

وزارة التعليم العالي و البحث العلمي

Université Badji Mokhtar –Annaba-

Badji Mokhtar –Annaba- University



جامعة باجي مختار- عنابة-

Année: 2015

Faculté des sciences de l'ingénieur
Département d'informatique

THESE

Pour obtenir le diplôme de
Docteur 3^{ème} cycle

**Traitement décentralisé des informations visuelles dans les
réseaux de capteurs sans fil multimédia**

Filière : Informatique

Spécialité : Reconnaissance de Formes et Intelligence Artificielle

Par : Fouad Bousetouane

Directrice de thèse : Dib Lynda

Professeur

Université d'ANNABA

Devant le jury

PRESIDENT : Labiba Souici

Professeur

Université d'ANNABA

EXAMINATEURS :

Cina Motamed

Professeur

Université du Littoral Cote
d'Opale (France)

Saddek Bensalem

Professeur

Université Joseph Fourier
Grenoble (France)

Noureddine Doghmane

Professeur

Université d'ANNABA

Imed Bouchrika

MCA

Université de Souk Ahras

Remerciements

Le plus court chemin d'une personne à une autre... c'est un brin de gentillesse.
-Anonyme-

Je remercie les membres de mon jury, Pr. Cina Motamed, Pr. Saddek Bensalem, Pr. Nouredine Doghmane et Dr. Imed Bouchrika en tant qu'examineurs et Pr. Souici-Meslati labiba en tant que présidente d'avoir acceptés d'examiner cette thèse.

Je souhaite également remercier mes anciens collègues et amis Abdullah, Ezedine, Adel, Said, Youcef, Ishak, Kamel, Ahmed, Ilyes, Yanis, Karime, Djeloul, Reda, Franck, Marine, Claire, Monika, pour leurs soutiens incondtionnels et leurs encouragements.

Enfin, je remercie bien sûr ceux qui ont fait de moi ce que je suis, mes frères et sœurs, et mes parents Chabaan et Fatiha, qui ont toujours su me soutenir et me donner beaucoup d'amour, et à qui je dédie cette thèse.

Thoughts :

Intelligents Machines Vs Human Being!!! The Hidden Revolt

Ambiguity, selfishness, limitation, imperfection, etc., of the human being are the strength of robots and intelligents machines. We should change our attitudes and behaviours before the revolt of the dreamer!! (as described in the famous book of Isaac Asimov "the dreamer"). But!! how we can change our attitudes ? sir, we have very important and amazing meta-rules defined by the creator of this weak world. These meta-rules can guide us to the success key of the both worlds !.

The important question (still hidden) we should ask today in the new unjust and material world is : are scholars really looking for the set of tools to make machines intelligents?, or they looking for the inverse? **Embed stupid machines into each intelligent human!!!!**.

By *Fouad Bousetouane*

Verse 135 of Surah Al Nisa

O you who have believed, be persistently standing firm in justice, witnesses for Allah , even if it be against yourselves or parents and relatives. Whether one is rich or poor, Allah is more worthy of both. So follow not [personal] inclination, lest you not be just. And if you distort [your testimony] or refuse [to give it], then indeed Allah is ever, with what you do, Acquainted.

Résumé :

LES progrès technologiques dans la conception des capteurs, en communication et en vision artificielle, permettent de favoriser le développement des nouvelles applications qui vont transformer les systèmes de vision classique en réseaux des caméras intelligentes omniprésentes. Les applications permises par les réseaux multi-caméra incluent les maisons intelligentes, la surveillance, la navigation robotique visuelle, la réalité virtuelle, etc. Cependant, l'analyse de l'information visuelle à partir de plusieurs flux vidéo dans des scénarios de la vie réelle est un problème très complexe, qui nécessite le développement des nouveaux algorithmes et des méthodes sophistiquées d'inférence collaborative capable d'analyser l'information visuelle répartie ou décentralisée en temps réel. Dans un réseau multi-caméra trois types de configuration par rapport aux champs de captures sont possibles : (1) Caméras avec chevauchement dans les champs de captures. (2) Caméras sans chevauchement dans les champs de captures ou caméras à champs disjoints. (3) Caméras hybrides. Les contraintes matérielles et économiques limitent en général le nombre des caméras et empêchent une couverture totale de l'espace, ce qui engendre des discontinuités dans le champ de vue du réseau. Un des principaux défis des réseaux multi-caméra à champs disjoints est la ré-identification et le suivi des cibles traversant le réseau : lorsqu'un objet apparaît dans le champ de vision d'une caméra, il s'agit de déterminer si cet objet a déjà été observé et suivi par l'une des caméras du réseau. Dans ce travail de thèse, nous nous intéressons spécifiquement à la ré-identification et le suivi d'objets en mouvement dans un réseau multi-caméra à champs disjoints sans fil. Dans la littérature, plusieurs approches ont été proposées pour la détection et le suivi d'objets en mouvement dans un système mono-caméra. En revanche, ces approches ne sont pas suffisamment robustes dans des conditions réelles d'application. Cette insuffisance pose un grand problème, en mesurant que ces approches forment les briques de base d'un suivi dans un réseau multi-caméra. La première partie de cette thèse est consacrée à l'amélioration des approches utilisées pour la détection, la représentation et le suivi d'objets en mouvement dans un système mono-caméra afin de les rendre plus robustes en présence des conditions complexes. Dans la deuxième partie, nous proposons un système totalement décentralisé pour la ré-identification et le suivi d'objets en mouvement dans un réseau multi-caméra ad-hoc à champs de vision disjoints sans fil. Ce système repose sur le paradigme multi-agent où chaque caméra est modélisée par un agent intelligent (agent cognitif), autonome dans sa décision locale et dans ses traitements. Les caméras intelligentes du réseau collaborent mutuellement afin

de parvenir à un consensus sur les identités des cibles qui traversent le réseau multi-caméra. Chaque agent incorpore cinq modules essentiels : (1) La détection d'objets en mouvement basée sur un mélange gaussiens. (2) Adaptation de la couleur entre chaque paire des caméras basée sur la fonction de transfert de luminosité moyenne (MBTF). (3) Extraction de caractéristiques et représentation des cibles. (4) La ré-identification collaborative d'objets inter-caméra basée sur un nouvel algorithme que nous nous proposons dans cette thèse (Consensus Intelligent). (5) le suivi intra-caméra basé sur un nouveau filtre particulière à état mixte adaptatif. Une mixture des descripteurs locaux est aussi proposée pour une représentation invariante intra/inter-caméra et discriminante inter-objet, et qui se base sur des moments statistiques d'ordre 1 et sur l'analyse de la distribution spatiale et la cooccurrence dans les régions d'intérêts. Dans le système qu'on propose, aucune hypothèse n'est imposée sur la topologie du réseau et sur le nombre des nœuds, aucune calibration n'est nécessaire pour les caméras. Notre système est invariant à l'échelle, robuste au changement des conditions d'éclairage, aux occultations, aux changements d'échelle et aux mouvements non-rigide des cibles. La nature distribuée du système que nous nous proposons mène à une approche complètement répartie pour la ré-identification et le suivi multi-objet dans un réseau ad-hoc multi-caméra à champs disjoints sans fil.

Enfin, un ensemble d'expérimentations réalisées dans des conditions réelles d'applications prouvant la validité et la puissance de la solution proposée pour la détection, le suivi et la ré-identification décentralisée d'objets mobiles en temps réel dans des conditions très complexes.

Mots clés :

Détection de Mouvement, Suivi Visuel Multi-cible, Filtrage Baysien récursif, Monte Carlo, Mélange Gaussiens, Estimateur à Noyau, Inférence Répartie, Réseau Multi-caméra Ad-Hoc, Ré-identification inter-caméra, intelligence artificielle distribuée

Distributed Multi-object Tracking and re-identification across camera networks with disjoint views

Abstract :

WITH the technological advances in visual sensors design, in communication and in dynamic computer vision are stimulating the development of new applications that will transform traditional mono-camera systems into pervasive intelligent camera networks. The multi-camera networks are the basis of several applications including video surveillance, visual robot navigation, smart homes, military and scientific applications, etc. However, analyzing the distributed visual information from multiple video streams in real-life scenarios is a very complex problem. This operation requires the development of new algorithms and sophisticated techniques for collaborative inferences able to analyze in real-time the decentralized and distributed visual information. In multi-camera network, three types of configuration are possible relative to the overall views of the network : (1) Overlapping multi-camera Networks. (2) Non-overlapping multi-camera networks or a network of disjoint cameras. (3) Hybrid multi-camera networks. Material and economic constraints limiting in general the number of cameras in the network and prevent a full coverage of a large geographical area, which creates discontinuities in the field of view of the network. A major challenge in networks with disjoint cameras is inter-camera re-identification operation : when an object appears in the field of view of a camera, it comes to determine if this object has already been observed and tracked by one of the network cameras. Following the limits of centralized approaches proposed in literature for inter-camera re-identification in terms of difficulty in analyzing a huge amount of data centrally, dynamic camera network, overloaded bandwidths, etc, it is also desirable that the inter-camera re-identification mechanism be distributed. In this thesis, we provide a multi-agent system for multi-object tracking and re-identification across cameras network with disjoint views, where the smart-cameras would have to act as autonomous agents and decisions about the objects identities would have to be taken in a distributed manner. However, to be able to attribute valid identities to all objects in the area of interest, the smart-cameras should be working cooperatively with each other. A consensus-based algorithm for distributed inter-camera re-identification is proposed in this work, where the smart-cameras collaborate to reach a consensus about objects identities. Each agent is considered as autonomous relative to the local decision and data processing. Each agent has a processing unit independent of the other agents. A modular modeling of each agent is proposed, where each agent incorporates five essential modules : (1) Moving object extraction which is based on back-

ground subtraction with statistical modeling of the background followed by adaptive post processing and common region labeling step. (2) Inter-camera color adaptation based on Mean Brightness Transfer Function (MBTF) for inter-camera color mapping. (3) Features extraction and objects representation based on a combination of statistic moments and low-level contextual information computed through the co-occurrence for invariant inter-camera object description. (4) The proposed distributed inter-camera re-identification algorithm which is based on the consensus principal ensured through the collaboration between agents (smart-cameras). (5) Intra-camera tracking based on mixed state condensation for estimating the trajectory of an object after the attribution of a valid identity. Unlike many approaches proposed in literature where many restrictions are used about the topologies of the network, cameras calibration, travelling time, closed network, exit/entrance locations, off-line learning steps, etc. The proposed approach for distributed inter-camera tracking and re-identification works in most cases without any restrictions, where many realized experiments proved the robustness of the proposed approach in many complex conditions.

Keywords : Visual Tracking, Inter-camera Re-identification, Distributed Inferences, Inter-Camera Color-Callibration, Particle Filter, Consensus-Based Algorithm, Low-level Contextual Information.

تتبع موزع متعدد الكائنة وإعادة تحديد الهوية عبر الشبكات الكاميرا مع إطلالة منفصلة

ملخص

مع التقدم التكنولوجي في مجال تصميم أجهزة الاستشعار البصرية، في مجال لاتصالات و في النضر الاصطناعي تم تحفيز تطوير تطبيقات الجديد من شأنها أن تحول أنظمة أحادية الكاميرا التقليدية إلى شبكة الكاميرات الذكية . شبكات متعددة الكاميرا هي أساس العديد من التطبيقات بما في ذلك المراقبة بالفيديو، البصرية الروبوت الملاحه، والمنازل الذكية و العسكرية و التطبيقات العلمية.

مع ذلك ، فإن تجميع وتفسير المعلومات البصرية الموزعة من تيارات الفيديو متعددة في سيناريوهات من واقع الحياة هو مشكلة معقدة جدا . هذه العملية تتطلب تطوير خوارزميات جديدة وتقنيات متطورة للاستدلالات تعاونية قادرة على تحليل في الوقت الحقيقي.

العوائق المادية والاقتصادية التي تحد بشكل عام عدد من الكاميرات v في الشبكة، ومنع التغطية الكاملة لمنطقة جغرافية واسعة، مما يخلق انقطاعات في مجال رؤية الشبكة. أحد التحديات الرئيسية في شبكات منفصلة مع الكاميرات هو إعادة تحديد الهوية بين الكاميرات : عندما يظهر كائن في مجال الرؤية للكاميرا، لأنه يأتي لتحديد ما إذا كان هذا الكائن بالفعل لوحظ وتتبعها واحدة من الكاميرات الشبكية.

بعد حدود نهج مركزية الكاميرا المقترحة في الأدب إعادة تحديد من حيث صعوبة في تحليل كمية هائلة من البيانات مركزيا، شبكة كاميرا دينامية، عرض الموجات فوق طاقتها، وما إلى ذلك، بل هو أيضا من المرغوب فيه أن المشترك بين كاميرا إعادة تحديد سيتم توزيع آلي في هذه الأطروحة، نقدم نظام متعدد وكيل لتتبع متعدد الكائن وإعادة تحديد عبر شبكة الكاميرات مع إطلالة متقطعة ، حيث الكاميرات الذكية يكون بمثابة وكلاء ذاتية الحكم واتخاذ القرارات حول هوية الكائنات يجب أن تؤخذ بطريقة متوزعة.

كما نقترح خوارزمية على أساس توافق الآراء لتوزيعها بين كاميرا إعادة تحديد الهوية في هذا العمل، حيث تتعاون الذكية الكاميرات من أجل التوصل إلى توافق في الآراء حول الكائنات.

على عكس العديد من الأساليب المقترحة في الأدب حيث يتم استخدام العديد من القيود حول طبولوجيا الشبكة والكاميرات والمعايرة، ووقت السفر، أغلقت الشبكة ،مدخل المواقع، خطوات التعلم خارج الخط، الخ .

النهج المقترح لتتبع الموزعة بين الكاميرا وإعادة تحديد يعمل في معظم الحالات بدون أي قيود، حيث أثبتت العديد من التجارب أدركت مدى متانة النهج المقترح في كثير من الظروف المعقدة.

كلمات البحث

إعادة تحديد الهوية بين الكاميرا ، والاستدلالات الموزعة، معايرة اللون بين الكامير، تصفية الجسيمات ، خوارزمية على أساس توافق الآراء.

