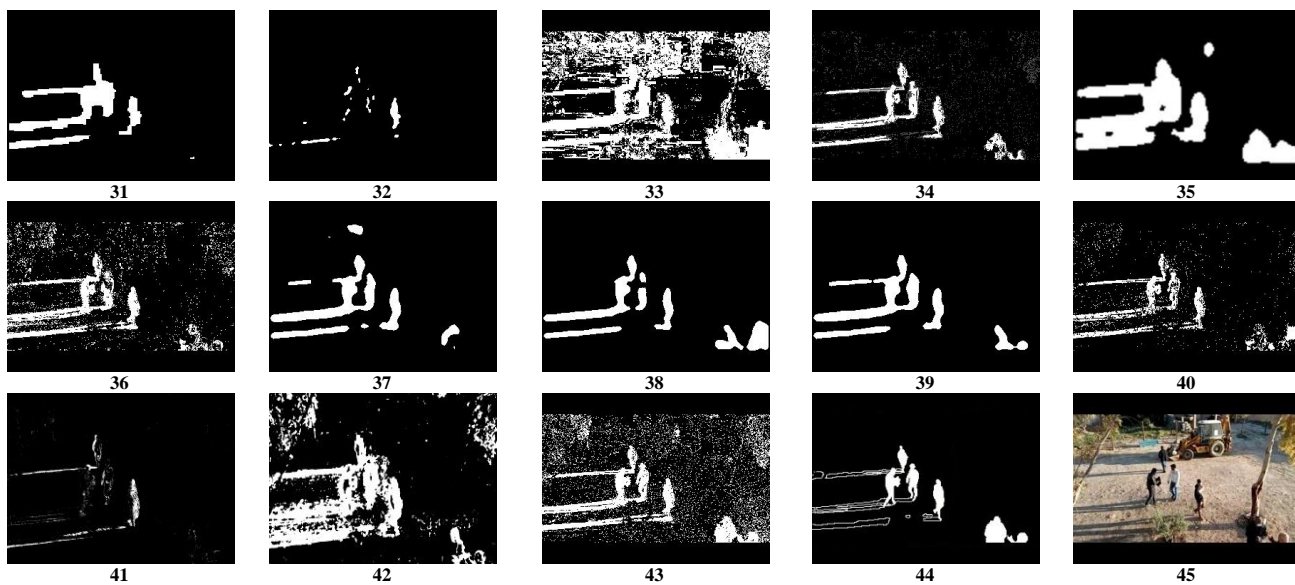
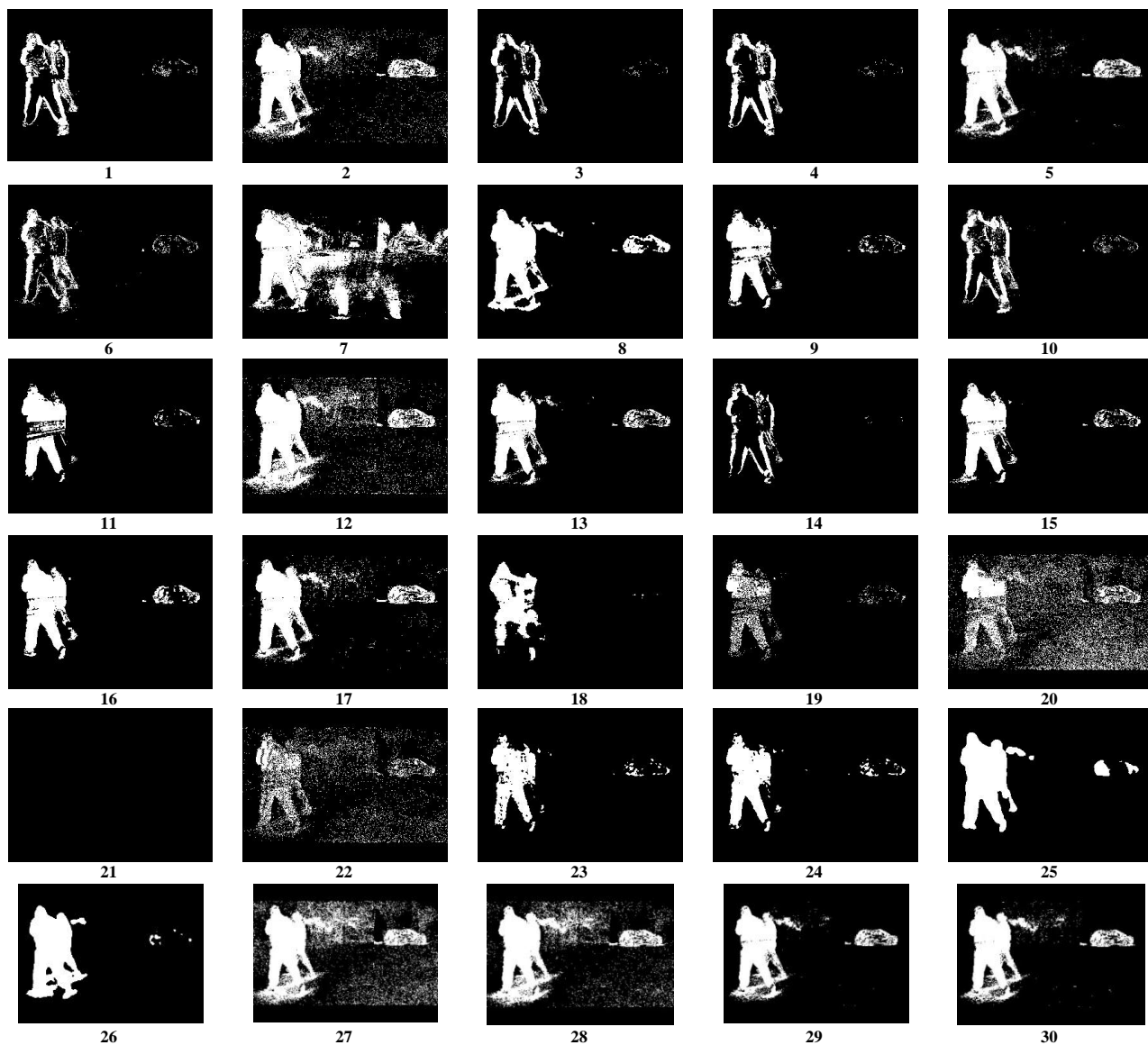


Tableau 4 : Masques de premier plan obtenus à partir de la Vidéo n°2.



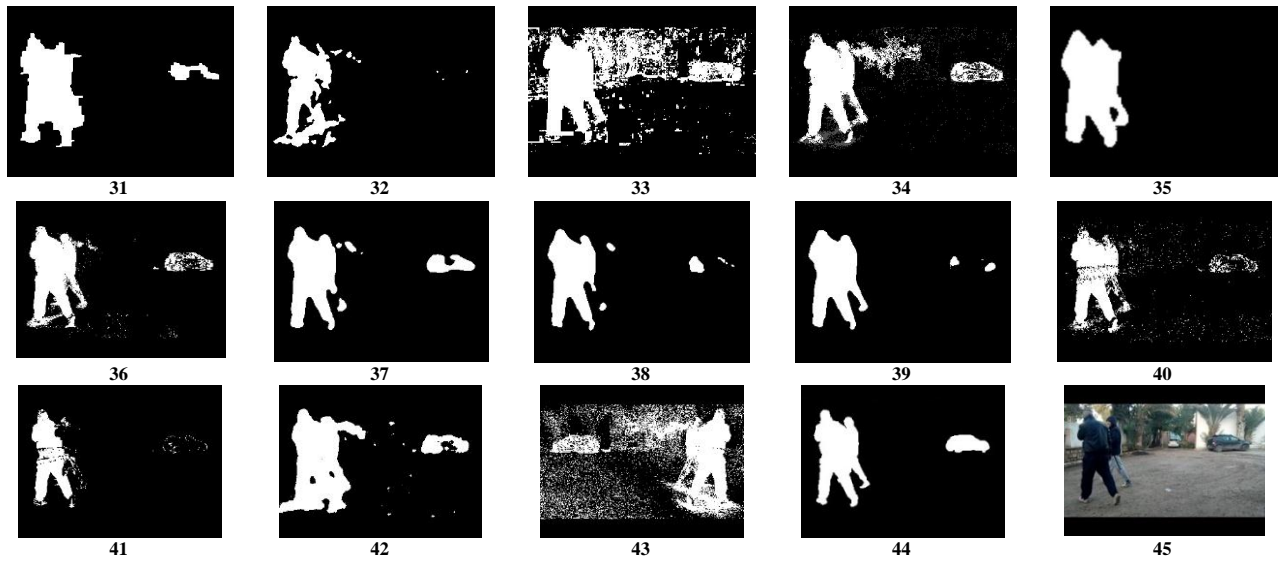
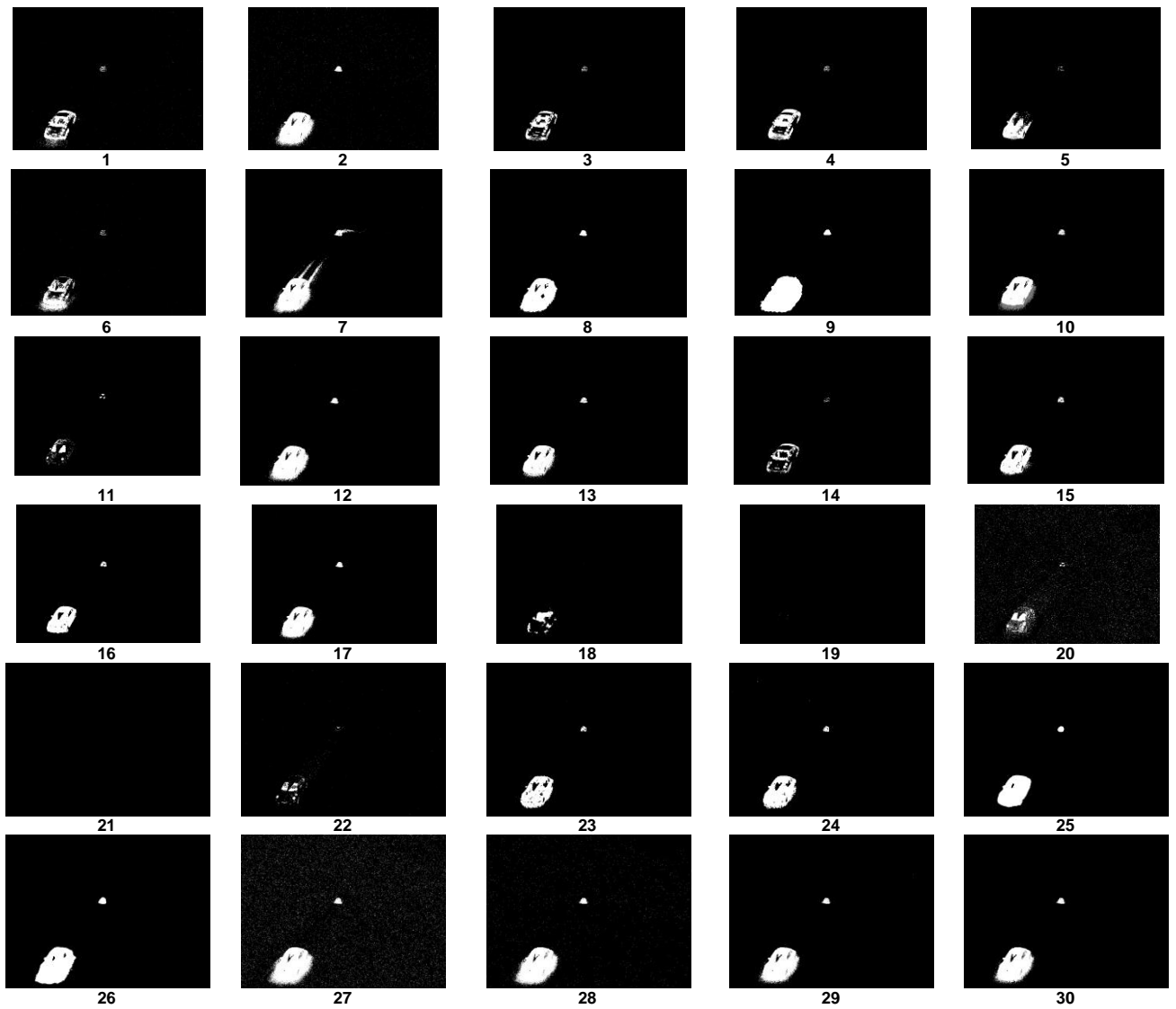
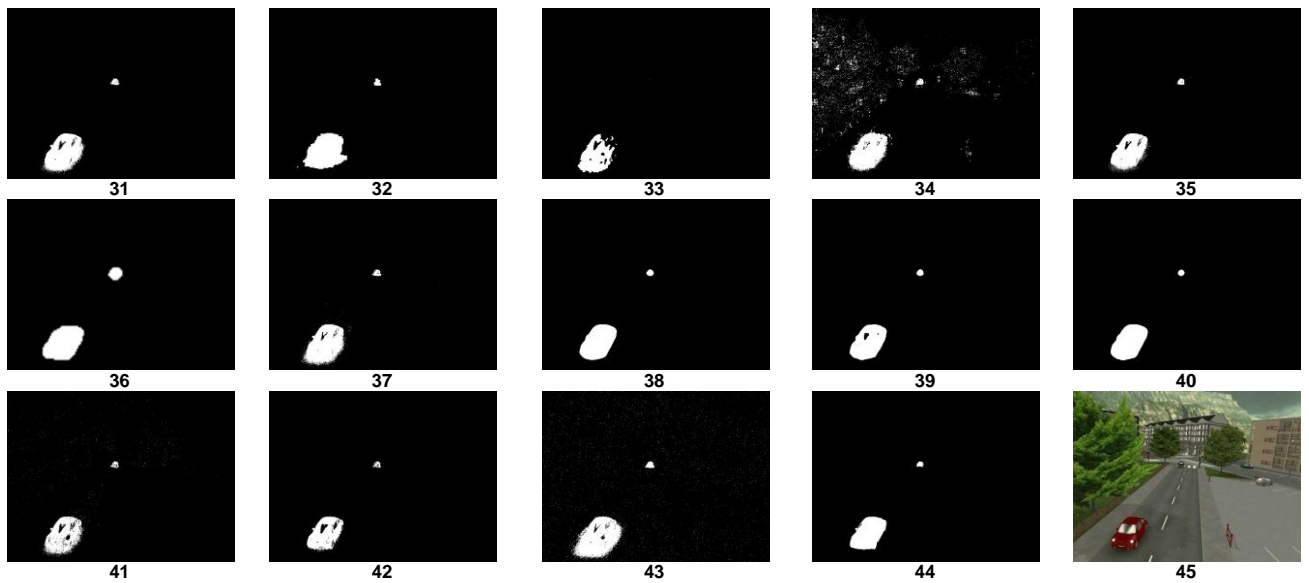