Table des matières

1	Introduction et état de l'art	17
1.1	Introduction	17
1.1.1	Historique	17
1.1.2	Applications	18
1.2	Contexte du travail	20
1.2.1	Description de la problématique	21
1.2.2	Complexité du Problème	27
1.3	Formalisation du problème et Travaux précédents	28
1.3.1	Formalisation du problème	28
1.3.2	Travaux précédents	29
1.4	Présentation du système proposé	33
1.5	Notre approche au sein de l'état de l'art	34
1.6	Liste des Contributions identifiées	35
1.7	Organisation de la thèse	36
2	Détection d'Objets en mouvement	37
2.1	Introduction	37
2.2	Etat de l'art sur la détection de mouvement	38
2.2.1	Taxinomie	38
2.2.2	Détection sans modélisation de l'arrière-plan	39
2.2.3	Détection avec modélisation de l'arrière-plan	41
2.2.4	Modélisation récursive	42
2.2.5	Modélisation non-récursive	46
2.3	Détection de l'avant-plan	47
2.4	Approche proposée avec post-traitement	48
2.5	Résultats de la détection d'objets mobiles	53
2.6	Évaluation	54
2.7	Conclusion	56
3	Extraction des caractéristiques et représentation des cibles	57
3.1	Introduction	58
3.2	Etat de l'art sur l'extraction des caractéristiques	59
3.2.1	Taxonomie	59
3.2.2	Caractéristiques de Bas-Niveau	59
3.2.3	Caractéristiques de Niveau Intermédiaire	67
3.2.4	Caractéristiques de Haut Niveau	68
3.3	Etat de l'art sur la représentation des cibles	69

3.3.1	Taxonomie	69
3.3.2	La Forme	69
3.3.3	L'apparence	70
3.4	Mesures de similarités	71
3.4.1	Distance de Bhattacharya	71
3.4.2	Divergence de Kullback-leibler	72
3.5	Modèle de représentation des cibles proposé	72
3.6	Evaluation	74
3.6.1	L'invariance inter-image et intra-caméra	76
3.6.2	Capacité discriminatoire du modèle proposé inter-Objet	76
3.6.3	L'invariance aux conditions d'éclairages inter-Camera .	76
3.6.4	L'invariance aux autres modalités de captures	77
3.7	Conclusion	80
4	Suivi d'objets Mobiles intra-caméra	81
4.1	Introduction	82
4.2	Etat de l'art sur le suivi d'objets mobiles	83
4.2.1	Taxonomie	83
4.2.2	Méthodes déterministes d'estimation des positions . . .	83
4.2.3	Méthodes stochastiques : Filtrage récursif d'états . . .	88
4.3	Filtre particulière adaptatif proposé	93
4.3.1	Indices visuels et modèle de Vraisemblance	95
4.3.2	Suivi Multi-objet	99
4.4	Résultats et Evaluation	101
4.4.1	Résultats de suivi Mono/Multi-Objet	101
4.4.2	Evaluation	107
4.5	Conclusion	108
5	Suivi et ré-identification d'objets mobiles inter-caméra	111
5.1	Introduction	112
5.2	Description de la solution proposée	113
5.2.1	Architecture Générale	113
5.2.2	Détection d'objets mobiles	117
5.2.3	Adaptation de la Couleur inter-caméra	118
5.2.4	Extraction d'indices visuels et représentation des cibles	122
5.2.5	Ré-identification distribuée des cibles Inter-caméra . .	124
5.2.6	Suivi d'objets Mobiles Intra-Caméra	129
5.3	Expérimentations	130
5.3.1	Résultats et discussions	131
5.4	Evaluation	138
5.5	Conclusion	141

Table des matières	11
<hr/>	
6 Conclusion et perspectives	143
6.1 Synthèse des travaux effectués	143
6.2 Perspectives	144
Bibliographie	147

Table des figures

1.1	Système d'analyse des videos distribuées	25
1.2	Reconstruction 3D Multi-caméra [Andreas 2008]	25
1.3	Notre système Multi-caméra à champs disjoints sans fil	26
2.1	Classification des méthodes de détection de mouvement	39
2.2	Exemple d'estimation du flot optique, [Bruhn 2005].	41
2.3	détection de mouvement basé sur la soustraction de fond	44
2.4	Schéma Global de la chaine de traitement proposée	49
2.5	Schéma global de détection de mouvement	50
2.6	Génération de la carte de soustraction Étiquetée	53
2.7	Comparaison des résultats de détection de mouvement	55
2.8	la Courbe ROC entre les TFD et les TDC	56
3.1	Classification des techniques d'extraction des caractéristiques	60
3.2	Matrices de Co-occurrence	65
3.3	Classification des techniques de représentation des cibles	70
3.4	Les séquences d'images utilisées dans l'évaluation.	75
4.1	Taxonomie des algorithmes de suivi	84
4.2	Résultats de suivi utilisant le MeanSHift Contextuel	88
4.3	Schéma graphique de la dépendance en suivi visuel	89
4.4	Estimation de la distribution de probabilité a postériori	90
4.5	Approximation Numérique de la distribution de probabilité	93
4.6	Technique proposée pour l'adaptation de l'espace de recherche	96
4.7	Résultats de suivi sur les séquences [CAVIAR 2004]	102
4.8	Résultats de suivi sur les séquences [SPEVI 2007]	103
4.9	Comparaison entre notre algorithme et le filtre classique	104
4.10	Résultats de suivi sur les séquences [Pets 2009]	105
4.11	Résultats de suivi multi-objet	106
4.12	Erreurs dans l'estimation des positions	108
4.13	Evaluation de l'algorithme proposé APF-MC	109
5.1	Modèle Générique d'agent proposé	115
5.2	Architecture générale du système distribué proposé	116
5.3	Histogrammes de couleur	119
5.4	Observations d'objets d'intérêts	123
5.5	Fonctions de transfert de luminosité moyenne	124
5.6	Histogrammes de couleur après l'adaptation	125

5.7	Algorithme Proposé	128
5.8	Chemins des scénarios utilisés dans les expérimentations	132
5.9	Resultats de Ré-identification (1 ^{er} scénario)	134
5.10	Resultats de Ré-identification (2 ^{me} scénario)	136
5.11	Resultats de Ré-identification (3 ^{me} scénario)	137
5.12	Courbes <i>ROC</i>	140
5.13	Courbes <i>RP</i>	140
5.14	Un Faux positif (Re-identification)	141

Liste des tableaux

3.1	Erreurs dans la similarité 1	78
3.2	Erreurs dans la similarité 2	78
3.3	Invariance Inter-caméra	79
3.4	Valeurs des instances du modèle	79

Introduction et état de l'art

Sommaire

1.1	Introduction	17
1.1.1	Historique	17
1.1.2	Applications	18
1.2	Contexte du travail	20
1.2.1	Description de la problématique	21
1.2.2	Complexité du Problème	27
1.3	Formalisation du problème et Travaux précédents	28
1.3.1	Formalisation du problème	28
1.3.2	Travaux précédents	29
1.4	Présentation du système proposé	33
1.5	Notre approche au sein de l'état de l'art	34
1.6	Liste des Contributions identifiées	35
1.7	Organisation de la thèse	36

Le présent chapitre introduit le contexte général de nos travaux de recherche, ses enjeux et applications. Ensuite, un état de l'art des techniques déjà existantes dans le domaine de suivi visuel et la ré-identification d'objets en mouvement dans un réseau multi-caméra à champs disjoints est présenté. Après avoir illustré le système proposé et positionné nos travaux par rapport à la littérature, nous détaillons le plan de notre thèse.

1.1 Introduction

1.1.1 Historique

L'étude de mouvement d'objets dans l'univers n'est pas un sujet récent, cela remonte à l'antiquité. L'astronomie est l'un des premiers domaines qui s'articule autour du mouvement apparent des planètes dans le ciel. Le mouvement humain a aussi fait l'objet de plusieurs recherches en biomécanique, philosophie et en physique, mais il faut attendre jusqu'au XXème siècle où

le photographe [Muybridge 1979] a concrétisé les travaux d'analyse du mouvement humain et des animaux en utilisant la photographie. Avec l'évolution technologique et la maîtrise des techniques d'acquisition de l'information visuelle, le suivi visuel devient l'un des axes les plus actifs en vision artificielle dynamique. Les premiers travaux dans ce domaine datent des années 80s, où les premières tentatives de suivi d'objets en mouvement se basant sur des informations visuelles ont été réalisées par [Hogg 1983]. Des problèmes de plus en plus compliqués ont été abordés par la suite surtout avec l'amélioration des capteurs visuels et la capacité des processeurs, nous pouvons citer par exemple : le suivi des yeux [Liu 2005], des mains [Laptev 2006] et des visages humains [Schwerdt 2000], etc. Récemment, avec la disponibilité des capteurs couvrant une large zone géographique, le suivi d'objets en mouvement et l'analyse des contenus vidéo distribués deviennent une tâche indispensable pour des raisons sécuritaires, scientifique et militaire. Plusieurs travaux remarquables pour le suivi dans un réseau multi-caméra sont ceux réalisés à la fin des années 90s, tel que : [Jain 1995] proposent un système composé de plusieurs caméras calibrées avec chevauchement dans les champs de vision pour l'estimation de la position réelle en 3D d'une personne, dans cette même période, [Cai 1999] proposent un système multi-caméra pour la surveillance. A partir des années 2000, plusieurs investissements de la communauté ont été consacrés dans le suivi et la ré-identification d'objets dans des réseaux multi-caméra à champs disjoints tel que, les travaux de [Javed 2008a] et [Chen 2011]. Jusqu'à nos jours ce domaine reste au cœur de la recherche et en évolution remarquable, mais les problèmes de suivi et de ré-identification d'objets dans un réseau multi-caméra à champs disjoints ou inter-caméra sont d'une grande complexité, tant au niveau de leur résolution théorique, que des aspects de concrétisation et d'implémentation.

1.1.2 Applications

L'analyse des vidéos dans des systèmes multi-caméra est la base de nombreuses applications, allant de la surveillance à distance et de la robotique vers les jeux immersifs interactifs. Les systèmes de suivi visuel dans les réseaux multi-caméra à champs disjoints sont aussi utilisés pour améliorer notre compréhension sur des grandes masses de vidéo. Bien que les frontières entre les applications du suivi et la ré-identification d'objets dans les réseaux multi-caméra à champs disjoints soient en quelque sorte floues, elles peuvent être regroupées en quatre axes essentiels :

-*La Surveillance Visuelle* : la surveillance visuelle est un domaine très important, spécialement quand un grand nombre de personnes occupent un

large ou petit espace géographique comme dans les aéroports, les stations de train et les places publiques. Dans le même contexte de la surveillance, les réseaux multi-caméra sont aussi utilisés pour l'analyse du comportement des personnes, la détection des événements critiques, la navigation routière, ainsi que dans les applications militaires. La figure suivante (Fig.1.1) illustre un exemple d'une architecture générale d'un système d'analyse des vidéos distribuées dans la vidéosurveillance.

Comme illustrée dans la figure (Fig.1.1), un système automatique d'analyse des vidéos dans le domaine de la vidéo surveillance est composé de plusieurs opérations : (1) La détection d'objets en mouvement, où plusieurs techniques de traitement d'images de bas niveau sont disponibles pour l'extraction du masque de mouvement d'un objet d'intérêt. (2) Le suivi d'objets en mouvement, cette opération assure une estimation des trajectoires d'objets d'intérêts. Elle est importante pour la localisation spatiale d'objets et l'analyse des comportements. (3) La reconnaissance d'objets où des identités sont attribuées automatiquement aux objets circulant dans le réseau multi-caméra. (4) La catégorisation d'évènements et l'analyse du comportement. Ces opérations assurent une détection automatique des événements critiques comme les accidents de la route, les incendies, etc. Le travail réalisé dans le cadre de cette thèse entre dans le contexte de la vidéosurveillance.

-La réalité augmentée : Le suivi d'objets en mouvement dans un réseau multi-caméra est un élément important pour les industries du cinéma et de diffusion. L'opération de correspondance de déplacement (match moving) est l'augmentation des prises de vue réelles avec la superposition des objets graphiques animés et des effets spéciaux restitués dans des films. Pour ajouter ces nouveaux éléments graphiques dans les séquences d'images composant les vidéos, la procédure d'intégration nécessite des informations en 3D sur la scène traitée. Ces informations peuvent être estimées automatiquement en utilisant des algorithmes de suivi d'objet en mouvement, qui calculent au fil du temps la position des caméras, l'orientation, la longueur focale, la trajectoire, etc. Les informations en 3D estimées proviennent de l'analyse des trajectoires d'objets dans un grand réseau multi-caméra avec ou sans chevauchement dans les champs de captures. Un exemple d'une telle application est illustrée dans la figure (Fig.1.2).

-Application Médicale et Biologique : Les méthodes de capture des informations en 3D présentées précédemment sont également utilisées pour l'analyse de mouvement humain, par exemple : l'amélioration des performances des athlètes, l'analyse de la marche des patients dans le domaine médicale. En Biologie, des algorithmes de suivi de mouvement sont largement utilisés à partir des séquences mono/multi-capteur dans le but de comprendre le comportement et l'état des cellules, des bactéries ou encore des virus.

-*Robotique et Véhicules intelligents* : Une autre application de suivi dans un réseau multi-caméra est la robotique. La technologie de la robotique incorpore le développement des robots humanoïde, la navigation visuelle collaborative, les véhicules aériens télé-pilotés, etc. La vision intelligente à partir d'une caméra ou d'un réseau multi-caméra embarqué dans des plateformes robotiques ou des véhicules permettent à ces derniers d'interagir ou de naviguer sur un environnement bien déterminé.

1.2 Contexte du travail

Ce travail a été réalisé dans les contextes de la vision artificielle, l'intelligence distribuée et le traitement d'image avec application dans la vidéo surveillance multi-caméra. La vidéosurveillance est un système composé de plusieurs caméras disposées dans un espace déterminé afin de contrôler les activités et le comportement d'objets ou d'autres informations dynamiques. Étant donné que la capacité visuelle de l'être humain est limitée, alors l'opérateur humain ne peut à la fois surveiller activement un grand nombre de caméras. Des problèmes peuvent également se produire spécialement dans un réseau à grand nombre de caméras couvrant une large zone géographique. Dans ce cas, plusieurs événements se produisent simultanément sur plusieurs caméras, cependant l'attention de l'opérateur reste concentrée sur une seule caméra. Se basant sur toutes ces contraintes, l'automatisation du processus de la surveillance à partir de plusieurs flux vidéo devient une nécessité indispensable. Les contraintes matérielles/économiques limitent en général le nombre de caméras et empêchent une couverture totale de l'espace, ce qui engendre des discontinuités dans le champ de vue du réseau. Le suivi d'objets dans ce réseau à champs disjoints est un problème très compliqué, car les observations des objets sont souvent largement séparées en temps et/ou en espace. Dans ce travail de thèse nous nous intéressons spécialement au problème de suivi et de ré-identification d'objets inter-caméra dans un réseau multi-caméra à champs disjoints. Certains chercheurs définissent le suivi d'objets dans un réseau multi-caméra à champs disjoints comme étant un problème de mise en correspondance des indices visuels d'objets inter-caméra [Javed 2008a]. D'autres chercheurs distinguent entre les deux opérations : la ré-identification et le suivi d'objets intra-caméra [Farenzena 2010]. Dans cette thèse, nous optons pour la deuxième définition où le suivi inter-caméra est décomposé en deux opérations consécutives : la ré-identification inter-caméra basée sur une mixture des descripteurs locaux et globaux et l'estimation des positions d'objets d'intérêts intra-caméra sans fil.

1.2.1 Description de la problématique

Le problème d'observation d'objets par une caméra et de son identification par une autre dans un réseau de caméras à champs disjoints est défini comme une opération de ré-identification ou reconnaissance ou encore comme une mise en correspondance inter-caméra. Le problème de suivi et de ré-identification d'objets dans un contexte décentralisé peut être décomposé en trois niveaux :

1. *Bas-Niveau* : Dans ce niveau, les traitements d'images de bas niveaux sont nécessaires. Ces traitements assurent l'extraction des caractéristiques des cibles et l'adaptation de la couleur inter-caméra, afin de créer des modèles d'observations d'objets invariants par rapport aux changements des conditions d'éclairage et de captures inter-caméra.

2. *Niveau-Intermédiaire* : Dans ce niveau, les algorithmes de détection et de suivi intra-caméra d'objets en mouvement doivent être testés pour les adapter avec les réseaux multi-caméra et spécialement dans un contexte distribué.