Tableau 5 : Les masques fournis ont été obtenus à partir de la vidéo BMC





D'après les tableaux précédents, les résultats montrent que l'extraction des objets en mouvement dans la vidéo synthétique est étai plus claire que dans les vidéos réelles. Une vidéo qui ne contient pas d'ombres est plus claire qu'une vidéo comportant des ombres. Ces résultats ne sont visibles qu'à l'œil nu, mais grâce à l'analyse de segmentation, ceci est obtenu en extrayant les valeurs de TP, TN, FP et FN. Cela se fait en comparant les masques de chaque algorithme avec sa vérité terrain, comme dans la figure suivante.

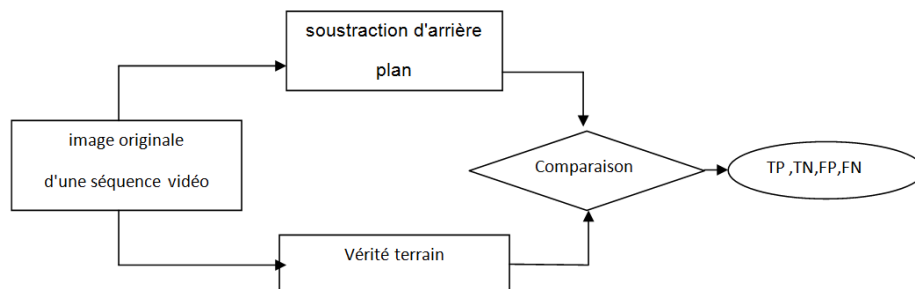


Figure 32. Calcul des métriques de classification

Les tableaux ci-dessous affichent les métriques de classification employées pour analyser les données vidéo issues de la comparaison entre les masques et la vérité terrain. Ces résultats mettent en évidence l'efficacité de l'extraction des objets en mouvement à partir de séquences vidéo.

Tableau 6 : Résultats des métriques d'évaluation utilisées pour analyser les données vidéo n°1.

Methods	RC	PC	TPR	FPR	FNR	FM	AC
FrameDifference	0.2979	0.6510	0.2979	0.0258	0.7021	0.4087	0.2568
StaticFrameDifference	0.5254	0.2870	0.5254	0.2108	0.4746	0.3713	0.2279
WeightedMovingMean	0.2528	0.6594	0.2528	0.0211	0.7472	0.3655	0.2236
WeightedMovingVariance	0.2749	0.6621	0.2749	0.0227	0.7251	0.3885	0.2411
MixtureOfGaussianV1	/////	/////	/////	/////	/////	/////	/////
MixtureOfGaussianV2	0.2848	0.6009	0.2848	0.0306	0.7152	0.3864	0.2395
AdaptiveBackgroundLearning	0.5097	0.3450	0.5097	0.1563	0.4903	0.4115	0.2590
AdaptiveSelectiveBackgroundLearning	0.4812	0.5608	0.4812	0.0609	0.5188	0.5179	0.3494
GMG	/////	/////	/////	/////	/////	/////	/////
KNN	0.3015	0.6603	0.3015	0.0251	0.6985	0.4140	0.2610
DPAadaptiveMedian	0.3534	0.5968	0.3534	0.0386	0.6466	0.4439	0.2853
DPGrimsonGMM	0.5066	0.2868	0.5066	0.2035	0.4934	0.3662	0.2242
DPZivkovicAGMM	0.4613	0.5063	0.4613	0.0727	0.5387	0.4827	0.3182
DPMean	0.1920	0.6960	0.1920	0.0136	0.8080	0.3010	0.1772
DPWrenGA	0.4255	0.5162	0.4255	0.0644	0.5745	0.4665	0.3042
DPPratiMediod	0.4221	0.5930	0.4221	0.0468	0.5779	0.4932	0.3273
DPEigenbackground	0.5283	0.3532	0.5283	0.1563	0.4717	0.4233	0.2685
DPTtexture	0.1669	0.6300	0.1669	0.0158	0.8331	0.2639	0.1520
T2FGMM_UM	0.2187	0.6809	0.2187	0.0166	0.7813	0.3310	0.1983
T2FGMM_UV	0.4917	0.2261	0.4917	0.2718	0.5083	0.3098	0.1833
T2FMRF_UM	0	NaN	0	0	1	NaN	0
T2FMRF_UV	0.3984	0.2425	0.3984	0.2010	0.6016	0.3015	0.1775
FuzzySugenoIntegral	0.3195	0.6206	0.3195	0.0315	0.6805	0.4218	0.2673
FuzzyChoquetIntegral	0.3480	0.5905	0.3480	0.0390	0.6520	0.4379	0.2803
Multi-Layer	0.3154	0.5907	0.3154	0.0353	0.6846	0.4112	0.2588
PixelBasedAdaptiveSegmenter	0.4034	0.6136	0.4034	0.0410	0.5966	0.4868	0.3217
LBSimpleGaussian	0.6080	0.2306	0.6080	0.3276	0.3920	0.3344	0.2008
LBFuzzyGaussian	0.5907	0.2408	0.5907	0.3008	0.4093	0.3422	0.2064
LBMixtureOfGaussians	0.4941	0.4105	0.4941	0.1146	0.5059	0.4484	0.2890
LBAadaptiveSOM	0.5171	0.3626	0.5171	0.1468	0.4829	0.4263	0.2709
LBFuzzyAdaptiveSOM	0.4868	0.4402	0.4868	0.1000	0.5132	0.4623	0.3007
LBP_MRF	0.3700	0.6763	0.3700	0.0286	0.6300	0.4783	0.3143
VuMeter	0.1139	0.7790	0.1139	0.0052	0.8861	0.1987	0.1103
KDE	0.6450	0.2232	0.6450	0.3626	0.3550	0.3316	0.1988
IndependentMultimodal	0.5483	0.2879	0.5483	0.2190	0.4517	0.3776	0.2327
MultiCue	0.6709	0.4848	0.6709	0.1152	0.3291	0.5629	0.3917
SigmaDelta	0.5244	0.2895	0.5244	0.2079	0.4756	0.3731	0.2293
SuBSENSE	0.4160	0.6726	0.4160	0.0327	0.5840	0.5141	0.3459
LOBSTER	0.4046	0.6686	0.4046	0.0324	0.5954	0.5042	0.3370
PAWCS	0.4481	0.7200	0.4481	0.0282	0.5519	0.5524	0.3816
TwoPoints	0.4443	0.3230	0.4443	0.1504	0.5557	0.3740	0.2300
ViBe	0.3830	0.5742	0.3830	0.0459	0.6170	0.4595	0.2983
CodeBook	0.5924	0.2380	0.5924	0.3064	0.4076	0.3396	0.2045

Tableau 7 : Résultats des métriques d'évaluation utilisées pour analyser les données vidéo n°2.