3. *Haut-Niveau* : Dans ce niveau, une opération de ré-identification inter-caméra est nécessaire afin de reconnaître les objets, où les caméras collaborent afin de parvenir à un consensus sur l'identité d'un objet déjà suivi par une autre caméra du réseau.

Dans cette thèse, nous considérons un réseau composé de plusieurs caméras à champs disjoints sans fil. Nous supposons que le réseau est totalement dynamique où les positions des caméras peuvent être changées et ou d'autres caméras peuvent être ajoutées à n'importe quel moment. Aucune unité centrale de traitement et de calcul commune n'est utilisée. Chaque caméra est dotée d'une unité de calcul séparée des autres caméras. La figure (Fig.1.3) illustre le réseau multi-caméra Ad-hoc à champs disjoints sans fil réalisé au département d'informatique de l'université Badji Mokhtar.

Dans ce genre de réseau multi-caméra à champs disjoints plusieurs scénarios sont possibles mais la solution que nous proposons dans cette thèse satisfait tous ces cas :

1. Si un objet entre dans le champ de couverture d'une caméra donnée (eg. E-CAM-1) alors il doit être détecté.
2. Au moment de sa détection il doit avoir une identité valide :
 - a. Si cet objet avec les caractéristiques visuelles e_i , a été déjà identifié dans une autre caméra du réseau, cette identité doit être la même quelle que soit la différence en temps de capture du même objet dans les deux caméras.
 - b. Les caractéristiques e_i de chaque observation générée par un objet et capturées par une caméra doivent être invariantes par rapport aux changements des conditions d'éclairage inter-caméra et discriminantes par rapport aux autres objets.

c. Si cet objet n'a jamais été détecté dans les caméras du réseau alors une identité lui sera affecté. Cette dernière doit être unique, non existante dans le réseau.

d. Aucune unité centrale ne doit être utilisée, la ré-identification d'objets inter-caméra doit être totalement distribuée. De ce fait, de nouveaux algorithmes intelligents sont proposés, dans cette thèse, favorisant la collaboration entre les caméras du réseau.

3. Après identification et ré-identification d'un objet Obj_x avec une identité valide, par une caméra donnée (eg. E-CAM-1) alors cette dernière doit estimer les positions spatiales (trajectoires) en $2D$ de cet objet en se basant sur les mêmes caractéristiques e_i (suivi intra-caméra).

4. La détection, le suivi et la ré-identification d'objets mobiles doivent être en temps-réel.

Pour réaliser la détection, le suivi et la ré-identification d'objets en mouvement dans ces conditions, nous devons améliorer et adapter d'abord les techniques utilisées dans un système mono-caméra aux réseaux multi-caméra. Les thématiques que nous adressons à travers cette thèse sont les suivantes :

1. *Réseau multi-caméra Ad-Hoc* : l'analyse et l'interprétation de plusieurs flux vidéo distribués dans les réseaux multi-caméra ad-hoc sont des problèmes très complexes, non seulement par rapport à la dynamique de réseau mais aussi par rapport à la limite de la bande passante. L'analyse des vidéos distribuées dans un tel réseau exige la création de nouvelles techniques de compression local de l'information visuelles afin d'assurer une transmission à temps réel inter-caméra et diminuer les surcharges des canaux de transmission sans fil. Une partie de cette thèse à été consacrée au codage de l'apparence d'objets d'intérêts dans le but de créer des modèles d'observation visuelles simple en terme de dimensionnalité et riches en représentation des cibles.

2. *Réseau multi-caméra à champs disjoints* : Comme illustré dans la figure (Fig.1.3), un réseau multi-caméra à champs disjoints est un réseau sans chevauchement dans les champs de vision des caméras. Une discontinuité ou une zone noire existe entre les champs de capture des caméras, ce qui rend les opérations de ré-identification et de suivi inter-caméra un grand défi.

3. *Détection d'objets en mouvement* : La détection d'objets en mouvement ou la segmentation spatio-temporelle est une opération très importante pour l'extraction et le comptage d'objets d'intérêts, le calcul des descripteurs locaux ainsi que la ré-identification et l'initialisation des algorithmes de suivi. Dans ce travail, une nouvelle chaine de traitement est proposée pour la détection d'objets en mouvement. Cette chaine repose sur la modélisation statistique de l'arrière-plan par un mélange gaussien. Un post-traitement adaptatif et une opération de regroupement des régions connexes sont utilisées pour la génération de la carte de soustraction binarisée et étiquetée [Bousetouane 2011].

4. *Modélisation de l'apparence des cibles* : Cette étape est très importante pour la ré-identification et le suivi d'objets en mouvement inter-caméra. Dans un réseau multi-caméra où les conditions d'éclairage sont différentes d'une caméra à une autre, l'apparence d'un même objet d'intérêt est forcément différente d'une caméra à une autre (voir la figure 1.3), ce qui nécessite la proposition des modèles d'apparence des cibles invariants. Dans ce travail, une nouvelle idée d'exploitation du contexte spatial calculé à travers les caractéristiques de texture de Haralick est proposée [Bousetouane 2013b] pour la représentation des cibles. Une mixture ad-hoc des caractéristiques est aussi proposée pour une représentation invariante inter-image et discriminante inter-objet, afin d'améliorer le suivi intra-caméra [Bousetouane 2012].

5. *Adaptation de la couleur inter-caméra* : Tant que les caméras sont dans des endroits géographiquement différents, alors il y'aura certainement une grande variation en captures, poses, couleur, etc. La plupart des descripteurs d'apparence sont calculés soit à partir des histogrammes de couleur ou du niveau de gris, alors il est important d'adapter la couleur entre l'ensemble des caméras avant le calcul des descripteurs. Dans ce travail, une technique d'adaptation de la couleur entre chaque paire des caméras est adoptée qui se base sur la fonction de transfert de la luminosité moyenne (MBTF) proposée dans [Javed 2008a].

6. *Ré-identification d'objets inter-caméra* : Dans cette opération, une unique et valide identité est attribuée à chaque objet traversant le réseau. Dans la littérature, la plus part des approches proposées pour la ré-identification d'objet inter-caméra sont centralisés. Ces approches sont coûteuses en calcul et en gestion, vu qu'elles nécessitent une unité centrale consultée par l'ensemble des caméras composant le réseau à l'instant t . Dans cette thèse, sans utilisation d'unité centrale de traitement, un nouveau algorithme est proposé pour la ré-identification d'objets inter-caméra. L'algorithme proposé est totalement décentralisé, où un ensemble de règles de collaboration sont définis entre l'ensemble des caméras du réseau afin de parvenir à un consensus sur l'identité d'un ou plusieurs objets circulant dans le réseau.

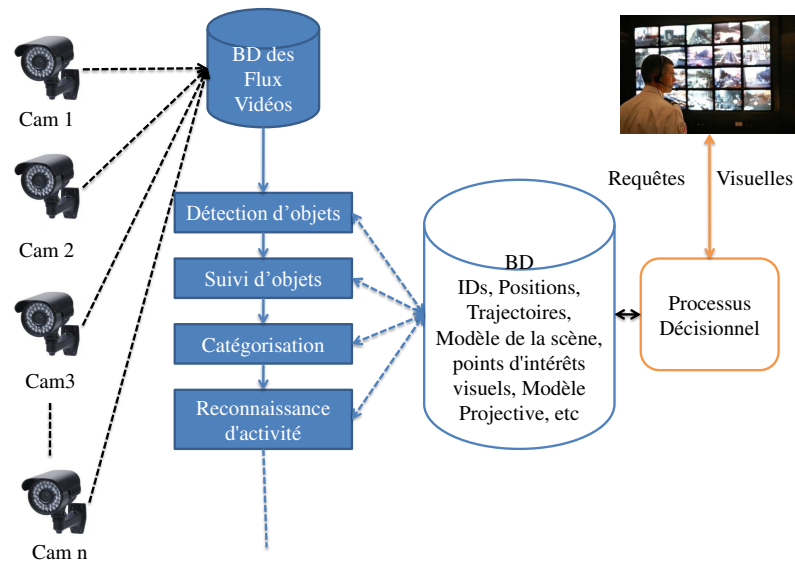
7. *Suivi Intra-Caméra* : Cette opération permet de suivre plusieurs objets en temps réel à travers l'estimation des positions d'objets en 2D au fil du temps. Dans un contexte multi-caméra les algorithmes classiques de suivi utilisés dans un système mono-caméra doivent être améliorés et adaptés à ce problème. Dans cette thèse, un filtre particulière adaptatif à états mixte est proposé. Il se base sur l'intégration de multiples indices visuelles pour l'estimation des positions d'un ou plusieurs objets intra-caméra. L'approche proposée pour le suivi d'objets intra-caméra utilise explicitement les identités attribuées à ces objets après l'étape de la ré-identification. Plusieurs instances du filtre particulière proposé sont utilisées pour le suivi multi-objet, où une

instance est utilisée par objet. .

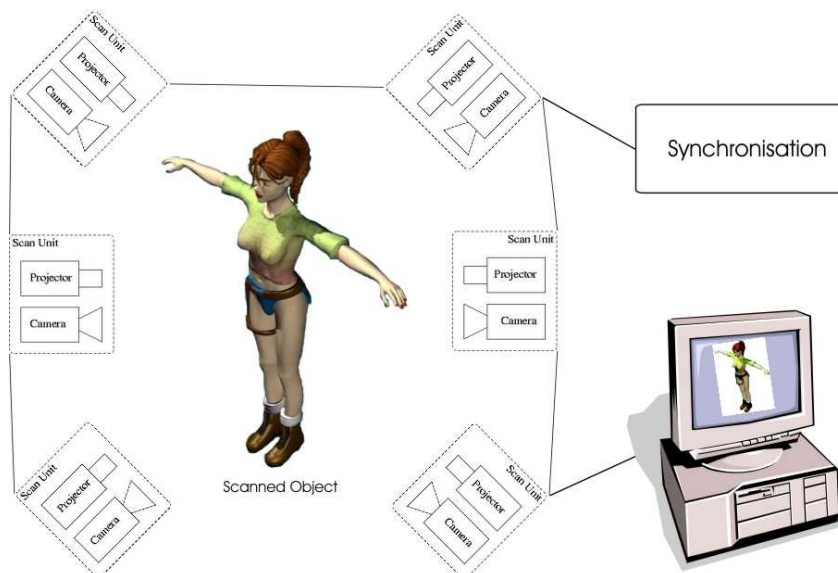
8. *Performance en Temps Réel* : Dans plusieurs applications de suivi il est nécessaire de suivre les objets d'intérêts en temps réel. Le suivi visuel d'objets en mouvement dans un réseau multi-caméra à champs disjoints est une opération très compliquée, qui inclut plusieurs processus comme : la détection d'objets en mouvement, la reconnaissance, la détection d'occultations, l'extraction des descripteurs, la représentation des cibles, etc. La complexité en calcul et en résolution de chaque processus font du suivi inter-caméra en temps réel un grand défi, qui nécessite l'utilisation des simples algorithmes avec beaucoup d'efficacités. Il y' a toujours un compromis entre la performance en temps réel, le cout du système (réseau multi-caméra) et la précision des algorithmes de suivi et de ré-identification.

9. *Traitements distribués* : Contrairement aux différentes approches centralisées proposées dans ce contexte jusqu'à présent, l'approche que nous proposons dans cette thèse est totalement distribuée, où le raisonnement réparti est favorisé. A travers cette approche distribuée, la performance en temps réel du système est remarquablement augmentée.

10. *Inférence répartie et intelligence artificielle distribuée* : La résolution des problèmes de nature distribuée, comme le problème abordé dans cette thèse, est un sous-domaine de l'intelligence artificielle distribuée où l'accent est mis sur la collaboration entre un ensemble d'agents pour résoudre des problèmes nécessitant un effort collectif. En raison d'une répartition des ressources inhérentes, telles que les connaissances, la capacité, l'information et l'expertise entre les agents, un agent dans un système distribué n'est pas en mesure d'accomplir ses propres tâches seul, où au moins peut mieux accomplir ses tâches lorsqu'il travaille avec d'autres agents (plus rapide, précis, ou certainement).



1.1: Architecture générale d'un système automatique d'analyse des vidéos distribuées en vidéosurveillance.



1.2: Un Système Multi-caméra pour la reconstruction 3D [Andreas 2008]



1.3: Le réseau multi-caméra a champs disjoints sans fil composé de sept caméras, monté dans le cadre de cette thèse.

1.2.2 Complexité du Problème

Le grand défi de suivi d'objets dans un réseau multi-caméra à champs disjoints dans un contexte distribué réside dans la mise en correspondance inter-caméra des observations d'objets et le suivi de ces objets intra-caméra. Les difficultés sont causées par :

- La Complexité de positionnement et la calibration de l'ensemble des caméras.
- Le paramétrage intrinsèque et extrinsèque des caméras est différent d'une caméra à une autre surtout dans les réseaux multi-caméra hétérogène.
- Les observations d'un objet captuées par les caméras sont largement séparées en temps et en espace.
- Les cibles disparaissent dans tous les points de vue des caméras pendant un moment (zones noire ou discontinuité dans les champs de vision des caméras).
- Le changement dans la luminosité entre l'ensemble des caméras où chaque caméra est positionnée dans un endroit différent avec des conditions différentes.
- La signature visuelle générée par un objet et capturée par une caméra est totalement différente de la signature visuelle du même objet dans une autre caméra du réseau.
- Absence d'une unité centrale de traitement ou de collecte des données.
- Le suivi mono-objet est considéré comme un problème complexe, et les algorithmes proposés jusqu'à maintenant sont insuffisants dans des conditions complexes.
- La détection d'objet en mouvement est aussi un problème au cœur de la recherche, aucune approche proposée jusqu'à présent n'est valide dans tous les cas spécialement si l'arrière-plan est dynamique.
- La complexité de calcul des algorithmes en traitement d'image et en vision artificielle rend la performance en temps réel un grand défi.
- Absence des connaissances préalables sur la topologie du réseau, le nombre des nœuds (caméras), informations sur les cibles, le traitement est en ligne (on-line).

Vis-à-vis de toutes ces complications, l'opération de ré-identification et le suivi d'objets en mouvement dans un réseau ad-hoc multi-caméra à champs disjoints est un grand challenge. A travers ce travail de thèse nous proposons une solution robuste qui pourrait intégrer l'ensemble des aléas citées précédemment.