Methods	RC	PC	TPR	FPR	FNR	FM	AC
FrameDifference	0.4838	0.7637	0.4838	0.0224	0.5162	0.5924	0.4208
StaticFrameDifference	0.7903	0.3808	0.7903	0.1925	0.2097	0.5139	0.3458
WeightedMovingMean	0.3992	0.7402	0.3992	0.0210	0.6008	0.5186	0.3501
WeightedMovingVariance	0.4324	0.7391	0.4324	0.0229	0.5676	0.5456	0.3751
MixtureOfGaussianV1	/////	/////	/////	/////	/////	/////	/////
MixtureOfGaussianV2	0.4976	0.7437	0.4976	0.0257	0.5024	0.5962	0.4248
AdaptiveBackgroundLearning	0.7845	0.3535	0.7845	0.2149	0.2155	0.4874	0.3222
AdaptiveSelectiveBackgroundLearning	0.7667	0.7546	0.7667	0.0373	0.2333	0.7606	0.6137
GMG	/////	/////	/////	/////	/////	/////	/////
KNN	0.5040	0.7361	0.5040	0.0271	0.4960	0.5983	0.4269
DPAdaptiveMedian	0.5848	0.9021	0.5848	0.0095	0.4152	0.7096	0.5499
DPGrimsonGMM	0.8201	0.3257	0.8201	0.2543	0.1799	0.4662	0.3040
DPZivkovicAGMM	0.7350	0.7914	0.7350	0.0290	0.2650	0.7621	0.6157
DPMean	0.3153	0.7628	0.3153	0.0147	0.6847	0.4462	0.2872
DPWrenGA	0.7055	0.8443	0.7055	0.0195	0.2945	0.7687	0.6243
DPPratiMediod	0.7124	0.8724	0.7124	0.0156	0.2876	0.7844	0.6452
DPEigenbackground	0.7800	0.5305	0.7800	0.1034	0.2200	0.6315	0.4614
DPTtexture	0.5397	0.9111	0.5397	0.0079	0.4603	0.6778	0.5127
T2FGMM_UM	0.5113	0.8210	0.5113	0.0167	0.4887	0.6302	0.4600
T2FGMM_UV	0.6845	0.2443	0.6845	0.3171	0.3155	0.3600	0.2195
T2FMRF_UM	0	NaN	0	0	1	NaN	0
T2FMRF_UV	0.6110	0.2837	0.6110	0.2310	0.3890	0.3875	0.2403
FuzzySugenoIntegral	0.6061	0.8870	0.6061	0.0116	0.3939	0.7201	0.5626
FuzzyChoquetIntegral	0.6435	0.8905	0.6435	0.0118	0.3565	0.7471	0.5963
Multi-Layer	0.7510	0.8246	0.7510	0.0239	0.2490	0.7861	0.6475
PixelBasedAdaptiveSegmenter	0.7302	0.7389	0.7302	0.0386	0.2698	0.7345	0.5804
LBSimpleGaussian	0.8412	0.2847	0.8412	0.3165	0.1588	0.4254	0.2702
LBFuzzyGaussian	0.8325	0.3268	0.8325	0.2569	0.1675	0.4693	0.3066
LBMixtureOfGaussians	0.7773	0.6702	0.7773	0.0573	0.2227	0.7198	0.5622
LBAadaptiveSOM	0.7911	0.5791	0.7911	0.0861	0.2089	0.6687	0.5023
LBFuzzyAdaptiveSOM	0.7704	0.6390	0.7704	0.0652	0.2296	0.6986	0.5367
LBP_MRF	0.7849	0.8082	0.7849	0.0279	0.2151	0.7964	0.6617
VuMeter	0.5844	0.7642	0.5844	0.0270	0.4156	0.6623	0.4951
KDE	0.8323	0.3815	0.8323	0.2020	0.1677	0.5232	0.3543
IndependentMultimodal	0.8121	0.4390	0.8121	0.1554	0.1879	0.5699	0.3985
MultiCue	0.7284	0.7295	0.7284	0.0404	0.2716	0.7290	0.5735
SigmaDelta	0.7690	0.7047	0.7690	0.0483	0.2310	0.7354	0.5816
SuBSENSE	0.7845	0.8479	0.7845	0.0211	0.2155	0.8150	0.6877
LOBSTER	0.6871	0.8834	0.6871	0.0136	0.3129	0.7730	0.6300
PAWCS	0.7001	0.8958	0.7001	0.0122	0.2999	0.7859	0.6473
TwoPoints	0.7144	0.4901	0.7144	0.1113	0.2856	0.5814	0.4098
ViBe	0.6623	0.8692	0.6623	0.0149	0.3377	0.7518	0.6023
Methods	0.8332	0.2714	0.8332	0.3350	0.1668	0.4095	0.2574

Tableau 8 : Résultats des métriques d'évaluation utilisées pour analyser les données vidéo BMC.

Methods	RC	PC	TPR	FPR	FNR	FM	AC
FrameDifference	0.6845	0.4092	0.6845	0.0257	0.3155	0.5122	0.3443
StaticFrameDifference	0.9264	0.1580	0.9264	0.1285	0.0736	0.2699	0.1560
WeightedMovingMean	0.6043	0.8619	0.6043	0.0025	0.3957	0.7105	0.5509
WeightedMovingVariance	0.6937	0.8293	0.6937	0.0037	0.3063	0.7555	0.6071
MixtureOfGaussianV1	0.6909	0.9385	0.6909	0.0012	0.3091	0.7958	0.6609
MixtureOfGaussianV2	0.6668	0.5596	0.6668	0.0137	0.3332	0.6085	0.4373
AdaptiveBackgroundLearning	0.9278	0.5376	0.9278	0.0208	0.0722	0.6808	0.5160
AdaptiveSelectiveBackgroundLearning	0.9047	0.7330	0.9047	0.0086	0.0953	0.8098	0.6804
GMG	0.9553	0.6100	0.9553	0.0159	0.0447	0.7446	0.5931
KNN	0.9201	0.6888	0.9201	0.0108	0.0799	0.7878	0.6499
DPAdaptiveMedian	0.4136	0.9379	0.4136	7.1289e-04	0.5864	0.5741	0.4026
DPGrimsonGMM	0.9193	0.6732	0.9193	0.0116	0.0807	0.7772	0.6356
DPZivkovicAGMM	0.8864	0.8327	0.8864	0.0046	0.1136	0.8587	0.7524
DPMean	0.4800	0.8611	0.4800	0.0020	0.5200	0.6164	0.4455
DPWrenGA	0.8549	0.8714	0.8549	0.0033	0.1451	0.8631	0.7591
DPPratiMediod	0.8525	0.9419	0.8525	0.0014	0.1475	0.8950	0.8099
DPEigenbackground	0.9159	0.7320	0.9159	0.0087	0.0841	0.8137	0.6859
DPTtexture	0.3768	0.9395	0.3768	6.3151e-04	0.6232	0.5379	0.3679
T2FGMM_UM	0.0392	0.9812	0.0392	1.9531e-05	0.9608	0.0753	0.0391
T2FGMM_UV	0.6205	0.0430	0.6205	0.3597	0.3795	0.0803	0.0419
T2FMRF_UM	0	NaN	0	0	1	NaN	0
T2FMRF_UV	0.3872	0.5659	0.3872	0.0077	0.6128	0.4598	0.2985
FuzzySugenoIntegral	0.8364	0.8690	0.8364	0.0033	0.1636	0.8524	0.7427
FuzzyChoquetIntegral	0.8695	0.8451	0.8695	0.0041	0.1305	0.8571	0.7500
MultiLayer	0.9222	0.9041	0.9222	0.0025	0.0778	0.9130	0.8400
PixelBasedAdaptiveSegmenter	0.9423	0.6804	0.9423	0.0115	0.0577	0.7902	0.6532
LBSimpleGaussian	0.9424	0.0624	0.9424	0.3687	0.0576	0.1170	0.0621
LBFuzzyGaussian	0.9378	0.2239	0.9378	0.0846	0.0622	0.3615	0.2206
LBMixtureOfGaussians	0.9189	0.6897	0.9189	0.0108	0.0811	0.7880	0.6502
LBAAdaptiveSOM	0.9198	0.6943	0.9198	0.0105	0.0802	0.7913	0.6547
LBFuzzyAdaptiveSOM	0.9119	0.7404	0.9119	0.0083	0.0881	0.8172	0.6910
LBP_MRF	0.9398	0.7640	0.9398	0.0076	0.0602	0.8429	0.7284
VuMeter	0.7281	0.7764	0.7281	0.0055	0.2719	0.7515	0.6019
KDE	0.9401	0.1613	0.9401	0.1272	0.0599	0.2753	0.1596
IndependentMultimodal	0.9140	0.6821	0.9140	0.0111	0.0860	0.7812	0.6410
MultiCue	0.9769	0.7029	0.9769	0.0107	0.0231	0.8175	0.6913
SigmaDelta	0.9226	0.6455	0.9226	0.0132	0.0774	0.7596	0.6124
SuBSENSE	0.9291	0.8976	0.9291	0.0028	0.0709	0.9131	0.8400
LOBSTER	0.9079	0.9053	0.9079	0.0025	0.0921	0.9066	0.8292
PAWCS	0.9232	0.9008	0.9232	0.0026	0.0768	0.9118	0.8380
TwoPoints	0.8539	0.3270	0.8539	0.0457	0.1461	0.4729	0.3097
ViBe	0.8475	0.9273	0.8475	0.0017	0.1525	0.8856	0.7947
CodeBook	0.9299	0.1123	0.9299	0.1912	0.0701	0.2005	0.1114

D'après les tableaux 5, 6, et 7, les méthodes qui ont démontré une bonne capacité d'extraction en termes de valeurs sont celles que nous avons mises en évidence en jaune. Nous avons observé que les résultats extraits de la vidéo BMC dans le Tableau 7 ont obtenu des scores très élevés, car une valeur proche de 1 indique toujours un résultat optimal. Certains algorithmes, tels que SuBSENSE, LOBSTER et PAWCS, ont atteint une valeur de 0,9, bien que ces méthodes nécessitent un temps d'exécution plus long (environ 6 à 8 minutes pour une simulation de 11 secondes). Les algo SuBSENSE, LOBSTER et PAWCS n'ont pas été décrits avant. En revanche, d'autres méthodes ont obtenu des valeurs inférieures dans les métriques de notation pour ce type de vidéo, parmi lesquelles CodeBook, LBSimpleGaussian et StaticFrameDifference se sont distinguées grâce à leurs temps d'exécution plus courts. Les algo LBSimpleGaussian et StaticFrameDifference

Concernant la vidéo n°1 du Tableau 5, nous avons remarqué que ce type de vidéo génère des valeurs plus basses pour la mesure F et d'autres échelles de notation par rapport au Tableau 7. Par exemple, l'algorithme AdaptiveSelectiveBackgroundLearning a obtenu la valeur F la plus élevée, à savoir 0,56. Cela indique que la présence d'ombres diminue la qualité de l'extraction des objets en mouvement. Nous avons également noté qu'une augmentation du taux de faux négatifs entraîne une diminution du taux de vrais positifs.

En ce qui concerne la vidéo n°2 du Tableau 6, les résultats montrent que les performances d'évaluation de cette vidéo sont meilleures que celles de la vidéo avec ombres du Tableau 5. La valeur la plus élevée de la mesure F, soit 0,81, a été obtenue par l'algorithme SuBSENSE.