1.3 Formalisation du problème et Travaux précédents

1.3.1 Formalisation du problème

La formalisation du problème de suivi d'objets dans un réseau multi-caméra à champs disjoints décrite dans cette section a été proposée par [Javed 2008a]. Supposant qu'on a un système ad-hoc multi-caméra S , composé de N caméras $Cam_1, Cam_2, \dots, Cam_N$ avec un champ de vision sans chevauchement. Supposant qu'il y'a k Objets en mouvement dans le réseau p_1, p_2, \dots, p_k . Soit O l'ensemble d'observations, $O_{Cam_j} = O_{Cam_j^1}, O_{Cam_j^2}, \dots,$

$O_{Cam_j^k}$, où $O_{Cam_j^1}$ est l'observation générée par la cible p_1 et observée par la caméra Cam_j . Soit $O_{Cam_i^a}$ l'observation d'un objet quelconque qui sort du champ de vision de la caméra Cam_i et entre dans le champ de vision de la caméra Cam_j avec une nouvelle observation $O_{Cam_j^b}$, sous de telles hypothèses, le problème de suivi inter-caméra est défini comme une opération de mise en correspondance entre l'observation $O_{Cam_i^a}$ et l'observation $O_{Cam_j^b}$. Si une grande similarité est calculée entre les deux observations $O_{Cam_i^a}$ et $O_{Cam_j^b}$ alors l'observation correspond au même objet p_k . Ce processus est nommé par d'autres chercheurs un processus de ré-identification. Dans un contexte probabiliste, la probabilité que l'observation $O_{Cam_i^a}$ observée par la caméra Cam_i correspond à l'observation $O_{Cam_j^b}$ observée par la caméra Cam_j peut être décrite par : $P(a = b | O_{Cam_i^a}, O_{Cam_j^b})$. La correspondance la plus probable doit maximiser la similarité entre les deux observations :

$$K = ArgMax(P(a = b | O_{Cam_i^a}, O_{Cam_j^b})) \quad (1.1)$$

$K \in L_s$, L_s est l'ensemble d'objets dans la liste de transfert (candidats).

L'observation d'un objet peut être basée sur des descripteurs locaux ou globaux : Si le suivi inter-caméra est basé sur des descripteurs locaux alors cette opération est nommée d'après plusieurs chercheurs une opération de ré-identification inter-caméra. Si le suivi inter-caméra se base sur des descripteurs globaux alors le suivi inter-caméra est défini par d'autres chercheurs comme une opération de mise en correspondance inter-caméra. Jusqu'à présent plusieurs descripteurs ont été proposés dans l'état de l'art, par exemple : couleurs, texture, mouvement, spatio-temporelle, etc. L'observation globale d'un objet p_k basée sur ces descripteurs capturés par la caméra Cam_j peut être décrite par :

$$O_{Cam_j^a} = O_{Cam_j^a}(couleur) \wedge O_{Cam_j^a}(Texture) \wedge O_{Cam_j^a}(Spat - temp) \wedge, \dots \quad (1.2)$$

L'observation générale d'un objet p_k basée sur ces descripteurs capturée par la caméra Cam_i peut être décrite par :

$$O_{Cam_i}^b = O_{Cam_i}^b(\text{couleur}) \wedge O_{Cam_i}^b(\text{Texture}) \wedge O_{Cam_i}^b(\text{Spat} - \text{temp}) \wedge, \dots \quad (1.3)$$

Si nous décomposons les descripteurs en deux catégories : descripteurs d'apparence d'objet p_k et descripteurs spatio-temporelle. Alors l'espace de solution où la correspondance générale la plus probable qui maximise la similarité entre les deux observations a et b est décrit par :

$$K' = ArgMax \prod (P(a = b | O_{Cam_i}^a(App), O_{Cam_j}^b(App)) \\ (P(a=b | O_{Cam_i}^a(Spat - Temp), O_{Cam_j}^b(Spat - Temp)))) \quad (1.4)$$

K' est équivalent de la maximisation du terme k'' :

$$K'' = ArgMax \sum (\ln(P(a = b | O_{Cam_i}^a(App), O_{Cam_j}^b(App))) \\ (\ln P(a = b | O_{Cam_i}^a(Spat - Temp), O_{Cam_j}^b(Spat - Temp)))) \quad (1.5)$$

1.3.2 Travaux précédents

Pour estimer l'espace de solution K'' ou la maximisation de la similarité entre les deux observations a et b (la mise en correspondance inter-caméra), plusieurs approches ont été proposées ; allant de l'utilisation des descripteurs globaux (la mise en correspondance inter-caméra) vers l'utilisation des descripteurs locaux (la ré-identification inter-caméra). Dans cette partie nous présentons en détail l'état de l'art des approches utilisées jusqu'à présent pour la mise en correspondance et la ré-identification d'objets inter-caméra.

1.3.2.1 La mise en correspondance inter-caméra

Récemment, Il y a eu des remarquables travaux dans la mise en correspondance inter-caméra basant sur des descripteurs globaux. [Collins 2001] ont développés un système composé de plusieurs caméras calibrées avec un modèle de l'espace utilisé, où une corrélation d'objets détectés et leurs positions en 3D ont été utilisées pour le suivi d'objets. [Huang 1997], proposent une approche probabiliste pour le suivi des véhicules dans deux caméras calibrées à champs disjoints. Cette approche ne fonctionne que dans des conditions spécifiques d'application : les véhicules circulent dans une direction particulière, les véhicules doivent être dans une voie d'une route déterminée, la calibration préalable des caméras est nécessaire, la couleur est modélisée par la moyenne

de la couleur de tous les objets, etc. Ensuite, [Pasula 1999] ont approfondi cette approche en une approche multi-caméra en proposant un algorithme d'approximation polynomial du temps de transfert basant sur une simulation Monte-Carlo fondée sur les chaînes de Markov (MCMC). [Kettner 1999], utilisent une formulation bayésienne du problème de la mise en correspondance inter-caméra, où les liens de circulation inter-caméra ont été reconstruits et l'apparence d'objets a été modélisée comme étant un histogramme de couleur. Le système proposé dans ce travail se base sur l'utilisation d'une topologie du réseau déterminée, où tous les liens possibles de transfert d'objets inter-caméra sont modélisés et manuellement initialisable. [Porikli 2003a], proposent un modèle de calibration de la couleur inter-caméra afin d'estimer la fonction d'alignement optimale entre l'histogramme d'apparence d'objets dans différentes vues. Dans le même travail, une combinaison des histogrammes avec un indice spatio-temporelle a été utilisée pour le suivi d'objets au cours du temps. [Makris 2004], proposent une approche pour la détermination de la topologie du réseau utilisée basée sur deux étapes : (1) Les zones d'entrées/sorties de chaque caméra sont déterminées. (2) Les liens entre ces zones sont trouvés utilisant la cooccurrence d'objets qui entrent et sortent de ces zones. Cette approche reste limitée dans des conditions réelles d'application où le comportement d'objets circulant dans le réseau est totalement imprédictible. Contrairement à cette approche, [Stauffer 2005], proposent une approche plus sophistiquée pour la recherche des liens inter-caméra où une hypothèse de la corrélation entre les événements entrant/sortant de ces zones peuvent ou pas contenir des transitions d'objets valide. Cette approche a été testée afin de couvrir le cas où les événements entrant/sortant sont corrélés mais pas avec des objets valides. Une extension de cette approche a été proposée par [1544940], où un multi-modèle transition d'états dans le temps a été proposé. [Rahimi 2004], proposent une méthode de reconstruction complète des liens inter-caméra d'un objet qui traverse le réseau. La dynamique de l'objet dans ce travail est modélisée par un processus markovien. Basant sur ces liens, les auteurs ont confirmé qu'étant donné la villosité et la localisation d'objets dans un plan de référence à partir des caméras composant le réseau alors la trajectoire est facilement estimée en utilisant un schéma de minimisation non-linéaire. Dans ce même contexte, [Chen 2011], ont proposé une approche basée sur l'apprentissage par lot pour la sélection automatique des liens de fortes fréquences entre les points d'entrées/sorties inter-caméra. [Porikli 2003b], proposent une fonction de transfert de luminosité (BTF) pour la mise en correspondance d'objets inter-caméra à champs disjoints. La fonction de transfert de luminosité (BTF) est calculée pour toute paire de caméras dans cette approche. Cette fonction (BTF) permet de cartographier la couleur entre chaque couple des caméras dans le réseau à travers la mise en correspondance entre

la couleur d'une observation dans une caméra avec la couleur de la même observation dans une autre caméra. Une fois la cartographie est connue entre chaque paire de caméras, alors le problème de la mise en correspondance inter-caméra revient à mettre en correspondance (en anglais, *matching*) entre les observations d'objets d'intérêts. Cette approche a fait l'objet de plusieurs travaux de recherche mais reste invalide dans un brusque changement d'éclairage dans le réseau au fil du temps. Une extension de cette approche a été proposée par [Prosser 2008], où l'histogramme moyen a été utilisé dans la cartographie ou l'adaptation de la couleur inter-caméra afin de réduire le temps de calcul. [Javed 2008a], proposent une extension de la BTF vers la mise en correspondance non seulement de la luminosité mais aussi les caractéristiques spatio-temporelle (temps de transfert, vitesse des objets, identités des caméras, etc). Dans cette approche, la topologie du réseau et les probabilités des liens d'entrées/sorties sont appris utilisent un estimateur à noyau basé sur la fenêtre de parzen avec une mise en correspondance initialement manuel. Les même auteurs [Javed 2008a], démontrent que toutes les fonctions de transferts de la luminosité (BTF) d'une caméra à une autre sont liées à un sous espace de dimension inférieure. Une analyse en composante principale stochastique (ACP) a été proposée par Javed et al [Javed 2008a], pour apprendre le sous espace de la luminosité ou la fonction de transfert de couleur est liée. [Chen 2008], proposent de mettre à jour la BTF au fil du temps en utilisant les dernières observations des cibles suivies. L'algorithme de Clustering (K-means) a été utilisé par Madden et al [Madden 2007], pour faciliter le calcul de la moyenne de la luminosité d'objets. Des techniques de consistance de la couleur ont été utilisées par [Colombo 2008] afin de compenser le changement de la luminosité. L'apprentissage incrémentale a été proposé par Gilbert et Bowden [Gilbert 2006], pour modéliser la variation de la couleur inter-caméra et la probabilité à posteriori des liens spatio-temporelle inter-caméra. Cette approche est la plus recommandée pour la mise en correspondance inter-caméra, adaptable au changement de la luminosité, par opposé elle nécessite une grande masse de données d'apprentissage. [Dick 2005], proposent l'utilisation d'une matrice de transition stochastique pour décrire la forme de mouvement d'objets afin de mettre en correspondance entre les mêmes objets inter-caméra. Parmi les approches présentées précédemment pour l'adaptation de la couleur entre les caméras composant le réseau (en anglais, *inter-camera matching*), la fonction de transfert de luminosité (BTF) avec ses variantes (MBTF, CBTF, ABTF, BBTF, etc.) est la plus utilisée jusqu'à présent. Cet axe de recherche « la mise en correspondance inter-caméra (*inter-camera Matching*) basé sur des descripteurs globaux » est trop actif jusqu'à l'heure actuel. Contrairement aux autres approches présentées précédemment où la luminosité est constante dans tous les caméras, le grand défi de la fonction de mise en correspondance

inter-caméra réside dans le changement des conditions d'éclairage au fil du temps et spécialement dans les réseaux dynamiques.

1.3.2.2 Descripteurs locaux et ré-identification d'objets

La ré-identification d'objets dans un réseau multi-caméra à champs dis-joints est une opération d'attribution d'identités valides aux objets circulant dans le réseau. Si un objet sort d'un champ de couverture d'une caméra et entre dans un autre champ de couverture d'une autre caméra, il s'agit de déterminer si c'est le même objet en se basant sur les descripteurs locaux de l'objet (comme la biométrie pour les cibles humain, le mouvement, la texture, couleur, etc). Cette orientation de la mise en correspondance inter-caméra a fait l'objet de plusieurs recherches et travaux publiés. Nous citons dans cette section les travaux récents, intéressant spécialement aux cibles humaines. [Jaffré 2004], proposent une approche basée sur la description invariante de la signature visuelle d'une cible utilisant l'histogramme de couleur de la région en dessous du visage de la cible, mais la couleur reste encore un descripteur limité dans des conditions de captures constantes. Dans ce même contexte, [Seigneur 2004], proposent une approche plus générique modélisant les vêtements des cibles par des descripteurs de couleurs. [Farenzena 2010], proposent une modélisation de l'apparence des cibles par un histogramme de couleur stochastique créée par l'utilisation d'un classifieur de type K-Plus-Proche-voisine Floue (Fuzzy-KNN) appris off-line. Cette modélisation permet de mettre en correspondance entre les deux modèles d'apparences d'une même personne dans un réseau multi-caméra. Dans [Schwartz 2009], les auteurs proposent un modèle de signature de haut-niveau composé de la texture, la couleur et le gradient des cibles. Le contexte spatial a été aussi utilisé par [Zheng 2009], et cela à fin d'enrichir la description contextuelle des cibles. Dans [Nakajima 2003], les auteurs proposent une combinaison de plusieurs descripteurs locaux et globaux extraits à partir des images d'apprentissage des cibles et appris utilisent un SVM multi-classes avec plusieurs schémas d'apprentissage. Dans [Hamdoun 2008], les auteurs proposent un schéma de ré-identification des cibles basé sur le détecteur des points d'intérêts (descripteur SURF) collectés à partir d'une petite séquence vidéo. [Lo Presti 2009], estiment dans chaque caméra du réseau un modèle d'allocation de Dirichlet basé sur l'ensemble des descripteurs extraits à partir d'une cible d'intérêt. [Truong Cong 2009], proposent une concaténation des caractéristiques d'une cible d'intérêt combinée avec un classifieur SVM pour améliorer la ré-identification dans un réseau multi-caméra. Un filtre particulaire à état mixte basé sur un ensemble d'histogramme de couleur comme un modèle d'observation a été proposé par [Meden 2011], et cela, pour la ré-identification des cibles inter-caméra et l'estimation de leurs positions en

même temps. Cette approche reste fonctionnelle dans un réseau multi-caméra fermé car les identités des cibles ont été apprises préalablement (off-line). Une stratégie d'agrégation des descripteurs locaux basée sur le descripteur de HOG (histograms of oriented gradients), la couleur et les informations structurelles a été proposée par [Doretto 2011], pour générer une signature multi-shots afin de ré-identifier les cibles inter-caméra. Plusieurs autres investissements ont eu lieu pour l'amélioration de la capacité des classifieurs, comme les travaux de [Bak 2010], où une combinaison d'un ensemble de classifieurs faibles a été utilisée et qui se base sur l'Adoboost. La ré-identification d'objets inter-caméra reste jusqu'à ce jour un axe très actif mais les défis à soulever restent dans le choix des descripteurs (invariants inter-caméra/inter-image et discriminants inter-objet) et la stratégie d'agrégation de ces descripteurs et ce la dans le but de créer des modèles d'observations informatifs.

1.4 Présentation du système proposé

Comme nous l'avons détaillé dans l'état de l'art, les approches proposées jusqu'à présent pour le suivi et la ré-identification d'objets mobiles inter-caméra restent limités dans des conditions spécifiques d'applications. La majorité des approches proposées, comme dans [Javed 2008a], [Chen 2011], se basent sur une infrastructure centralisée. La nécessité des solutions matériellement économiques, invariantes à l'échelle et capable d'analyser l'information visuelle répartie en temps réel rend les architectures centralisées des solutions insuffisantes. Le passage vers des architectures complètement décentralisées devenant une nécessité, cela dans le but de favoriser le calcul distribué et l'inférence répartie. La nature distribuée du problème de suivi et la ré-identification d'objets inter-caméra exige l'utilisation d'une architecture décentralisée où les caméras intelligentes composant le réseau collaborent entre elles afin de suivre et ré-identifier les cibles traversant le réseau, [Bousetouane 2013c]. Cette nature nous mène à modéliser le problème par un système multi-agent, où un modèle générique d'agent cognitif (caméra intelligente) est proposé dans cette thèse pour la modélisation des caméras. Physiquement, une instance du modèle proposé (agent cognitif) est utilisée par caméra. Le traitement de chaque instance (caméra intelligente) s'effectue par l'exécution de cinq modules importants : La détection d'objets en mouvement et la génération de la carte de soustraction binarisée et étiquetée (chapitre.2). Extraction des caractéristiques et représentation des cibles. Adaptation de la couleur inter-caméra basée sur la fonction de transfert de luminosité moyenne (CBTF). La ré-identification collaborative d'objets inter-caméra basée sur le nouveau algorithme proposé dans cette thèse (Consensus Intelligent). Le suivi intra-caméra (chapitre.4).