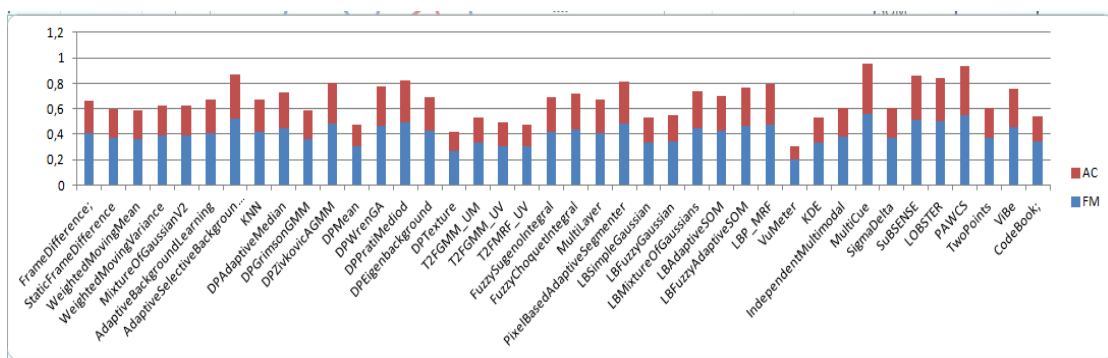


Figure 33. La mesure F et la précision des différentes méthodes pour la video1

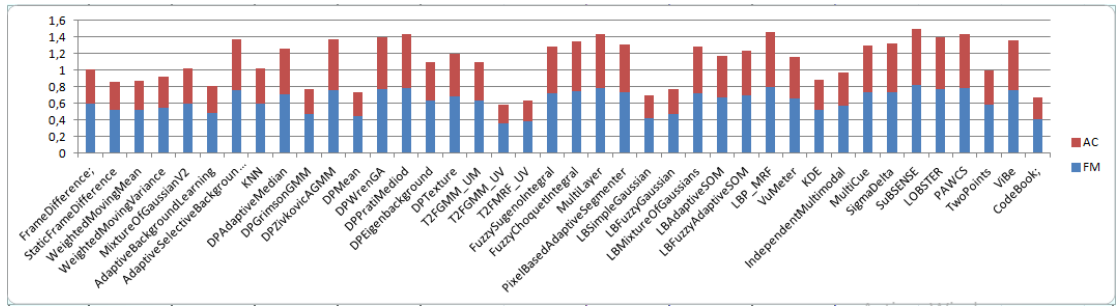


Figure 34. La mesure F et la précision des différentes méthodes pour la video2

Les deux figures précédentes(33, 34) illustrent les valeurs de FM et de précision AC. A partir de ces deux figures, nous pouvons remarquer la différence entre la présence d'ombres et l'absence d'ombres.

On calcule la qualité de l'extraction des méthodes par les métriques précédents. Etant donné que l'idéal est d'avoir (Fm=1, TPR =1 , FPR =0). Après avoir lu les valeurs obtenues dans les tableaux et les figures précédents. On remarque que la qualité de l'extraction des vidéos (1,2) est globalement plus mauvaise avec la qualité de l'extraction de la vidéo BMC, Cela est dû à la présence d'ombres, ce qui réduit la qualité d'extraction.

5.3.2 Évaluation des Performances des Algorithmes en Fonction de Leurs Caractéristiques

Les caractéristiques dépendent du temps d'exécution, lequel varie d'un algorithme à l'autre. En effet, certains offrent une réponse rapide tandis que d'autres exigent davantage de temps, de mémoire et de puissance de traitement. Le tableau ci-dessous présente ces caractéristiques pour chaque algorithme.

Tableau 9 : Caractéristiques de chaque algorithme.

Methods	moyen de temps d'exécution	moyenne de la Mémoire	Moyen de CPU (%)
Frame Difference;	21 s	22	17
Static Frame Difference	18 s	22	21
Weighted Moving Mean	29 s	38	30
Weighted Moving Variance	50 s	51	29
Mixture Of GaussianV1	44.05	57	21
Mixture Of GaussianV2	29 s	47	33
Adaptive Background Learning	26 s	32	23
Adaptive Selective Background Learning	23 s	24	21
GMG	51s	205	33
KNN	45s	44	59
DP Adaptive Median	21s	21	21
DP Grimson GMM	54 s	38	25
DP Zivkovic AGMM	31 s	35	20
DP Mean	26 s	23	21
DP Wren GA	26 s	26	17
DP Prati Mediod	50 s	83	25

DP Eigen background	52 s	276	25
DP Texture	6.11 min	115	26
T2FGMM_UM	1min	38	25
T2FGMM_UV	1.01min	38	25
T2FMRF_UM	1.25min	315	25
T2FMRF_UV	1.20min	315	25
Fuzzy Sugeno Integral	2.38min	43	35
Fuzzy Choquet Integral	2.39min	43	40
Multi Layer	51s	205	39
Pixel Based Adaptive Segmenter	1.32min	105	25
LB Simple Gaussian	27s	34	21
LB Fuzzy Gaussian	41s	34	25
LB Mixture Of Gaussians	38s	70	21
LB Adaptive SOM	59s	116	25
LB Fuzzy Adaptive SOM	59s	116	25
LBP_MRF	2.45min	128	33
Vu Meter	27s	52	21
KDE	29s	83	21
Independent Multi modal	33s	135	22
Multi Cue	39s	69	26
Sigma Delta	27s	23	23
Su BSENSE	6.53min	154	26
LOBSTER	4.23min	109	26
PAWCS	8.40min	673	26
Two Points	21s	20	17
Vi Be	22s	34	21
CodeBook	25s	97	25

Dans le tableau précédent, on peut observer les résultats d'évaluation des performances moyennes, notamment le temps nécessaire à l'exécution de chaque algorithme, la consommation moyenne de mémoire et le processeur CPU moyen. Les algorithmes affichant les temps d'exécution les plus courts sont mis en évidence en jaune, tandis que ceux présentant les valeurs de consommation mémoire les plus élevées sont mis en rouge.

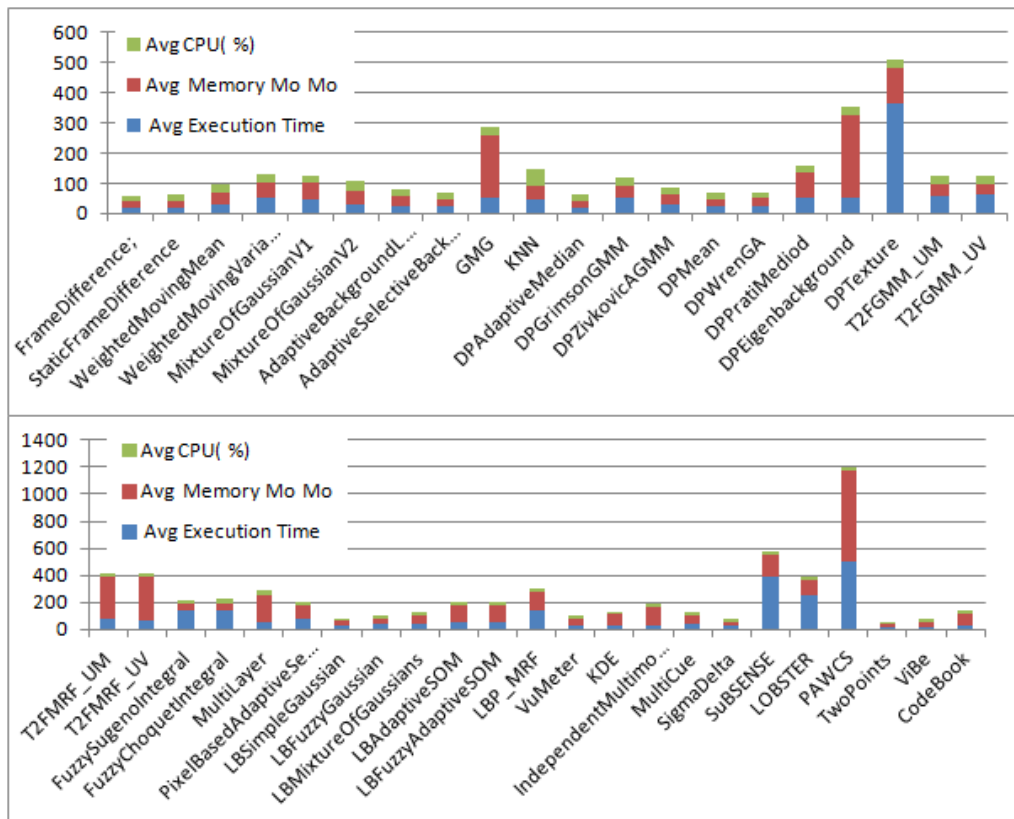


Figure 35. Valeur moyenne du temps, de la mémoire et de la consommation CPU

La figure précédente présente des colonnes graphiques pour toutes les méthodes en termes de temps d'exécution, de consommation mémoire et de processeur. L'objectif de ces colonnes est de permettre une observation visuelle des méthodes ayant les valeurs les plus faibles.

Il est notable que les méthodes qui ont affiché les temps d'exécution les plus courts ainsi que les niveaux de consommation mémoire les plus bas sont (LB Simple Gaussian). À l'inverse, les méthodes qui présentent la plus grande consommation mémoire et les temps d'exécution les plus longs sont celles de la méthode DP Texture.

5.4 Détection et Suivi Automatiques d'Objets en Mouvement

Suite à l'exposition théorique présentée dans les chapitres précédents sur la détection et le suivi d'objets, nous avons mis en œuvre l'application du filtre de Kalman sur un ensemble de séquences d'images capturées dans un atelier. Cela nous a permis d'évaluer les performances de notre algorithme de suivi d'objets. L'implémentation a été réalisée en langage C++ sous Visual Studio 2017. Dans la suite, nous détaillerons notre approche, puis nous présenterons les résultats obtenus à partir des différentes séquences vidéo.

Avant d'effectuer le suivi, il est essentiel de détecter la présence d'objets en mouvement dans la scène. Pour cette phase, en accord avec notre analyse comparative des algorithmes de détection précédents, nous avons choisi d'utiliser la méthode de soustraction de fond basée sur le modèle de mélange gaussien. Quant au processus de suivi, il repose sur une estimation temporelle et spatiale de la position des objets en mouvement dans le champ de vision. Cette approche a pour but de définir la trajectoire des objets en mouvement à travers les images successives de la séquence, grâce à l'utilisation du filtre de Kalman.

5.4.1 Algorithme de Détection

L'algorithme de détection d'objets en mouvement repose sur plusieurs étapes. L'ensemble de ces étapes ont été implémentées.

- Etape 1 : Phase d'initialisation

Le modèle d'arrière-plan est créé en calculant l'intensité moyenne des pixels pour chaque image de la séquence, comme spécifié dans l'équation 80.

$$\mu_n(x, y) = \frac{1}{N_e} \sum_i I_{i,c}(x, y) \quad (80)$$

N_e Le nombre d'images séquentielles dans une scène vidéo, (x,y) les coordonnées des pixels.

- Etape 2 : Détection de changement

La détection de changement est obtenue selon l'équation gaussienne (cf. équation 9/ chapitre 2).

- Etape 3 : Mise à jour

L'actualisation de l'arrière-plan est faite en utilisant la distribution gaussienne mise à jour conformément aux équations précédentes du chapitre 2 (équations 21, 22, 23).

- Etape 4 : Binarisation

La binarisation est obtenue en créant un masque de mouvement binaire pour chaque canal en utilisant l'équation 34.

- Etape 5 : Elimination du bruit

L'élimination du bruit et autres effets indésirables est obtenue en appliquant des filtres morphologiques.

- Etape 6 : Détermination des boîtes englobantes

Elle est réalisée en séparant les pixels du premier plan en carrés.

5.4.2 Algorithme de Suivi

Le suivi de mouvement implémente un filtre de Kalman pour anticiper les futurs emplacements , en supposant que chaque objet se déplace à une vitesse

constante. Chaque itération du processus de suivi commence par une phase de prédiction du système dynamique linéaire, suivie d'une phase de correction. L'algorithme de suivi est formulé comme suit :

Initialisation :

La trajectoire est suivie en calculant le centre de gravité (G) de l'objet détecté en fonction des coordonnées cartésiennes . comme suit :

$$x_G = \frac{1}{N} \sum_{i=0}^{N-1} x_i \tag{81}$$

$$y_G = \frac{1}{N} \sum_{i=0}^{N-1} y_i \tag{82}$$

Où N est le nombre de pixels qui composent l'objet.

Le vecteur des variables d'état est composé de la position et de la vitesse supposée constante $x = (x, y, v_x, v_y)^T$. La position et la vitesse initiales supposées nulles, nous posons donc l'état initial estimé $\hat{x}_0 = 0$.

D'autre part, la matrice de covariance initiale pour ce vecteur d'état est P_0 , et les deux matrices de covariance Q et R sont associées respectivement au bruit d'état v et au bruit de mesure w .

Le filtre de Kalman impose le choix du modèle en fonction de la nature du mouvement de l'objet. Si un objet est en mouvement sur un plan à vitesse constante ($x = x + v\Delta t, \Delta t = 1$), $(x, y, v_x, v_y)^T$, ceci revient à définir la matrice d'état F :

$$F = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{83}$$

D'autre part, la matrice d'observation ne prend en compte que les composantes de position du vecteur d'état, comme le montre sa représentation matricielle :

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{84}$$

Itération :

Le processus de prédiction démarre de $k=1$ avec une estimation initiale \hat{x}_0 de l'état x_0 et sa matrice de covariance associée $P_0 = Q_0 = I$ étant la matrice d'identité.

Prédiction :

Pour la prédiction, nous calculons le vecteur d'état estimé $\hat{x}_{k|k-1}$ et sa Covariance $P_{k|k-1}$.

Correction:

Pour la mise à jour du vecteur d'état estimé \hat{x}_k et sa covariance P_k , il est nécessaire d'obtenir le gain du filtre K_k qui dépend directement des matrices de covariance obtenues pour l'état prédit.

5.4.3 Résultats

Les résultats de la détection et du suivi sont présentés dans les figures ci-dessous.

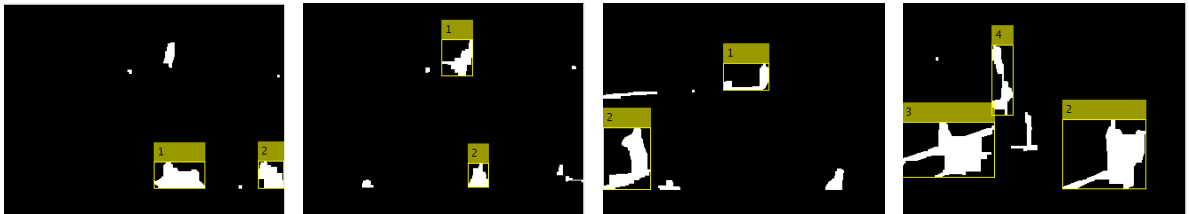


Figure 36. Détecter le mouvement des personnes dans la séquence de la vidéo1, séquence1

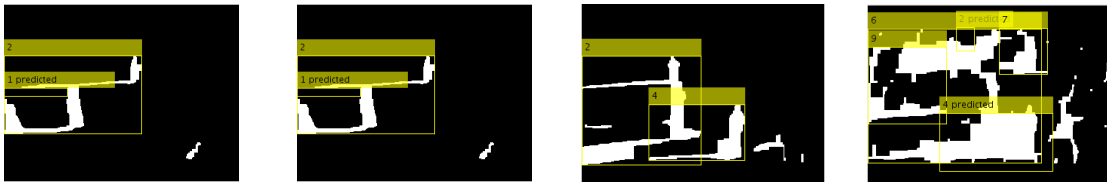


Figure 37. Détecter le mouvement des personnes dans la séquence de la vidéo1 séquence2



Figure 38 . Détecter le mouvement des personnes dans la séquence de la vidéo2, séquence1

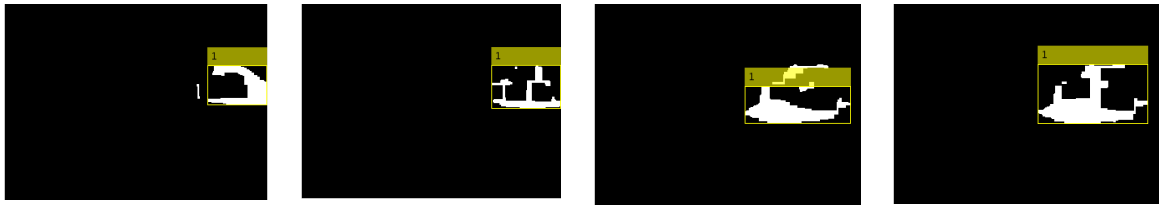


Figure 39. Détecter le mouvement des personnes dans la séquence de la vidéo2, séquence2

Les figures 35, 36, 37 et 38 illustrent les résultats obtenus grâce à l'algorithme de détection d'objets en mouvement. Ces résultats révèlent que l'algorithme parvient efficacement à détecter les objets en mouvement. Cependant, il présente une limitation notable, à savoir qu'il ne parvient pas à distinguer entre les personnes et les ombres. Cette distinction est particulièrement évidente sur les figures 36 et 37, où l'on observe la présence d'ombres et leur absence.



Figure 40. Séquence 1 de la vidéo 1



Figure 41. Séquence 2 de la vidéo 1



Figure 42. Séquence 1 de la vidéo 2



Figure 43. Séquence 2 de la vidéo 2

Les figures 39, 40, 41 et 42 illustrent les résultats de l'algorithme de suivi des objets en mouvement. Ces résultats ont permis de mettre en évidence les avantages et les inconvénients du filtre de Kalman en matière de suivi d'objets en mouvement. Toutefois, il est important de noter que cet algorithme présente des difficultés lorsqu'il est confronté à des ombres superposées, comme le montre la figure 40. Même en l'absence d'ombres, l'algorithme a tendance à fusionner le mouvement de deux personnes proches l'une de l'autre en une seule entité, comme illustré dans la figure 41. Lorsqu'il s'agit de suivre un seul objet en mouvement, l'algorithme se comporte de manière optimale, comme le montre la figure 42.

Cependant, il est important de souligner que ces comportements ne sont pas compatibles avec la réalité des vidéos de surveillance où il est essentiel de suivre précisément des individus, du trafic, etc. Ces limitations sont considérées comme des lacunes du filtre de Kalman dans ce contexte spécifique.

5.5 Conclusion

Dans ce chapitre, nous avons mené une étude approfondie de notre proposition en évaluant et en comparant 43 techniques de détection d'objets en mouvement dans diverses conditions. Pour assurer une évaluation optimale en temps réel, nous avons pris en considération des paramètres tels que la consommation de mémoire, les performances de hachage, le temps d'exécution et l'utilisation du processeur. Nous avons réalisé nos expériences en utilisant deux types de vidéos. Le premier type consistait en de véritables vidéos capturées dans un environnement contrôlé avec différentes conditions d'éclairage, tandis que le deuxième type utilisait l'ensemble de données BMC (Background Modeling Challenge) en introduisant du bruit dans des vidéos synthétiques.

Les résultats expérimentaux ont clairement montré que les vidéos synthétiques de l'ensemble de données BMC présentaient de meilleures performances, avec un taux de vrais positifs plus élevé et un taux de faux négatifs plus faible. Ces résultats ont été comparés à ceux obtenus avec des vidéos réelles, mettant en

lumière les défis posés par les ombres. En outre, nous avons identifié que les algorithmes capables de surmonter les images fantômes et le bruit, tels que MultiCue, DPWrenGA, PixelBasedAdaptiveSegmenter, DPZivkovicAGMM et AdaptiveSelectiveBackgroundLearning, ont obtenu des performances remarquables.

Parallèlement à cette étude sur la détection, nous avons également abordé deux défis majeurs dans l'analyse des mouvements à partir de séquences d'images : la détection et le suivi d'objets en mouvement dans les scènes vidéo. Pour la détection, nous avons utilisé un algorithme de rétro-soustraction basé sur des distributions gaussiennes, représentant la famille de méthodes de détection de mouvement la plus couramment utilisée. En ce qui concerne le suivi, nous avons opté pour le filtre de Kalman, qui fonctionne en deux étapes : la prédiction de la position de la silhouette et la correction de l'estimation résultant de cette prédiction. Les résultats ont mis en évidence à la fois les avantages et les inconvénients de l'algorithme du filtre de Kalman, soulignant qu'il ne résout pas de manière optimale tous les problèmes de suivi inhérents.

CONCLUSION

Au terme de cette thèse, nous avons abordé en profondeur la technologie de la vidéosurveillance intelligente ainsi que les étapes clés nécessaires à la construction d'un système de surveillance complet. Notre attention s'est concentrée en particulier sur l'étape cruciale de découverte des objets en mouvement au sein du flux vidéo. Cette étape, en tant que pierre angulaire du système, a fait l'objet d'une mise en œuvre exhaustive des méthodes existantes dans ce domaine. Notre objectif était de déterminer la méthode la plus précise et la plus performante pour répondre aux défis réels liés à l'extraction d'objets en mouvement.

Cette recherche nous a conduit à évaluer et comparer 43 techniques de détection d'objets en mouvement sous diverses conditions. Pour une évaluation complète, optimale et en temps réel, nous avons pris en compte des critères tels que la consommation de mémoire, les performances de hachage, le temps d'exécution et l'utilisation du processeur. Deux types de vidéos ont été sélectionnés pour ces expérimentations. Le premier type impliquait des vidéos réelles capturées dans un environnement contrôlé avec différentes conditions d'éclairage. Le second type a utilisé le jeu de données BMC (Background Modeling Challenge) en incorporant du bruit dans des vidéos synthétiques. Les résultats expérimentaux ont démontré que les vidéos synthétiques de l'ensemble de données BMC affichaient de meilleures performances, avec un taux de vrais positifs plus élevé et un taux de faux négatifs inférieur. Les résultats obtenus ont été comparés à ceux des vidéos réelles, mettant en évidence les défis posés par les ombres. Par ailleurs, nous avons constaté que les algorithmes réussissant à surmonter les ombres et le bruit, tels que « MultiCue », « DPWrenGA », « PixelBasedAdaptiveSegmenter », « DPZivkovicAGMM » et « AdaptiveSelectiveBackgroundLearning », ont obtenu des performances remarquables.

Parallèlement, nous avons abordé deux défis majeurs dans l'analyse du mouvement à partir d'une séquence d'images : la détection et le suivi des objets en mouvement dans des scènes simples. En matière de détection, nous avons employé un algorithme de soustraction a posteriori basé sur les distributions gaussiennes, qui représentent la famille la plus répandue de méthodes de détection de mouvement. Pour le suivi, nous avons opté pour un filtre de Kalman qui procède en deux étapes : prédiction de la position de la silhouette et correction

de l'estimation issue de la prédiction, améliorant ainsi la trajectoire suivie tout en réduisant le bruit. Le filtre de Kalman propose une solution itérative, optimale et aisément applicable au problème de filtrage dans le cas de systèmes linéaires gaussiens. Néanmoins, il est important de souligner que l'approche de suivi par filtre de Kalman ne résout pas de manière optimale tous les problèmes inhérents au suivi.