Cette structuration modulaire des différents traitements réalisés par chaque agent permet de décentraliser l'expertise de chaque instance d'agent, ce qui assure l'amélioration de la performance de chaque instance en terme de temps de calcul et de gestion. La figure (Fig.5.2) illustre l'architecture générale du système décentralisé que nous proposons dans cette thèse pour le suivi et la ré-identification des cibles inter-caméra dans un réseau multi-caméra ad-hoc à champs disjoints sans fil. La nature décentralisée du système proposé nous mène à la proposition d'une nouvelle approche totalement répartie pour le suivi et la ré-identification des cibles inter-caméra. Dans cette approche les caméras intelligentes (agents) collaborent mutuellement entre elles afin de parvenir à un consensus sur les identités des cibles traversant le réseau multi-caméra. L'approche décentralisée proposée est détaillée dans le chapitre (Chapitre.5).

1.5 Notre approche au sein de l'état de l'art

L'opération de la ré-identification d'objets dans un réseau multi-caméra à champs disjoints (suivi-inter-caméra) dans la littérature (voir la section.1.3.2) est décomposée en deux grandes orientations : (1) La mise en correspondance inter-caméra, consiste à trouver une correspondance entre deux observations capturées par deux caméras différentes. Cette opération se base sur des descripteurs globaux. (2) La ré-identification inter-caméra, cette opération se base sur des descripteurs locaux afin de trouver l'objet sortant du champ de vision d'une caméra donnée et entrant dans un nouveau champ de vision d'une autre caméra. La plus part des approches proposées jusqu'à présent restent limitées à cause des restrictions posées par les chercheurs dans le but de faciliter la nature du problème, comme : La topologie du réseau, le nombre des caméras, les réseaux de caméras fermés, la calibration, les réseaux homogènes, les plans de référence (homographie), les points d'entrées/sorties, les liens entre les caméras, l'utilisation des unités centrales pour le calcul et le traitement. Dans la littérature, aucune solution totalement décentralisée remarquable n'a été proposée. Face à toutes ces restrictions, l'application de ces approches dans des conditions réelles reste très limitée et compliquée en même temps. La nature distribuée du problème (suivi-inter-caméra) et la complexité de calcul en temps réel dans les systèmes centralisés exige la proposition de nouvelles approches plus sophistiquées compatibles avec la nature distribuée ou répartie du problème de suivi et la ré-identification d'objets inter-caméra. Dans le cadre de cette thèse nous proposons un système totalement distribué pour la détection, le suivi et la ré-identification d'objets en mouvement dans un réseau multi-caméra à champs disjoints sans fil. Le système proposé se base

sur l'inférence répartie (paradigme Multi-agent) où chaque caméra est modélisée une instance du modèle d'agent cognitif proposé. L'agent cognitif proposé incorpore cinq modules : la détection d'objet en mouvement, extraction des caractéristiques, l'adaptation de la couleur inter-caméra, l'identification/ré-identification d'objets inter/intra-caméra, le suivi intra-caméra. L'architecture décentralisée du système proposé mène à une approche totalement distribuée pour la ré-identification d'objets inter-caméra. Cette nouvelle approche assure une collaboration entre les caméras intelligentes (instances d'agent) afin de parvenir à un consensus sur les identités d'objets traversant le réseau. D'après nos connaissances, une ré-identification distribuée d'objets en mouvement basée sur le principe du consensus n'a pas été proposée auparavant dans la littérature. Le système proposé dans le cadre de cette thèse est détaillé dans le chapitre.5.

1.6 Liste des Contributions identifiées

Tous les travaux de cette thèse sont donc à raccorder directement ou indirectement aux travaux précédents pour le suivi et la ré-identification inter-caméra d'objets en mouvement dans un réseau multi-caméra à champs dis-joints. Nous identifions plusieurs contributions :

a. Un Modèle d'observation visuelle invariant inter-image et inter-caméra et discriminant inter-objets d'apparence des cibles est proposé. Le modèle proposé se base sur le contexte spatial calculé à travers la texture de Haralick combiné avec des moments statistique d'ordre 1, [Bousetouane 2013b].

b. Par rapport au suivi intra-caméra, nous proposons un filtre particulière adaptatif à état mixte. Ce filtre intègre une combinaison ad-Hoc d'indices visuels 'APF-MC' pour l'estimation des positions des cibles au fil de temps. Pour améliorer la capacité du filtre particulière classique par rapport au changement d'échelle, aux objets non-rigides et aux traitements temps réel, les états des particules dans le filtre que nous proposons sont adaptables avec les dimensions réelles des cibles suivies. Dans ce même contexte, un nouveau modèle de vraisemblance est proposé basé sur une combinaison de la couleur et la texture. Ce modèle vise à améliorer la convergence des particules afin d'éviter les minimums locaux (les positions faux positifs), [Bousetouane 2012],[Bousetouane 2013a].

c. Un nouveau algorithme est proposé pour la ré-identification distribuée d'objets inter-caméra, où il définit un ensemble de règles permettant aux caméras composant le réseau de collaborer mutuellement pour parvenir à un consensus sur les identités des cibles circulant dans le réseau, [Bousetouane 2013c].

d. Une fusion des techniques utilisées dans les deux axes (la ré-identification

inter-caméra et la mise en correspondance inter-caméra) est indirectement intégrée dans l'algorithme proposé afin de créer un modèle d'apparence des cibles invariants aux conditions d'éclairage inter-caméra.

e. La solution proposée est invariante à l'échelle et totalement dynamique par rapport au nombre des caméras, à la topologie du réseau, aux champs de captures.

1.7 Organisation de la thèse

Le manuscrit est structuré en six chapitres incluant le présent état de l'art.

Chapitre (2) : Dans ce chapitre nous présentons un état de l'art sur les techniques utilisées pour la détection d'objets en mouvement dans les vidéos et l'approche que nous avons choisi d'investir et met en exergue quelques caractéristiques propres à nos travaux.

Chapitre (3) : se focalise sur les techniques déjà utilisées pour l'extraction des caractéristiques et la représentation des cibles. Nous abordons aussi les descripteurs proposés dans le cadre de cette thèse pour la représentation invariante d'apparence des cibles intra/inter-caméra.

Chapitre (4) : présente un état de l'art sur les méthodes utilisées pour le suivi visuel et l'estimation des positions des cibles intra-caméra ; allant des méthodes déterministes, vers les méthodes stochastiques. Dans ce chapitre nous présentons aussi les versions améliorées du filtre particulaire et l'estimateur à noyau proposé dans le cadre de cette thèse.

Chapitre (5) : Dans ce chapitre, nous présentons le système proposé dans le cadre de cette thèse pour le suivi et la ré-identification décentralisée d'objets inter-caméra dans un réseau multi-caméra à champs disjoints sans fil et qui se base sur le paradigme multi-agent. Une présentation détaillée aura lieu aussi de l'algorithme proposé pour la collaboration entre les caméras intelligentes afin de parvenir à un consensus sur les identités des cibles circulent dans le réseau. Un ensemble d'expérimentations dans des conditions réelles d'application auront aussi lieu ; décrivant la plateforme utilisée, le nombre des nœuds, les outils technologiques utilisés et les conditions d'application. Enfin, une évaluation de l'algorithme proposé et quelques améliorations introduites pour la ré-identification d'objets inter-caméra sont présentés et discutés.

Chapitre (6) : Le dernier chapitre de cette thèse est réservé à une conclusion générale sur les contributions de nos recherches, ensuite nous présentons aussi quelques perspectives pour nos travaux futures.

Détection d'Objets en mouvement

Sommaire

2.1	Introduction	37
2.2	Etat de l'art sur la détection de mouvement	38
2.2.1	Taxinomie	38
2.2.2	Détection sans modélisation de l'arrière-plan	39
2.2.3	Détection avec modélisation de l'arrière-plan	41
2.2.4	Modélisation récursive	42
2.2.5	Modélisation non-récursive	46
2.3	Détection de l'avant-plan	47
2.4	Approche proposée avec post-traitement	48
2.5	Résultats de la détection d'objets mobiles	53
2.6	Évaluation	54
2.7	Conclusion	56

DANS ce chapitre, nous présenterons tout d'abord l'état de l'art sur la détection automatique de mouvement. Une classification possible des techniques proposées jusqu'à présent pour la détection d'objets en mouvement (segmentation de mouvement) est présentée. Nous introduisons ensuite l'approche proposée dans le cadre de cette thèse pour la détection et l'extraction d'objets en mouvement afin d'adapter les méthodes classiques avec notre application. Nous terminons ce chapitre par une présentation des résultats obtenus par l'application de notre approche sur des séquences vidéo capturées par les caméras du réseau considéré (chapitre.1, Fig.1.3).

2.1 Introduction

Les systèmes de vision artificielle ont été développés pour simuler notre système biologique qui a la capacité de faire face au changement de l'environnement, comme : le mouvement d'objets, changement de la luminosité, et aux scènes complexes. La détection d'objets mobiles est considérée comme une tâche de vision de bas niveau pour la compréhension des événements de haut

niveau tel que le comportement d'objets dans une scène. Dans les systèmes de vision modernes, comme la surveillance et la réalité augmentée, la détection automatique d'objets en mouvement devienne une tâche indispensable pour la localisation spatiale des cibles, l'extraction des caractéristiques, l'initialisation automatique des algorithmes de suivi d'objets, la ré-identification d'objets dans les réseaux multi-caméra, etc. La détection d'objets en mouvement assure une classification des pixels dans une séquence vidéo en avant-plan (objets en mouvement) et en arrière-plan. Dans la littérature plusieurs méthodes ont été proposées pour aboutir à une telle classification ; allant de l'estimation de mouvement et les dérivées temporelles ; jusqu'à la modélisation sophistiquée de l'arrière-plan couvrant des cas complexes comme le changement d'éclairage et les arrière-plans dynamiques. L'une des méthodes les plus utilisées dans ce contexte est la détection par soustraction de fond (arrière-plan). Cette opération se base sur une soustraction entre l'image courante de la séquence vidéo et l'image de l'arrière-plan estimée pour extraire les objets en avant-plan (mouvement). Pour la modélisation de l'arrière-plan au fil du temps plusieurs approches ont été proposées, allant de la modélisation par des simples images ; vers la modélisation statistique. Face aux multitudes de méthodes proposées dans la littérature, plusieurs états de l'art ont été publiés pour établir une taxonomie possible des algorithmes existants [Wang 2003], [Hu 2004]. Dans la section.2.2.1, nous présenterons une classification possible des techniques proposées jusqu'à présent pour la détection d'objets en mouvement. Dans la section.2.4 nous détaillons la chaîne de traitement que nous proposons dans cette thèse pour la détection d'Object en mouvement et la génération de la carte de soustraction binarisée et étiquetée.

2.2 Etat de l'art sur la détection de mouvement

2.2.1 Taxinomie

Depuis la naissance de cet axe de recherche qui s'articule autour de la détection automatique d'objets en mouvement ou la segmentation spatio-temporelle, plusieurs efforts ont été réalisés et plusieurs travaux ont été publiés afin de proposer des méthodes robustes aux conditions complexes de capture, comme : les objets non-rigide, arrière-plan dynamique, changement d'éclairage, etc. Vis-à-vis la multitude des méthodes proposées dans la littérature pour la détection d'objets en mouvement, la classification de ces méthodes n'est pas une tâche facile. La plus part des travaux actuels pour la détection d'objet en mouvement, sont orientés vers une application précise ou plus générale traitant les systèmes complets de surveillance. A partir des travaux

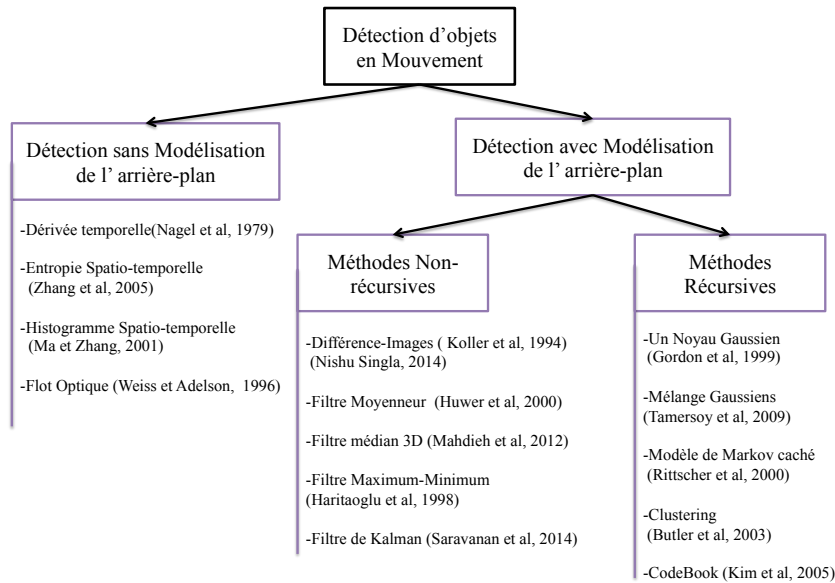


Fig. 2.1: Taxonomie Possible des méthodes de détection d'objets en mouvement.

présentés dans [Wang 2003], [Hu 2004], [Piccardi 2004], nous avons pu établir une taxonomie (classification) des méthodes proposées jusqu'à présent pour la détection d'objets en mouvement (la segmentation spatio-temporelle). En fonction de la modélisation de l'arrière-plan, nous distinguons deux grandes familles de méthodes : Détection sans modélisation de l'arrière-plan et détection avec modélisation de l'arrière-plan. La figure suivante (Fig.2.1) illustre une taxonomie possible des méthodes proposées jusqu'à présent pour la détection d'objets en mouvement.

2.2.2 Détection sans modélisation de l'arrière-plan

Ce type de détection est basé sur des mesures simples pour estimer la variation temporelle de la couleur entre deux images consécutives comme : l'histogramme spatio-temporelle et l'entropie spatio-temporelle. Ces méthodes sont souvent classées comme des estimateurs de mouvement que des détecteurs de mouvement. Dans cette section nous présentons en détail les méthodes les plus utilisées dans la détection de mouvement sans modélisation d'arrière-plan.

2.2.2.1 Dérivée temporelle

Cette méthode aussi connue comme la différence inter-image (en anglais, Frame differencing) [Radke 2005]. Elle permet de détecter les régions en mouvement et de mesurer le changement de la luminosité inter-pixel entre deux images (Frames) consécutives d'une séquence de vidéo.