Enfin, nous avons combiné la méthode de suivi par filtre de Kalman avec la méthode de détection de soustraction de fond. Cette combinaison a permis de doter notre système de la capacité à suivre simultanément plusieurs objets en mouvement. Les tests ont été réalisés sur des séquences d'images capturées en environnement contrôlé, et les résultats ont été minutieusement discutés.

Pour l'avenir, il est prévu d'étendre ces méthodes à des conditions plus complexes, telles que la neige, le brouillard, le vent, les foules, les ombres et le bruit. L'un des défis majeurs de la vision artificielle reste la conception et la mise à jour du modèle d'arrière-plan. En conséquence, nos futurs travaux se concentreront sur le développement d'un système de surveillance intelligent capable de détecter et d'actualiser avec précision et efficacité le modèle d'arrière-plan, tout en faisant face aux variations dynamiques de l'environnement et aux défis posés par les ombres. Ce travail de développement visera une exécution en temps réel, tout en maintenant des performances élevées et un algorithme simple.

REFERENCES

- [1] I. E. Olatunji and C.-H. Cheng, "Video Analytics for Visual Surveillance and Applications: An Overview and Survey," in *Machine Learning Paradigms: Applications of Learning and Analytics in Intelligent Systems*, G. A. Tsihrintzis, M. Virvou, E. Sakkopoulos, and L. C. Jain, Eds., in Learning and Analytics in Intelligent Systems. Cham: Springer International Publishing, 2019, pp. 475–515. doi: 10.1007/978-3-030-15628-2_15.
- [2] Valérie Gouaillier, "La vidéosurveillance intelligente : promesses et défis," 2009. Rapport de veille technologique et commerciale
- [3] H. H. Huynh, "Vidéosurveillance pour appartements intelligents : application à la détection de prise de médicaments," These de doctorat, Aix-Marseille 2, 2010. Accessed: Sep. 15, 2022. [Online]. Available: <https://www.theses.fr/2010AIX22133>
- [4] F. Bousetouane and L. Dib, "Traitement décentralisé des informations visuelles dans les réseaux de capteurs sans fil multimédia," These de doctorat, Badji Mokhtar, Annaba, 2015.
- [5] M. Brulin, "Analyse sémantique d'un trafic routier dans un contexte de vidéo-surveillance," These de doctorat, Bordeaux 1, 2012. Accessed: Feb. 17, 2022. [Online]. Available: <https://www.theses.fr/2012BOR14589>
- [6] X. Wang, "Intelligent multi-camera video surveillance: A review," *Pattern Recognition Letters*, vol. 34, no. 1, pp. 3–19, 2013, doi: 10.1016/j.patrec.2012.07.005.
- [7] A. Shafie, F. Hafiz, and M. H. Ali, "Motion detection techniques using optical flow," *International Journal of Electrical and Computer Engineering*, p. 3(8), 1555-1557., 2009.
- [8] J Withagen, P., "Object detection and segmentation for visual surveillance," 2006. Dissertation
- [9] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Comput. Surv.*, vol. 38, no. 4, pp. 13-es, 2006, doi: 10.1145/1177352.1177355.
- [10] L. Wang, W. Hu, and T. Tan, "Recent developments in human motion analysis," *Pattern Recognition*, vol. 36, no. 3, pp. 585–601, Mar. 2003, doi: 10.1016/S0031-3203(02)00100-0.
- [11] C. Stauffer, "Estimating Tracking Sources and Sinks," in *2003 Conference on Computer Vision and Pattern Recognition Workshop*, Jun. 2003, pp. 35–35. doi: 10.1109/CVPRW.2003.10036.
- [12] A. Bissacco, P. Saisan, and S. Soatto, "Gait Recognition using Dynamic Affine Invariants," p. 7. Computer Science, Engineering, Published 2004
- [13] A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz, "Robust Real-Time Unusual Event Detection using Multiple Fixed-Location Monitors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 3, pp. 555–560, Mar. 2008, doi: 10.1109/TPAMI.2007.70825.

- [14] P. Turaga, R. Chellappa, V. S. Subrahmanian, and O. Udrea, "Machine Recognition of Human Activities: A Survey," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 11, pp. 1473–1488, 2008, doi: 10.1109/TCSVT.2008.2005594.
- [15] M. O. Mehmood, "People detection methods for intelligent multi-Camera surveillance systems," p. 155. Thèse de doctorat, ECOLE CENTRALE DE LILLE, Septembre 2015.
- [16] A. Griesser, "Oscar: the opportunistic scanner," *Selected readings in vision and graphics*, vol. 047, 2007, doi: 10.3929/ethz-a-005487575.
- [17] J. K. Aggarwal and M. S. Ryoo, "Human activity analysis: A review," *ACM Comput. Surv.*, vol. 43, no. 3, p. 16:1-16:43, 2011, doi: 10.1145/1922649.1922653.
- [18] N. Verbeke, "Suivi d'objets en mouvement dans une séquence vid'eo," p. 146. These
- [19] T. Yang, S. Z. Li, Q. Pan, and J. Li, "Real-time and accurate segmentation of moving objects in dynamic scene," in *Proceedings of the ACM 2nd international workshop on Video surveillance & sensor networks*, in VSSN '04. New York, NY, USA: Association for Computing Machinery, 2004, pp. 136–143. doi: 10.1145/1026799.1026822.
- [20] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, "Pfinder: real-time tracking of the human body," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 780–785, Jul. 1997, doi: 10.1109/34.598236.
- [21] N. M. Oliver, B. Rosario, and A. P. Pentland, "A Bayesian computer vision system for modeling human interactions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 831–843, Aug. 2000, doi: 10.1109/34.868684.
- [22] Y. Weiss and E. H. Adelson, "A unified mixture framework for motion segmentation: incorporating spatial coherence and estimating the number of models," in *Proceedings CVPR IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 1996, pp. 321–326. doi: 10.1109/CVPR.1996.517092.
- [23] J.-M. Odobez and P. Bouthemy, "Direct incremental model-based image motion segmentation for video analysis," *Signal Processing*, vol. 66, no. 2, pp. 143–155, Apr. 1998, doi: 10.1016/S0165-1684(98)00003-6.
- [24] T. B. Moeslund, A. Hilton, and V. Krüger, "A survey of advances in vision-based human motion capture and analysis," *Computer Vision and Image Understanding*, vol. 104, no. 2, pp. 90–126, Nov. 2006, doi: 10.1016/j.cviu.2006.08.002.
- [25] W. Hu, T. Tan, L. Wang, and S. Maybank, "A survey on visual surveillance of object motion and behaviors," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 34, no. 3, pp. 334–352, 2004, doi: 10.1109/TSMCC.2004.829274.
- [26] S. S. Cheung and C. Kamath, "Robust techniques for background subtraction in urban traffic video," in *Visual Communications and Image Processing 2004*, SPIE, Jan. 2004, pp. 881–892. doi: 10.1117/12.526886.
- [27] M. Piccardi, "Background subtraction techniques: a review," in *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583)*, 2004, pp. 3099–3104 vol.4. doi: 10.1109/ICSMC.2004.1400815.
- [28] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, "Wallflower: principles and practice of background maintenance," in *Proceedings of the Seventh IEEE*

- International Conference on Computer Vision*, Sep. 1999, pp. 255–261 vol.1. doi: 10.1109/ICCV.1999.791228.
- [29] R. Pless, “Spatio-temporal Background Models for Outdoor Surveillance,” *EURASIP J. Adv. Signal Process.*, vol. 2005, no. 14, p. 101240, Aug. 2005, doi: 10.1155/ASP.2005.2281.
- [30] M. Heikkila and M. Pietikainen, “A texture-based method for modeling the background and detecting moving objects,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 4, pp. 657–662, Apr. 2006, doi: 10.1109/TPAMI.2006.68.
- [31] C. Stauffer and W. Grimson, “Adaptive Background Mixture Models for Real-Time Tracking,” in *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)(CVPR)*, 1999, p. 2246. doi: 10.1109/CVPR.1999.784637.
- [32] A. Elgammal, D. Harwood, and L. Davis, “Non-parametric Model for Background Subtraction,” in *Computer Vision — ECCV 2000*, D. Vernon, Ed., in Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2000, pp. 751–767. doi: 10.1007/3-540-45053-X_48.
- [33] B. Han, D. Comaniciu, and L. Davis, “Sequential kernel density approximation through mode propagation: Applications to background modeling,” *Proceedings of ACCV*, vol. 1, Jan. 2004.
- [34] A. M. McIvor, “Background Subtraction Techniques,” p. 6. article
- [35] S. J. McKenna, S. Jabri, Z. Duric, A. Rosenfeld, and H. Wechsler, “Tracking Groups of People,” *Computer Vision and Image Understanding*, vol. 80, no. 1, pp. 42–56, Oct. 2000, doi: 10.1006/cviu.2000.0870.
- [36] D. Mokhtari, “Détection des chutes par calcul homographique,” Mémoire en vue de l’obtention du grade de Maîtrise en Informatique, Département d’Informatique et de Recherche Opérationnelle Faculté des Arts et des Sciences, Université de Montréal, 2012.
- [37] W. E. L. Grimson, C. Stauffer, R. Romano, and L. Lee, “Using adaptive tracking to classify and monitor activities in a site,” in *Proceedings. 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No.98CB36231)*, Jun. 1998, pp. 22–29. doi: 10.1109/CVPR.1998.698583.
- [38] M. Harville, “A Framework for High-Level Feedback to Adaptive, Per-Pixel, Mixture-of-Gaussian Background Models,” in *Computer Vision — ECCV 2002*, A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, Eds., in Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2002, pp. 543–560.
- [39] C. Stauffer and W. E. L. Grimson, “Learning patterns of activity using real-time tracking,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 747–757, Aug. 2000, doi: 10.1109/34.868677.
- [40] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, “Pfinder: real-time tracking of the human body,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 780–785, Jul. 1997, doi: 10.1109/34.598236.
- [41] Z. Zivkovic, “Improved adaptive Gaussian mixture model for background subtraction,” in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.*, Aug. 2004, pp. 28-31 Vol.2. doi: 10.1109/ICPR.2004.1333992.
- [42] Z. Zivkovic and F. van der Heijden, “Efficient adaptive density estimation per image pixel for the task of background subtraction,” *Pattern Recognition Letters*, vol. 27, no. 7, pp. 773–780, May 2006, doi: 10.1016/j.patrec.2005.11.005.

- [43] E. López-Rubio and R. M. Luque-Baena, "Stochastic approximation for background modelling," *Computer Vision and Image Understanding*, vol. 115, no. 6, pp. 735–749, Jun. 2011, doi: 10.1016/j.cviu.2011.01.007.
- [44] F. J. López-Rubio, E. López-Rubio, M. A. Molina-Cabello, R. M. Luque-Baena, E. J. Palomo, and E. Domínguez, "The effect of noise on foreground detection algorithms," *Artif Intell Rev*, vol. 49, no. 3, pp. 407–438, Mar. 2018, doi: 10.1007/s10462-016-9525-3.
- [45] A. Sobral and A. Vacavant, "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos," *Computer Vision and Image Understanding*, vol. 122, pp. 4–21, May 2014, doi: 10.1016/j.cviu.2013.12.005.
- [46] H. zhang and D. xu, "Fusing Color and Texture Features for Background Model," in *Fuzzy Systems and Knowledge Discovery*, L. Wang, L. Jiao, G. Shi, X. Li, and J. Liu, Eds., in Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2006, pp. 887–893. doi: 10.1007/11881599_110.
- [47] M. SUGENO, "Theory of fuzzy integrals and its applications," *Doct. Thesis, Tokyo Institute of Technology*, 1974, Accessed: Dec. 16, 2020. [Online]. Available: <https://ci.nii.ac.jp/naid/10017209011/>
- [48] T. Bouwmans and F. El Baf, "Modeling of Dynamic Backgrounds by Type-2 Fuzzy Gaussians Mixture Models," *MASAUM Journal of of Basic and Applied Sciences*, vol. 1, no. 2, pp. 265–276, Feb. 2010.
- [49] G. Choquet, "Theory of capacities," *Annales de l'Institut Fourier*, vol. 5, pp. 131–295, 1954, doi: 10.5802/aif.53.
- [50] M. M. Azab, H. A. Shedeed, and A. S. Hussein, "A new technique for background modeling and subtraction for motion detection in real-time videos," in *2010 IEEE International Conference on Image Processing*, Sep. 2010, pp. 3453–3456. doi: 10.1109/ICIP.2010.5653748.
- [51] M. H. Sigari, N. Mozayani, and H. R. Pourreza, *Fuzzy Running Average and Fuzzy Background Subtraction: Concepts and Application*. International Journal of Computer Science and Network Security, VOL.8 No.2, February 2008.
- [52] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Real-time foreground–background segmentation using codebook model," *Real-Time Imaging*, vol. 11, no. 3, pp. 172–185, Jun. 2005, doi: 10.1016/j.rti.2004.12.004.
- [53] M. Julian, M. Cyril, and R. Yassine, "Vers une meilleure détection des objets mobiles par Codebook et invariance couleur - Archive ouverte HAL," 2013.
- [54] A. Ilyas, M. Scuturici, and S. Miguet, "Real Time Foreground-Background Segmentation Using a Modified Codebook Model," in *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*, Sep. 2009, pp. 454–459. doi: 10.1109/AVSS.2009.85.
- [55] D. Culibrk, O. Marques, D. Socek, H. Kalva, and B. Furht, "Neural Network Approach to Background Modeling for Video Object Segmentation," *IEEE Transactions on Neural Networks*, vol. 18, no. 6, pp. 1614–1627, Nov. 2007, doi: 10.1109/TNN.2007.896861.
- [56] L. Maddalena and A. Petrosino, "A Self-Organizing Approach to Background Subtraction for Visual Surveillance Applications," *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1168–1177, Jul. 2008, doi: 10.1109/TIP.2008.924285.
- [57] L. Maddalena and A. Petrosino, "A fuzzy spatial coherence-based approach to background/foreground separation for moving object detection," *Neural Comput & Applic*, vol. 19, no. 2, pp. 179–186, Mar. 2010, doi: 10.1007/s00521-009-0285-8.

- [58] A. Garcia-Martin and J. M. Martinez, "Robust Real Time Moving People Detection in Surveillance Scenarios," in *2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance*, 2010, pp. 241–247. doi: 10.1109/AVSS.2010.33.
- [59] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting moving objects, ghosts, and shadows in video streams," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1337–1342, Oct. 2003, doi: 10.1109/TPAMI.2003.1233909.
- [60] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam, "Image change detection algorithms: a systematic survey," *IEEE Transactions on Image Processing*, vol. 14, no. 3, pp. 294–307, Mar. 2005, doi: 10.1109/TIP.2004.838698.
- [61] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski, "A Database and Evaluation Methodology for Optical Flow," *Int J Comput Vis*, vol. 92, no. 1, pp. 1–31, Mar. 2011, doi: 10.1007/s11263-010-0390-2.
- [62] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques," *International Journal of Computer Vision*, vol. 12, no. 1, pp. 43–77, Feb. 1994, doi: 10.1007/bf01420984.
- [63] D. L. Almanza-Ojeda, "Détection et suivi d'objets mobiles perçus depuis un capteur visuel embarqué," p. 141. THESE 2011
- [64] K.-W. Chen, C.-C. Lai, P.-J. Lee, C.-S. Chen, and Y.-P. Hung, "Adaptive Learning for Target Tracking and True Linking Discovering Across Multiple Non-Overlapping Cameras," *IEEE Transactions on Multimedia*, vol. 13, no. 4, pp. 625–638, 2011, doi: 10.1109/TMM.2011.2131639.
- [65] F. Bousetouane, L. Dib, and H. Snoussi, "Improved mean shift integrating texture and color features for robust real time object tracking," *Vis Comput*, vol. 29, no. 3, pp. 155–170, Mar. 2013, doi: 10.1007/s00371-012-0677-0.
- [66] F. Bousetouane, C. Motamed, and L. Dib, "Contextual adaptive particle filtering for robust real-time non-rigid object tracking," in *2013 8th International Workshop on Systems, Signal Processing and their Applications (WoSSPA)*, 2013, pp. 127–132. doi: 10.1109/WoSSPA.2013.6602349.
- [67] A. Hanbury and J. Serra, "Colour Image Analysis in 3D-Polar Coordinates," in *Pattern Recognition*, B. Michaelis and G. Krell, Eds., in Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2003, pp. 124–131. doi: 10.1007/978-3-540-45243-0_17.
- [68] A. Loza, A. Achim, D. Bull, and N. Canagarajah, "Statistical Image Fusion with Generalised Gaussian and Alpha-Stable Distributions," in *2007 15th International Conference on Digital Signal Processing*, Jul. 2007, pp. 268–271. doi: 10.1109/ICDSP.2007.4288570.
- [69] B. BAILEY, "Moments in image processing," Nov 2002.
- [70] Md. M. Rahman, P. Bhattacharya, and B. C. Desai, "A Framework for Medical Image Retrieval Using Machine Learning and Statistical Similarity Matching Techniques With Relevance Feedback," *IEEE Transactions on Information Technology in Biomedicine*, vol. 11, no. 1, pp. 58–69, 2007, doi: 10.1109/TITB.2006.884364.
- [71] A. Bekkari, S. Idbraim, D. Mammass, and M. E. Yassa, "Exploiting spectral and space information in classification of high resolution urban satellites images using Haralick features and SVM," in *2011 International Conference on Multimedia Computing and Systems*, Apr. 2011, pp. 1–4. doi: 10.1109/ICMCS.2011.5945611.

- [72] S. Chitrakala, P. Shamini, and D. Manjula, "Multi-class Enhanced Image Mining of Heterogeneous Textual Images Using Multiple Image Features," in *2009 IEEE International Advance Computing Conference*, Mar. 2009, pp. 496–501. doi: 10.1109/IADCC.2009.4809061.
- [73] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. SMC-3, no. 6, pp. 610–621, 1973, doi: 10.1109/TSMC.1973.4309314.
- [74] F. Bousetouane, L. Dib, and H. Snoussi, "Robust detection and tracking pedestrian object for real time surveillance applications," in *International Conference on Graphic and Image Processing (ICGIP 2011)*, SPIE, Sep. 2011, pp. 64–70. doi: 10.1117/12.913034.
- [75] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person re-identification by symmetry-driven accumulation of local features," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2010, pp. 2360–2367. doi: 10.1109/CVPR.2010.5539926.
- [76] H. Deng, W. Zhang, E. Mortensen, T. Dietterich, and L. Shapiro, "Principal Curvature-Based Region Detector for Object Recognition," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2007, pp. 1–8. doi: 10.1109/CVPR.2007.382972.
- [77] A. C. Berg, T. L. Berg, and J. Malik, "Shape matching and object recognition using low distortion correspondences," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Jun. 2005, pp. 26–33 vol. 1. doi: 10.1109/CVPR.2005.320.
- [78] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, "Robust Object Recognition with Cortex-Like Mechanisms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 3, pp. 411–426, Mar. 2007, doi: 10.1109/TPAMI.2007.56.
- [79] B. Wu, X. Song, V. K. Singh, and R. Nevatia, "Evaluation of USC Human Tracking System for Surveillance Videos," in *Multimodal Technologies for Perception of Humans*, R. Stiefelhagen and J. Garofolo, Eds., in Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2007, pp. 183–189. doi: 10.1007/978-3-540-69568-4_14.
- [80] I. Benraya and N. Benblidia, "Comparison of Background Subtraction methods," in *2018 International Conference on Applied Smart Systems (ICASS)*, 2018, pp. 1–5. doi: 10.1109/ICASS.2018.8652040.
- [81] I. Benraya, N. Benblidia, and Y. Amara, "Real-time performance evaluation of BGSLibrary algorithms for intelligent surveillance," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 24, no. 3, Art. no. 3, Dec. 2021, doi: 10.11591/ijeecs.v24.i3.pp1491-1498.
- [82] P. Pérez, C. Hue, J. Vermaak, and M. Gangnet, "Color-Based Probabilistic Tracking," in *Computer Vision — ECCV 2002*, A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, Eds., in Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2002, pp. 661–675. doi: 10.1007/3-540-47969-4_44.
- [83] G. Takacs, V. Chandrasekhar, S. Tsai, D. Chen, R. Grzeszczuk, and B. Girod, "Unified Real-Time Tracking and Recognition with Rotation-Invariant Fast Features," in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2010, pp. 934–941. doi: 10.1109/CVPR.2010.5540116.
- [84] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," in *Proceedings IEEE Conference on Computer Vision*

and Pattern Recognition. *CVPR 2000 (Cat. No. PR00662)*, Jun. 2000, pp. 142–149 vol.2. doi: 10.1109/CVPR.2000.854761.

[85] P. Pérez, C. Hue, J. Vermaak, and M. Gangnet, “Color-Based Probabilistic Tracking,” in *Computer Vision — ECCV 2002*, A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, Eds., in Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2002, pp. 661–675. doi: 10.1007/3-540-47969-4_44.

[86] M.-Y. Ju, C.-S. Ouyang, and H.-S. Chang, “Mean shift tracking using fuzzy color histogram,” in *2010 International Conference on Machine Learning and Cybernetics*, Jul. 2010, pp. 2904–2908. doi: 10.1109/ICMLC.2010.5580780.

[87] I. Leichter, M. Lindenbaum, and E. Rivlin, “Mean Shift tracking with multiple reference color histograms,” *Computer Vision and Image Understanding*, vol. 114, no. 3, pp. 400–408, Mar. 2010, doi: 10.1016/j.cviu.2009.12.006.

[88] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, “A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking,” *IEEE Transactions on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002, doi: 10.1109/78.978374.