Soit $I(x, y, t - 1)$ l'image capturée au temps $t - 1$ et $I(x, y, t)$ l'image capturée au temps t . L'estimation de la dérivée temporelle à l'instant t est donné par :

$$\forall (x, y) \in \mathbb{R}^2 \quad \forall (t) > 0 \quad I_t(x, y, t) \approx |I(x, y, t - 1) - I(x, y, t)| \quad (2.1)$$

Malgré la simplicité et l'efficacité de cette méthode dans des applications et des conditions très spécifiques, elle reste insuffisante dans des conditions réelles d'application, par exemple en présence : de mouvement lent ou saccadé. Cette méthode est la base de toutes les autres méthodes proposées dans la littérature pour la soustraction de fond.

2.2.2.2 Flot optique

En physique, le flot optique (en anglais, Optical Flow) est le mouvement apparent des objets, surface et contours d'une scène visuelle causée par le mouvement relatif entre l'observateur (œil humain ou camera) et la scène observée [Radke 2005]. En traitement et analyse des vidéos, le flot optique est défini comme un champ de vecteurs à 2D inter-pixel pour passer d'une image dans une séquence vidéo à une autre au fil du temps. L'estimation de mouvement en 2D à travers le mouvement apparent est une tâche très complexe parce que le mouvement apparent est très sensible par rapport aux bruits et au changement de la luminosité. Plusieurs méthodes ont été proposées dans la littérature pour l'estimation du flot optique inter-pixel allant des méthodes différentielles ; vers des méthodes basées sur la corrélation. La méthode différentielle proposée par Lucas et Kanade [Baker 2011] est la plus utilisée par rapport aux autres méthodes existantes et cela pour sa robustesse et sa rapidité. Il s'agit de calculer au temps t , le déplacement $d = (d_x, d_y)^t$ d'un point $X = (x, y)^t$ par rapport aux deux axes x et y , sous l'hypothèse de la conservation de la luminosité au fil du temps, [Verbeke 2007] :

$$\forall X \in \mathbb{R}^2 \quad \forall t > 0 \quad I(X + d, t) - I(X, t - 1) = 0 \quad (2.2)$$

L'estimation du flot optique au point X est le vecteur d' qui minimise la fonction d'erreur quadratique ξ calculée sur un voisinage $N(x)$ et définie par :

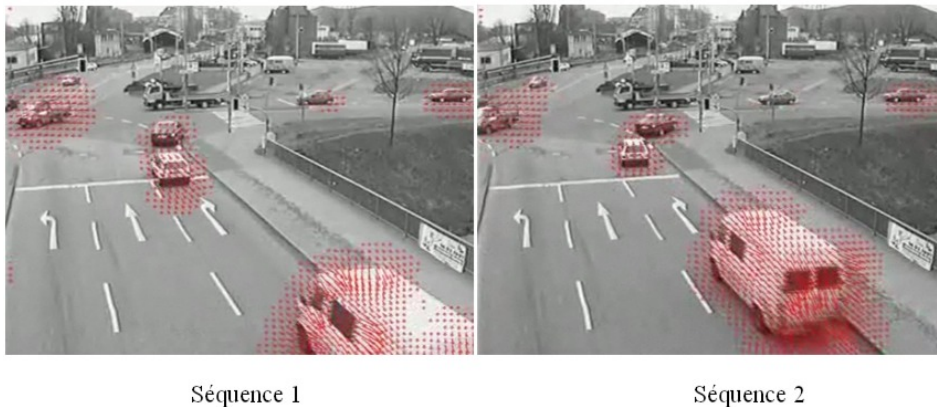


Fig. 2.2: Exemple d'estimation du flot optique, [Bruhn 2005].

$$\xi(d) = \sum_{x' \in N(x)} [I(x' + d, t) - I(x', t - 1)]^2 \quad (2.3)$$

La minimisation peut être obtenue d'une manière récursive, voir les travaux de [Baker 2011]. Un exemple de l'estimation du flot optique présenté par [Bruhn 2005] est illustré dans la figure (Fig.2.2).

2.2.3 Détection avec modélisation de l'arrière-plan

La soustraction de fond (en anglais, Background subtraction) est une classe des techniques d'extraction d'objets en mouvement utilisées dans plusieurs applications comme la surveillance. La validité de ces techniques reste conditionnée par l'arrière-plan utilisé, le développement des arrière-plans valide est une tâche très complexe :

- Premièrement, l'arrière-plan doit être robuste au changement de la luminosité et aux mouvements lents.
- Deuxièmement, il faut éviter la détection des objets mobiles dans le fond et les ombres projetées par les objets mobiles.

Un bon modèle de fond devrait rapidement répondre aux changements en arrière-plan et s'adapter automatiquement aux changements survenant dans le fond. Dans la littérature, plusieurs méthodes ont été proposées pour la modélisation de l'arrière-plan. Ces méthodes peuvent être décomposées en deux grandes familles : les méthodes récursives et les méthodes non-récursives. Les méthodes récursives, ne nécessitant pas une trace de l'historique de la luminosité de chaque pixel. Les méthodes non-récursive, nécessitent un buffer pour garder une trace de l'historique de la luminosité de chaque pixel. Dans cette partie, nous présentons en détail quelques méthodes, les plus recommandées

jusqu'à présent pour la modélisation de l'arrière-plan.

2.2.4 Modélisation récursive

A) *Simple gaussienne*

La méthode de modélisation de l'arrière-plan par une simple gaussienne ou Pfinder, permet de modéliser la distribution temporelle de l'intensité de chaque pixel par une simple distribution gaussienne. Pour chaque pixel dans l'image courante d'une séquence vidéo au temps t , le modèle utilise les n valeurs des pixels précédents comme des échantillons pour créer une fonction de distribution de probabilité caractérisée par une moyenne et un écart type. Si l'intensité des n pixels précédents de chaque image qui compose la séquence ont été stockés pour calculer la moyenne et l'écart type, alors la mémoire nécessaire pour cette méthode est très importante ce qui limite l'application de cette méthode en temps réel. Face à ce problème, d'autres chercheurs [Wren 1997] proposent une méthode plus robuste basée sur la moyenne glissante.

Soit $I_t = I(x, y, t)$ l'intensité du pixel X à l'instant t . La Moyenne glissante pour la mise à jour du modèle B_t est défini par :

$$B_t = (1 - \alpha)B_{t-1} + \alpha I_t \quad (2.4)$$

Où α est le taux d'apprentissage prédéterminé. I_t est la valeur de l'intensité précédente d'un pixel et $B(t-1)$ est la moyenne glissante précédente du modèle de l'arrière-plan. L'autre paramètre du modèle gaussien est l'écart type σ qui peut être mis à jour de la même façon en utilisant l'équation suivante :

$$\sigma_t = (1 - \alpha)\sigma_{t-1} + \alpha(I_t - B_t)^2 \quad (2.5)$$

Un simple seuillage est utilisé par la suite pour classifier les pixels avant-plan ou arrière-plan, les pixels classifiés comme avant-plan représentant le masque d'objets en mouvement. La méthode de soustraction entre l'arrière-plan estimé et l'image courante pour extraire l'avant-plan est présentée dans la section 2.3. La modélisation de l'arrière-plan par une simple gaussienne est robuste si est seulement si la dynamique de la scène est inférieure par rapport au changement introduit par les objets en mouvements. Cette méthode échoue quand un mouvement rapide aura lieu dans l'arrière-plan, ce qui limite la robustesse de cette méthode dans des conditions réelles d'application.

B) *Mélanges gaussiens*

La modélisation de l'arrière-plan (image de référence) par un simple gaussien est insuffisante surtout lorsqu'un changement rapide et important aura

lieu dans l'arrière-plan. Dans ce cas une distribution multimodale est nécessaire pour décrire la dynamique de la scène. Face à ce problème, les auteurs de [Stauffer 2000], proposent une modélisation de l'arrière-plan par un mélange gaussien (en anglais, Mixture of Gaussian «GMM »). Cette modélisation qui permet la représentation de l'image en arrière-plan avec une distribution multimodale. Dans un mélange gaussien, la distribution temporelle au fil du temps de l'intensité de chaque pixel est modélisée para-métriquement par un mélange de K gaussiennes. Le mélange gaussien est caractérisé par une moyenne, une matrice de covariance et une probabilité a priori de chaque k gaussiennes. Ces paramètres sont mis à jour dans chaque image de la séquence vidéo. Pour chaque pixel I_t , chaque gaussienne des K distributions correspond à la probabilité d'observation d'une intensité particulière, ce qui rend le mélange gaussien (GMM) une représentation plus générale par rapport à un simple gaussien. La probabilité d'observer un pixel d'intensité I_t dans l'image courante est donnée par [Stauffer 2000] :

$$I(x, y, t) = I_t = \sum_{i=1}^K w_{i,t} * \eta(I(x, y, t), \mu_{i,t}, \Sigma_{i,t}) \quad (2.6)$$

K est le nombre de distribution gaussienne, $w_{i,t}$ est la probabilité a priori de la K^{me} gaussienne au temps t , $\eta(I(x, y, t), \mu_{i,t}, \Sigma_{i,t})$ est la fonction de distribution de probabilité de la k^{eme} gaussienne avec une moyenne μ et covariance Σ . Pour simplifier les calculs, la matrice de covariance est supposée diagonale, $\Sigma_{i,t} = \sigma_k^2 I_t$ où les composantes de la couleur sont supposées indépendantes avec la même variance. Pour chaque nouvelle observation I_t , la gaussienne qui est susceptible de représenter I_t avec une forte probabilité est mise à jour. Cette condition est valide si et seulement si $\|I_t - \mu_{i,t}\| \leq 2, 5\sigma_{i,t}$ [Stauffer 2000]. La mise à jour des paramètres de la K^{eme} gaussienne qui vérifie cette condition est donnée par [Verbeke 2007] :

$$\begin{aligned} w_{i,t} &= (1 - \alpha)w_{i,t-1} + \alpha; \\ \mu_{i,t} &= (1 - \rho)\mu_{i,t-1} + \rho I_t \\ \sigma_{i,t}^2 &= (1 - \rho)\sigma_{i,t-1}^2 + \rho(I_t + \mu_{i,t}^T(I_t - \mu_{i,t})) \end{aligned} \quad (2.7)$$

$$\rho = \alpha \frac{P(i|I_t, \mu_{i,t}, \Sigma_{i,t})}{w_{i,t}} \quad (2.8)$$

ρ est une constante liée aux temps et qui détermine la vitesse avec laquelle les paramètres sont mis à jour. Dans les travaux de [Power 2002], la constante ρ est définie comme suit :

$$\begin{cases} \rho = \frac{\alpha}{w_{i,t}} & \text{Si } I_t \text{ représente la } i^{eme} \text{ distribution} \\ 0 & \text{Dans un autre Cas} \end{cases} \quad (2.9)$$

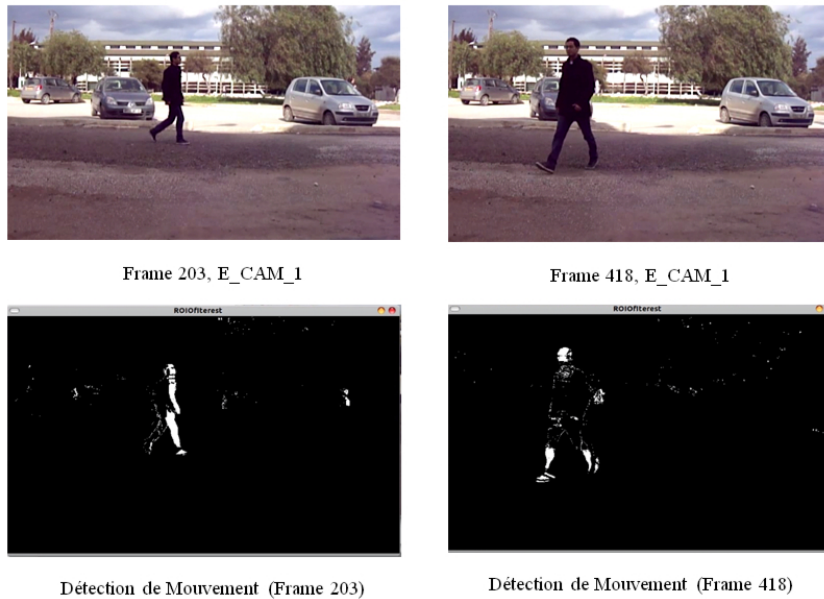


Fig. 2.3: Détection de mouvement en utilisant la soustraction de fond avec une Modélisation Statistique de l'arrière-plan par un mélange gaussiens, initialiser comme suit : $K = 4$, $\sigma = 300$ et $\mu = 100$.

Pour d'autres distributions qui ne représentent pas le pixel, les paramètres restent les mêmes. Après la mise à jour, les distributions sont ordonnancées par rapport aux poids et l'écart-type pour décider les quelles font parties du fond et celles qui représentent les objets en mouvement. Les premiers B distributions sont attribuées à l'arrière-plan.

$$B = \text{MaxArg}\left(\sum_{i=1}^b w_{i,t} > T_b\right) \quad (2.10)$$

T_b est un seuil prédéterminé. Plusieurs autres travaux publiés jusqu'à présent proposent d'autres paramétrages possibles pour améliorer cette méthode par rapport aux temps de calcul et la capacité de classification. Un exemple des résultats d'application de cette méthode en utilisant des séquences vidéo capturées par la caméra (E-CAM-1) de notre réseau multi-caméra est illustré dans la figure (Fig.2.3). Pour plus de détails sur l'initiation contextuelle du Mélanges gaussiens voir les travaux de [Zang 2006].

Malgré la recommandation de cette méthode par plusieurs chercheurs pour la détection de mouvement, les résultats d'application de cette méthode restent insuffisants dans des conditions réelles d'application en particulier en présence de changements d'éclairage, mouvement dans le fond de la scène et

mouvement non-rigide [Zang 2006], [Bouwman 2008].

C) CodeBook

"CodeBook" ou « en français, Livre de code », est une méthode proposée par [Kim 2005] pour la modélisation de l'arrière-plan. Cette méthode est basée sur une technique de quantification, où 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. Le nombre de codewords par pixel dépend de la variation du fond. 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 $I_L = \min(I_L, I_L), I_L = \max(I, I_L)$ représentent le minimum et le maximum respectivement de l'intensité attribuée au "codeword". F_L est la fréquence de la mise en correspondance de codeword. λ_L permet de décrire la durée de temps où ce codeword n'a pas été mis à jour. P_L et Q_L représentent le premier et le dernier temps d'accès au codeword respectivement. L'altération de la couleur λ entre le pixel et le codeword est calculée comme suit :

$$ALTCouleur(X_t, V_L) = \delta = \sqrt{\|X_t\|^2 - C_p^2} \quad (2.11)$$

$\|X_t\|^2$ et C_p^2 sont calculés utilisant les deux équations suivantes :

$$\|X_t\|^2 = R^2 + V^2 + B^2 \quad (2.12)$$

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 :

$$C_p^2 = \|X_t\|^2 \cos^2\theta = \frac{(R_i R + V_i V + B_i B)^2}{R_i^2 + V_i^2 + B_i^2} \quad (2.13)$$

Initialement, le codebook de chaque pixel est vide et ne contient pas des *codeword's*. 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. [Kim 2005] donnent plus de détails de cette méthode. Cette méthode a été prouvée dans le même travail comme robuste par rapport au mouvement

rapide et au changement d'éclairage. D'autres chercheurs [Ilyas 2009] 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. Dans le même cadre de la modélisation réursive d'arrière-plan pour la détection de mouvement, plusieurs méthodes ont été proposées. Cet axe reste très actif à l'heure actuelle et en développement remarquable ces dernières années. Plusieurs autres méthodes pour la modélisation de l'arrière-plan ont été proposées comme la modélisation par un modèle de Markov cachée [Rittscher 2000], les méthodes prédictives [Koller 1994], etc.

2.2.5 Modélisation non-réursive

Les méthodes non-réursives de la modélisation de l'arrière-plan se basent essentiellement sur l'historique de l'intensité des pixels afin de construire un modèle robuste de l'arrière-plan. Dans cette orientation il y a eu beaucoup d'investissement et plusieurs méthodes ont été proposées. La forme la plus basique d'un modèle d'arrière-plan est l'utilisation des images de références [François 1999] ou de l'image précédente (Frame differencing en anglais) [Radke 2005]. Ces méthodes restent très limitées dans plusieurs conditions comme : le mouvement rapide, le changement d'éclairage, la stabilité et la non-stabilité d'objets, etc. Face à ces problèmes d'autres chercheurs comme dans [François 1999], proposent la modélisation de l'arrière-plan utilisant un filtre moyennier où la valeur de chaque pixel de l'arrière-plan courant est égale à la moyenne de toutes les valeurs des pixels gardés dans l'historique. Dans ce même contexte d'autres chercheurs [Cucchiara 2003] proposent l'utilisation d'un filtre de type médian pour la mise à jour de l'arrière-plan. Le problème de ces méthodes est que les informations des pixels proviennent non seulement de l'arrière-plan précédent mais aussi de l'avant plan précédent. Un filtre de Kalman est aussi utilisé par [Weng 2006], où l'arrière-plan courant est calculé par la prédiction de la valeur des pixels en utilisant l'historique. Le grand souci de ces méthodes c'est le temps de calcul et le changement d'éclairage. D'autres chercheurs [Haritaoglu 1998], proposent la modélisation de chaque pixel par une variable aléatoire dans un espace de caractéristiques spécifiques puis l'associe à une fonction de distribution de probabilité. L'estimation de cette distribution de probabilité se fait par un estimateur à noyau. Cette méthode est considérée comme une forme générale de la mixture gaussienne non-paramétrique. Sa robustesse est strictement conditionnée par la dimension des données et l'espace des caractéristiques utilisées. La modélisation non réursive de l'arrière-plan pour la détection de mouvement a fait l'objet de plusieurs recherches. Ces méthodes nécessitent une grande mémoire et une capacité de calcul importante.

2.3 Détection de l'avant-plan

Après la modélisation de l'arrière-plan par l'une des méthodes présentées préalablement (récursives ou non-récursives), une opération de comparaison entre l'image courante de la séquence vidéo et l'arrière-plan estimé est utilisée pour la détection de l'avant-plan (détection d'objets en mouvement). Les pixels de l'avant-plan seront identifiés par rapport aux pixels de l'arrière-plan. Pour obtenir une telle classification, une opération de binarisation par seuillage de la carte de soustraction est utilisée. Le seuil de binarisation dépend de la scène, de l'application, des bruits de la caméra utilisée et des conditions d'éclairage. Plusieurs méthodes ont été proposées pour la génération de la carte de soustraction entre l'image courante et l'arrière-plan afin d'extraire les objets en mouvement. Nous citons dans cette section les plus récentes : Une soustraction normalisée a été proposée dans [García-Martín 2010] où la carte de soustraction est calculée par une valeur absolue de la différence entre les pixels correspondant dans l'arrière-plan estimé et l'image courante. Dans le même contexte la variabilité spatiale a été pris en compte dans [Cucchiara 2003] par l'utilisation de la différence souillée entre les images (arrière-plan et l'image courante). Dans [Mohamed 2010], les auteurs proposent l'utilisation de la différence relative plutôt que la différence absolue où le seuillage est en fonction de la localisation spatiale des pixels. Une simple gaussienne a été utilisée par [Horprasert 1999], où les pixels de l'image courante sont comparés avec l'arrière-plan estimé par la mesure de la vraisemblance dans l'espace de couleur. D'autres méthodes ont été aussi proposées dans la littérature afin de générer une carte de soustraction binarisée, nous pouvons citer aussi, les mélanges gaussiens [Bak 2010], l'estimateur à noyau [Migdal 2005], etc. La technique la plus utilisée jusqu'à présent pour la détection de l'avant plan et la génération de la carte de soustraction binarisée est la soustraction absolue avec seuillage. La formalisation mathématique de cette technique est définie comme suit :

$$F(I(x,y,t),B)=\begin{cases} 255 & \text{Si } |I(x,y,t) - B| > T^0 \\ 0 & \text{Dans un autre Cas} \end{cases} \quad (2.14)$$

T^0 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.

2.4 Approche proposée avec post-traitement

La détection d'objets en mouvement est l'une des tâches les plus importantes dans les systèmes de vision moderne comme la surveillance et la réalité virtuelle. Le résultat de cette tâche fera l'entrée d'un autre processus de haut niveau tel que : le suivi des cibles, la ré-identification, l'analyse de mouvement et la reconnaissance. Dans la littérature et comme nous l'avons détaillé précédemment, plusieurs méthodes ont été proposées. La méthode de soustraction de fond avec modélisation de l'arrière-plan est l'une des méthodes les plus utilisées jusqu'à présent. Ces méthodes restent limitées dans des conditions et des applications spécifiques, ce qui exige l'adaptation de ces méthodes avec nos besoins. Dans cette partie nous proposons une chaîne de traitement robuste pour la détection d'objets en mouvement et la génération de la carte de soustraction binarisée et étiquetée ([Bousetouane 2011],[Bousetouane 2013b]). La carte de soustraction étiquetée est utilisée dans l'extraction des caractéristiques, le suivi et la ré-identification des cibles inter-caméra dans notre système. La chaîne de traitement proposée pour la détection d'objets en mouvement se décompose en quatre modules essentiels :(1) Un pré-traitement, basé sur un filtre moyenneur pour équilibrer la dynamique de l'image.(2) Modélisation de l'arrière-plan par un mélange gaussien pour la génération de la carte de soustraction binarisée. (3) Un post-traitement, basé sur une convolution d'un noyau gaussien avec la carte de soustraction dilatée par un élément structurant elliptique pour remplir le vide entre les pixels de la même région. (4) En fin, une opération d'étiquetage des composantes connexes(en anglais, Commun Region Labeling) est utilisée pour l'étiquetage et le comptage des cibles. Le schéma général de la chaîne de traitement proposée est illustré dans la figure (Fig.2.4).

Comme illustré dans la figure (Fig.2.4), la chaîne de traitement proposée pour la détection d'objets en mouvement afin de construire une carte de soustraction binarisée et étiquetée est composée de quatre processus successifs :

1.Pré-traitement

Afin de régulariser la dynamique de l'image et la réduction des bruits, un filtrage de l'image courante de la séquence vidéo par un filtre moyenneur est effectué.

Soit $I(x, y, t)$ l'image courante d'une séquence vidéo au temps t de dimension $N \times M$, la convolution de cette image par un noyau moyen G de taille $n \times m$

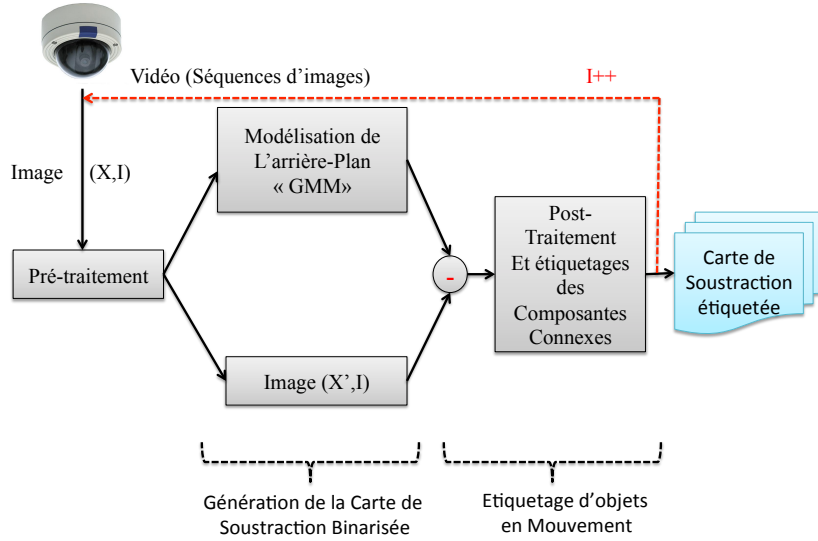


Fig. 2.4: Schéma global de la chaîne de traitement proposée pour la génération de la carte de soustraction binarisée et étiquetée

où ($n=3$ et $m=3$) est définie comme suit :

$$O(x, y) = I. * .G = \sum_{K=1}^m \sum_{l=1}^n I(x + K - 1, y + l - 1)G(K, l) \quad (2.15)$$

$$G(K, l) = \frac{1}{9} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

(2.16)

. * . est l'opération de convolution entre deux matrices, $x = 1$ jusqu'à $N - m + 1$ et $y = 1$ jusqu'à $M - n + 1$.

2. Génération de la Carte de soustraction binarisée

Ce processus incorpore deux opérations : La modélisation de l'arrière-plan par un mélange gaussiens (présentée précédemment dans la section.2.2.4) suivi par une opération de soustraction absolue avec seuillage entre l'image courante et l'arrière-plan estimé afin de construire une carte de soustraction binarisée. Le

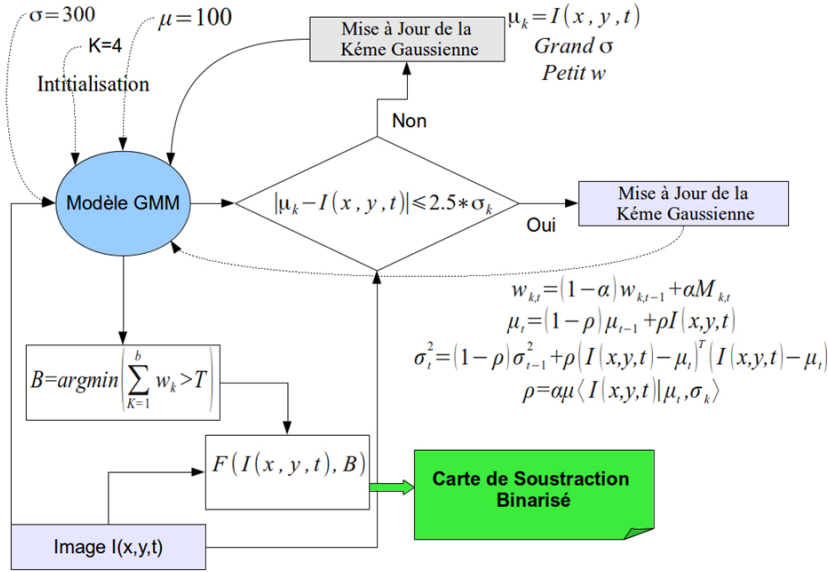


Fig. 2.5: Schéma global et paramétrage du processus de la génération de la carte de soustraction binarisée

schéma global de ce processus et le paramétrage utilisé sont illustrés dans la figure suivante (Fig.2.5).

3. Post-Traitement

Après la génération de la carte de soustraction binarisée, une opération de post-traitement est nécessaire pour plusieurs raisons : consolider les régions (blobs) les plus proches et l'élimination des régions isolées non-significatifs. Pour assurer une telle opération nous proposons : (1) un filtrage de la carte de soustraction binarisée utilisant une opération de convolution de cette carte par un noyau gaussien afin d'enlever les petites régions isolées. l'isolation de ces régions est causé par le mouvement dans le fond, mouvement des arabes, les ombres et la pluie. (2) Après le filtrage, une dilatation de la carte de soustraction filtrée par un élément structurant elliptique est effectuée afin de connecter les régions (blobs) les plus proches. Le post-traitement proposé est défini comme suit :

$$\psi_{B_R}(I(x, y, t)) = I(x, y, t) \oplus B_R$$

$$\text{GeD}(\psi_{B_R}(I(x, y, t)), G(i, j)) = \frac{1}{\sigma^2(2\pi)} e^{-\frac{x^2+y^2}{2\sigma^2}} * \psi_{B_R}(I(x, y, t)) \quad (2.17)$$

GeD est le post-traitement proposé (en anglais, Gaussian Elliptical Dila-

tion Filter), $\psi_B(I(x, y, t))$ est la dilatation de la carte de soustraction binarisée calculée au temps t par un élément structurant B de rayon R . $\cdot *_x \cdot$ est la convolution de l'image dilatée par rapport à l'axe x et $\cdot *_y \cdot$ est la convolution de l'image dilatée par rapport à l'axe y . Nous pouvons définir directement le noyau gaussien de taille 5×5 dans l'espace $2D$ avec $\sigma = 1$ par la matrice suivante :

$$G(i, j) = \frac{1}{273} \begin{bmatrix} 1 & 4 & 7 & 4 & 1 \\ 4 & 16 & 26 & 16 & 4 \\ 7 & 26 & 41 & 26 & 7 \\ 4 & 16 & 26 & 16 & 4 \\ 1 & 4 & 7 & 4 & 1 \end{bmatrix} \quad (2.18)$$

Les résultats d'application de ce post-traitement sur des cartes de soustraction binarisée générées à partir des séquences d'images capturées par la camera E-CAM-1 de notre réseau multi-caméra en présence des conditions complexes sont illustrées dans la figure (Fig.2.7(b)). Les résultats obtenus prouvent l'efficacité du post-traitement proposé pour l'élimination des zones isolées non-porteuse d'informations pertinentes et le remplissage des pixels entre les régions les plus proches. Comparant les résultats obtenus de la génération de la carte de soustraction binasiée avec le post-traitement proposé et les résultats d'application directe de mélange gaussiens (voir la figure (Fig.2.7(a))), nous pouvons visuellement remarquer la pertinence du post-traitement proposé.

4. Étiquetage des composantes connexes

L'étiquetage des composantes connexes (en anglais Connected-Component Labeling) est une étape très importante dans un système de reconnaissance. Cette étape assure une labialisation des composantes connectées dans une scène afin de faciliter l'extraction et l'interprétation automatique de ces composantes, par exemple : le nombre, les positions et la morphologie d'objets dans une scène. Après la génération de la carte de soustraction binarisée et filtrée, cette carte encode des informations importantes concernant le mouvement des cibles, les positions, les périmètres, le nombre des cibles, etc. Cette carte est interprétée facilement par l'être humain mais difficilement par une machine, ce qui nécessite une opération d'étiquetage afin de faciliter l'opération d'extraction automatique des informations concernant les cibles. Cette étape est très importante dans notre système pour l'extraction des caractéristiques, la représentation des cibles, le comptage d'objets en mouvement, l'extraction

de la dimension réelle des cibles pour l'adaptation de l'algorithme de suivi, l'initialisation de l'algorithme de suivi, la ré-identification inter-caméra, etc. Dans la littérature plusieurs algorithmes ont été proposés pour l'étiquetage des composantes connexes dans des images binaire afin d'extraire les objets en premier-plan. La plupart de ces algorithmes se basent sur la théorie des graphes où les nœuds d'un graphe contiennent des informations sur l'heuristique de la comparaison, les arrêtes du même graphe indiquant la connectivité dans le voisinage. Deux types de connectivité sont possibles : 4-connectivité et 8-connectivité. Union-Find est l'un des algorithmes le plus utilisé jusqu'à présent pour l'étiquetage des composantes connexes [Soman 2010], par rapport a sa simplicité et sa rapidité. Cet algorithme se base principalement sur deux étape : (1) enregistrement des équivalences et attribution des étiquettes temporaires. (2) Remplacement de chaque étiquette par sa classe équivalente.