[89] A. Agarwal, “Machine Learning for Image Based Motion Capture,” phdthesis, Institut National Polytechnique de Grenoble - INPG, 2006. Accessed: Sep. 24, 2022. [Online]. Available: <https://tel.archives-ouvertes.fr/tel-00390301>

[90] J. J. Pantrigo, A. S. Montemayor, and R. Cabido, “Scatter Search Particle Filter for 2D Real-Time Hands and Face Tracking,” in *Image Analysis and Processing – ICIAP 2005*, F. Roli and S. Vitulano, Eds., in Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2005, pp. 953–960. doi: 10.1007/11553595_117.

[91] S.-K. Weng, C.-M. Kuo, and S.-K. Tu, “Video object tracking using adaptive Kalman filter,” *Journal of Visual Communication and Image Representation*, vol. 17, no. 6, pp. 1190–1208, Dec. 2006, doi: 10.1016/j.jvcir.2006.03.004.

[92] B. Nicola and H. Huosheng, “People tracking with a mobile robot,” In Proc. of the 13th IASTED Int. Conf. on Robotics and Applications, 2007.

[93] D. P. Spanos, R. Olfati-Saber, and R. M. Murray, “Approximate distributed Kalman filtering in sensor networks with quantifiable performance,” in *IPSN 2005. Fourth International Symposium on Information Processing in Sensor Networks, 2005.*, Apr. 2005, pp. 133–139. doi: 10.1109/IPSN.2005.1440912.

[94] M. Isard and A. Blake, “CONDENSATION—Conditional Density Propagation for Visual Tracking,” *International Journal of Computer Vision*, vol. 29, no. 1, pp. 5–28, Aug. 1998, doi: 10.1023/A:1008078328650.

[95] Q. Ma, “Abnormal Event Detection in Videos Based on Deep Neural Networks,” *Scientific Programming*, vol. 2021, p. e6412608, Aug. 2021, doi: 10.1155/2021/6412608.

[96] R. Caroline, “Vidéosurveillance intelligente pour la détection de chutes chez les personnes âgées,” Thèse présentée à la Faculté des arts et des sciences en vue de l’obtention du grade de Philosophiæ Doctor (Ph.D.) en informatique, Université de Montréal, 2010.

[97] T. Lee, “An intelligent emergency response system: preliminary development and testing of automated fall detection - Tracy Lee, Alex Mihailidis, 2005,” 4): – 198, 2005 194AD.

<https://journals.sagepub.com/doi/abs/10.1258/1357633054068946> (accessed Oct. 06, 2022).

[98] H. Nait-Charif and S. J. McKenna, “Activity summarisation and fall detection in a supportive home environment,” in *Proceedings of the 17th International*

- Conference on Pattern Recognition, 2004. ICPR 2004.*, Aug. 2004, pp. 323-326 Vol.4. doi: 10.1109/ICPR.2004.1333768.
- [99] F. Mokhtarian, "Silhouette-based isolated object recognition through curvature scale space," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 5, pp. 539–544, May 1995, doi: 10.1109/34.391387.
- [100] R. Diaz De Leon and L. E. Sucar, "Human silhouette recognition with Fourier descriptors," in *Proceedings 15th International Conference on Pattern Recognition. ICPR-2000*, Sep. 2000, pp. 709–712 vol.3. doi: 10.1109/ICPR.2000.903643.
- [101] V. A. A. K. Roy-Chowdhury, and R. Chellappa, "Matching shape sequences in video with applications in human movement analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 12, pp. 1896–1909, Dec. 2005, doi: 10.1109/TPAMI.2005.246.
- [102] C. Bauckhage, J. K. Tsotsos, and F. E. Bunn, "Detecting abnormal gait," in *The 2nd Canadian Conference on Computer and Robot Vision (CRV'05)*, May 2005, pp. 282–288. doi: 10.1109/CRV.2005.32.
- [103] O. Richard, Duda, Peter E., Hart et David G., and Stork, "Pattern Classification - ProQuest."
<https://www.proquest.com/openview/278d79e0719c26331f9ced035aa5ee87/1?cbi=37651&pq-origsite=gscholar> (accessed Oct. 09, 2022).
- [104] G. Dreyfus, M. Martinez, J. Samuelides, M., Gordon, M. B., Badran, F., and Thiria, S., & Hérault, L., "LES RÉSEAUX DE NEURONES," p. 19, 2002.
- [105] I. L. Dryden and K. V. Mardia, "Multivariate Shape Analysis," *Sankhyā: The Indian Journal of Statistics, Series A (1961-2002)*, vol. 55, no. 3, pp. 460–480, 1993.
- [106] Rao, Gritai, Shah, and Syeda-Mahmood, "View-invariant alignment and matching of video sequences," in *Proceedings Ninth IEEE International Conference on Computer Vision*, Oct. 2003, pp. 939–945 vol.2. doi: 10.1109/ICCV.2003.1238449.
- [107] R. Niels and L. Vuurpijl, "Dynamic time warping applied to Tamil character recognition," in *Eighth International Conference on Document Analysis and Recognition (ICDAR'05)*, Aug. 2005, pp. 730-734 Vol. 2. doi: 10.1109/ICDAR.2005.96.
- [108] A. Corradini, "Dynamic time warping for off-line recognition of a small gesture vocabulary," in *Proceedings IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, Jul. 2001, pp. 82–89. doi: 10.1109/RATFG.2001.938914.
- [109] L. Jiao, Y. Wu, G. Wu, E. Y. Chang, and Y.-F. Wang, "Anatomy of a multicamera video surveillance system," *Multimedia Systems*, vol. 10, no. 2, pp. 144–163, Aug. 2004, doi: 10.1007/s00530-004-0147-2.
- [110] A. Sixsmith and N. Johnson, "A smart sensor to detect the falls of the elderly," *IEEE Pervasive Computing*, vol. 3, no. 2, pp. 42–47, Apr. 2004, doi: 10.1109/MPRV.2004.1316817.
- [111] A. Sobral, "BGSLibrary: An OpenCV C++ Background Subtraction Library," BOOK. - IX Workshop de Visão Computacional (WVC'2013).
- [112] N. J. B. McFarlane and C. P. Schofield, "Segmentation and tracking of piglets in images," *Machine Vis. Apps.*, vol. 8, no. 3, pp. 187–193, May 1995, doi: 10.1007/BF01215814.
- [113] S. Calderara, R. Melli, A. Prati, and R. Cucchiara, "Reliable background suppression for complex scenes," in *Proceedings of the 4th ACM international workshop on Video surveillance and sensor networks*, in VSSN '06. New York,

- NY, USA: Association for Computing Machinery, Oct. 2006, pp. 211–214. doi: 10.1145/1178782.1178814.
- [114] H. zhang and D. xu, “Fusing Color and Texture Features for Background Model,” in *Fuzzy Systems and Knowledge Discovery*, L. Wang, L. Jiao, G. Shi, X. Li, and J. Liu, Eds., in Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2006, pp. 887–893. doi: 10.1007/11881599_110.
- [115] F. El Baf, T. Bouwmans, and B. Vachon, “Fuzzy integral for moving object detection,” in *2008 IEEE International Conference on Fuzzy Systems (IEEE World Congress on Computational Intelligence)*, Jun. 2008, pp. 1729–1736. doi: 10.1109/FUZZY.2008.4630604.
- [116] M. H. Sigari, N. Mozayani, and H. R. Pourreza, *Fuzzy Running Average and Fuzzy Background Subtraction: Concepts and Application*. International Journal of Computer Science and Network Security, VOL.8 No.2, February 2008
- [117] Y. Benezeth, P. M. Jodoin, B. Emile, H. Laurent, and C. Rosenberger, “Review and Evaluation of Commonly-Implemented Background Subtraction Algorithms,” in *2008 19th International Conference on Pattern Recognition*, Dec. 2008, pp. 1–4. doi: 10.1109/ICPR.2008.4760998.
- [118] C. Stauffer and W. E. L. Grimson, “Adaptive background mixture models for real-time tracking,” in *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, Jun. 1999, pp. 246–252 Vol. 2. doi: 10.1109/CVPR.1999.784637.
- [119] P. KaewTraKulPong and R. Bowden, “An Improved Adaptive Background Mixture Model for Real-time Tracking with Shadow Detection,” in *Video-Based Surveillance Systems: Computer Vision and Distributed Processing*, P. Remagnino, G. A. Jones, N. Paragios, and C. S. Regazzoni, Eds., Boston, MA: Springer US, 2002, pp. 135–144. doi: 10.1007/978-1-4615-0913-4_11.
- [120] T. Bouwmans, F. El Baf, and B. Vachon, “Background Modeling using Mixture of Gaussians for Foreground Detection - A Survey,” *Recent Patents on Computer Science*, vol. 1, no. 3, pp. 219–237, Nov. 2008.
- [121] F. El Baf, T. Bouwmans, and B. Vachon, “Type-2 Fuzzy Mixture of Gaussians Model: Application to Background Modeling,” in *Advances in Visual Computing*, G. Bebis, R. Boyle, B. Parvin, D. Koracin, P. Remagnino, F. Porikli, J. Peters, J. Klosowski, L. Arns, Y. K. Chun, T.-M. Rhyne, and L. Monroe, Eds., in Lecture Notes in Computer Science. Berlin, Heidelberg: Springer, 2008, pp. 772–781. doi: 10.1007/978-3-540-89639-5_74.
- [122] F. E. Baf, T. Bouwmans, and B. Vachon, “Fuzzy statistical modeling of dynamic backgrounds for moving object detection in infrared videos,” in *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Jun. 2009, pp. 60–65. doi: 10.1109/CVPRW.2009.5204109.
- [123] Z. Zhao, T. Bouwmans, X. Zhang, and Y. Fang, “A Fuzzy Background Modeling Approach for Motion Detection in Dynamic Backgrounds,” in *Multimedia and Signal Processing*, F. L. Wang, J. Lei, R. W. H. Lau, and J. Zhang, Eds., in Communications in Computer and Information Science. Berlin, Heidelberg: Springer, 2012, pp. 177–185. doi: 10.1007/978-3-642-35286-7_23.
- [124] J. Yao and J.-M. Odobez, “Multi-Layer Background Subtraction Based on Color and Texture,” in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2007, pp. 1–8. doi: 10.1109/CVPR.2007.383497.
- [125] M. Hofmann, P. Tiefenbacher, and G. Rigoll, “Background segmentation with feedback: The Pixel-Based Adaptive Segmenter,” in *2012 IEEE Computer Society*

- Conference on Computer Vision and Pattern Recognition Workshops*, Jun. 2012, pp. 38–43. doi: 10.1109/CVPRW.2012.6238925.
- [126] A. B. Godbehere, A. Matsukawa, and K. Goldberg, “Visual tracking of human visitors under variable-lighting conditions for a responsive audio art installation,” in *2012 American Control Conference (ACC)*, Jun. 2012, pp. 4305–4312. doi: 10.1109/ACC.2012.6315174.
- [127] Y. Goyat, T. Chateau, L. Malaterre, and L. Trassoudaine, “Vehicle trajectories evaluation by static video sensors,” in *2006 IEEE Intelligent Transportation Systems Conference*, Sep. 2006, pp. 864–869. doi: 10.1109/ITSC.2006.1706852.
- [128] N. M. Oliver, B. Rosario, and A. P. Pentland, “A Bayesian computer vision system for modeling human interactions,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 831–843, 2000, doi: 10.1109/34.868684.