Une description générale de cet algorithme est illustrée dans l'algorithme (Algorithme.2.4)

Algorithm.1. Algorithme d'étiquetage des composantes connexes

INPUT : Carte de soustraction binarisée et filtrée $CS(x, y, t)$, la connectivité utilisée (4)

OUTPUT : Les étiquettes des Composantes de la carte de soustraction binarisée et filtrée $GS(x, y, t)$

début

Première étape (first pass)

1. Parcourir tous les pixels $P_{x,y}$ de la carte de soustraction binarisée et filtrée $CS(x, y, t)$ par colonne, puis par ligne.
2. Si le pixel P_{xy} n'appartient pas à l'arrière-plan ($P_{x,y} \neq 0$) alors :
 - a. Obtenez les pixels voisins par rapport au pixel courant $P_{x,y}$.
 - b. S'il ya pas des voisins, alors attribuer une étiquette au pixel courant $P_{x,y}$.
 - c. Dans un autre cas, trouver les pixels voisins (dont l'étiquette est la plus petite) et attribuer la même étiquette du pixel courant $P_{x,y}$.
 - d. Enregistrer les équivalences entre les étiquettes dans le voisinage.

Deuxième étape (second pass)

1. Parcourir tous les pixels $P_{x,y}$ de la carte de soustraction binarisée et filtrée $CS(x, y, t)$ après le 1^{er} étiquetage par colonne, puis par ligne
2. Si le pixel $P_{x,y}$ n'appartient pas à l'arrière-plan ($P_{x,y} \neq 0$) alors Ré-étiqueter le pixel avec l'étiquette la plus faible avec l'étiquette équivalent.

Fin

L'efficacité de cet algorithme est strictement conditionnée par la qualité de la classification avant-plan/arrière-plan obtenue. Dans cette étape de notre chaîne de traitement, cet algorithme a été utilisé par rapport à sa simplicité et son efficacité prouvée dans plusieurs applications. Les résultats d'application de cet algorithme (Union-Find) pour l'étiquetage des composantes connexes sur des cartes de soustraction binarisées et filtrées générées à partir des séquences vidéo capturées par la caméra E-CAM-1 de notre réseau multi-caméra sont illustrés dans la figure suivante, (Fig.2.6). A travers cette étape nous pouvons extraire automatiquement plusieurs informations concernant les objets en mouvement : les périmètres d'objets, le nombre d'objets mobiles dans la scène, la localisation spatiale d'objets, la dimension, etc. Cette étape est essentielle dans notre système.

2.5 Résultats de la détection d'objets mobiles

Dans cette partie nous présentons quelques résultats d'application de notre chaîne de traitement pour la détection de mouvement (génération de la carte de soustraction binarisée) sur des vidéos capturées par des caméras de notre réseau multi-caméra ad-hoc sans fil a champs disjoints (voir chap.1, Fig.1.3). La figure suivante (Fig.2.7) illustre une comparaison entre les résultats de notre chaîne de traitement et la soustraction de fond classique basé sur un mélange gaussiens pour la détection de mouvement.

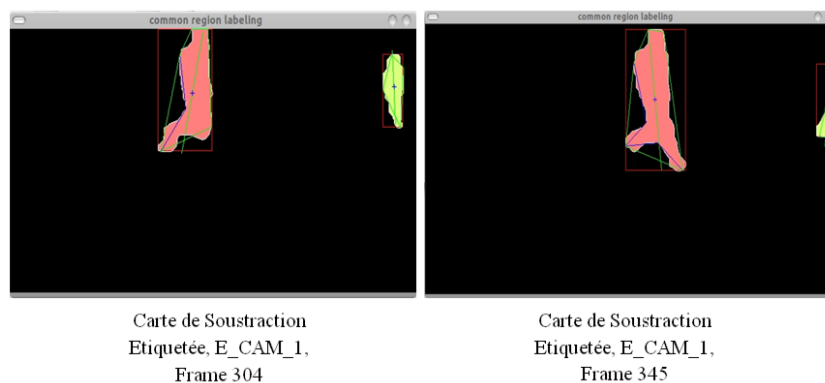


Fig. 2.6: Génération de la carte de soustraction Étiquetée

[h]

À partir de cette comparaison nous pouvons visuellement remarquer que notre approche est meilleure que la soustraction de fond basée sur un mélange gaussiens classique sans post-traitement pour la détection d'objets en

mouvement. La puissance de l'approche proposée réside au niveau du post-traitement proposé où les pixels non pertinents, comme les pixels isolés, ont été éliminés et d'autres pixels pertinents, en terme de connexité ont été groupés. Une évaluation quantitative de notre approche pour la détection d'objets en mouvement est détaillée dans la section suivante.

2.6 Évaluation

Pour l'évaluation de la chaîne de traitement proposée pour la détection d'objets en mouvement et la génération de la carte de soustraction binarisée, nous avons utilisé la méthode proposée dans [Benezeth 2008]. Cette méthode d'évaluation se base sur quatre étapes essentielles :

1. Un ensemble des séquences d'images sont sélectionnées pour l'évaluation, et tous les objets en mouvement présent dans les séquences sont automatiquement détectés et corrigés manuellement si nécessaire afin d'obtenir la vérité du terrain (en anglais, ground truth).
2. Les résultats de la détection d'objets en mouvement sur les séquences sélectionnées utilisant notre approche sont comparés avec la vérité du terrain.
3. Les erreurs de classification sont calculées.
4. Pour chaque type d'erreurs, des indices statistiques sont aussi calculés (moyenne, écart-type, etc).

Cette méthode se base sur la mise en correspondance entre les régions d'intérêts dans la vérité du terrain et les régions en mouvement détectées à travers l'algorithme de détection de mouvement. Plusieurs configurations spatiales (pattern) ont été proposées par les auteurs pour assurer la mise en correspondance inter-régions. Pour plus de détails sur cette méthode d'évaluation des algorithmes de détection de mouvement, voir la publication de [Benezeth 2008]. Dans notre cas, nous avons sélectionné des séquences d'images capturées par la caméra E-CAM-1 de notre réseau multi-caméra d'une taille de 50 frames où une image (frame) par seconde. La figure suivante illustre le graphe ROC des taux des détections correctes en fonctions des taux des fausses détections. A partir de la courbe ROC nous pouvons clairement remarquer que les taux des détections correctes en utilisant notre approche sont plus importants que les taux de détections correctes utilisant un mélange gaussiens pour la modélisation statistique de l'arrière-plan sans post-traitement. Cette expérimentation illustre l'efficacité de l'approche proposée pour la détection d'objets mobiles et sa robustesse en présence des conditions complexes comme la vibration et le mouvement saccadé en arrière-plan, mouvement non-rigide, ombres, arrière-plan texturée, etc.

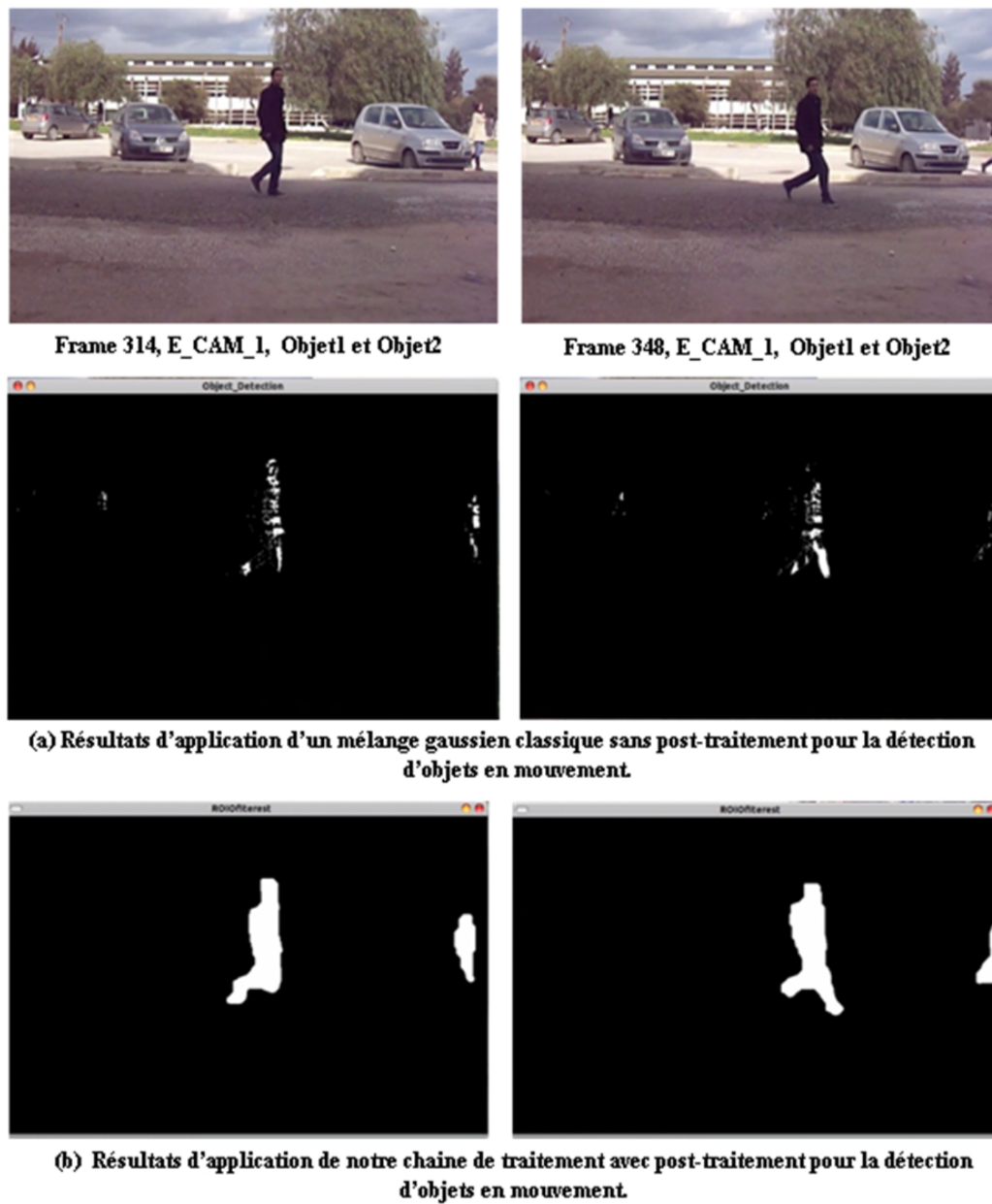


Fig. 2.7: Comparaison entre les résultats de la détection d'objets en mouvement utilisant un mélange gaussiens (a) et utilisant notre chaine de traitement avec post-traitement (b). $K = 4$, $\sigma = 300$, $\tau = 100$ initialement.

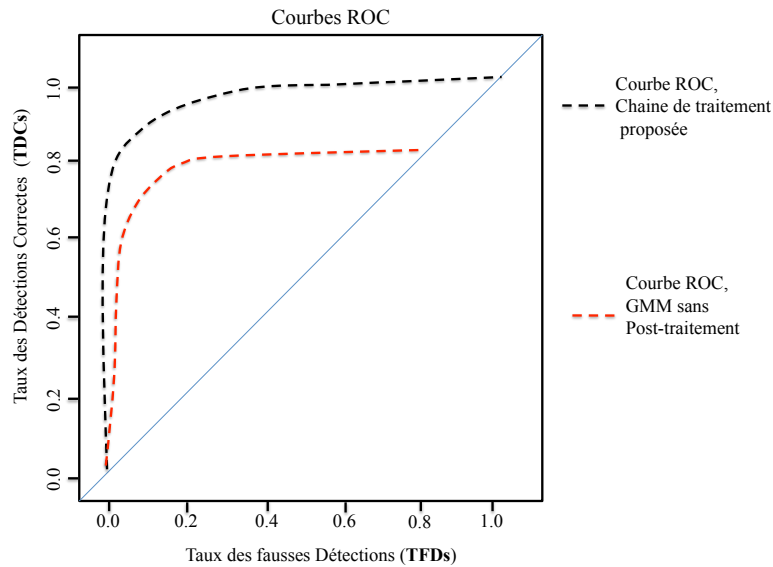


Fig. 2.8: la Courbe ROC entre les TFD et les TDC

2.7 Conclusion

Dans ce chapitre nous avons préalablement discuté un état de l'art sur les méthodes proposées jusqu'à présent pour la détection de mouvement (la segmentation de mouvement). A travers l'application de certaines méthodes de l'état de l'art pour la détection d'objets en mouvement nous avons constaté que la détection de mouvement reste un problème au cœur de la recherche. Les méthodes proposées jusqu'à présent restent limitées dans des applications très spécifiques ce qui nécessite l'adaptation de ces méthodes avec nos besoins. La deuxième partie de ce chapitre a été réservée à une nouvelle chaîne de traitement proposée pour améliorer les performances des méthodes déjà proposées et l'adaptation de ces méthodes avec notre application. Les résultats de détection d'objets mobiles obtenus sont très remarquables. La chaîne de traitement proposée pour la détection d'objets en mouvement peut être utilisée dans d'autres applications avec d'autres paramètres. Dans le chapitre suivant nous présenterons l'exploitation des cartes de soustraction étiquetées générées à partir des séquences vidéos capturées par notre réseau multi-caméra dans le processus d'extraction des caractéristiques pour la représentation des cibles.

Extraction des caractéristiques et représentation des cibles

Sommaire

3.1	Introduction	58
3.2	Etat de l'art sur l'extraction des caractéristiques	59
3.2.1	Taxonomie	59
3.2.2	Caractéristiques de Bas-Niveau	59
3.2.3	Caractéristiques de Niveau Intermédiaire	67
3.2.4	Caractéristiques de Haut Niveau	68
3.3	Etat de l'art sur la représentation des cibles	69
3.3.1	Taxonomie	69
3.3.2	La Forme	69
3.3.3	L'apparence	70
3.4	Mesures de similarités	71
3.4.1	Distance de Bhattacharya	71
3.4.2	Divergence de Kullback-leibler	72
3.5	Modèle de représentation des cibles proposé	72
3.6	Evaluation	74
3.6.1	L'invariance inter-image et intra-caméra	76
3.6.2	Capacité discriminatoire du modèle proposé inter-Objet	76
3.6.3	L'invariance aux conditions d'éclairages inter-Camera	76
3.6.4	L'invariance aux autres modalités de captures	77
3.7	Conclusion	80

DANS ce chapitre, nous présentons tout d'abord un état de l'art sur les techniques actuelles les plus utilisées pour l'extraction des caractéristiques (indices visuels) à partir des vidéos. Cette partie est suivie par les détails des indices visuels utilisés pour la représentation des cibles dans les systèmes de suivi visuel actuels. Nous terminons ce chapitre par la description du modèle de représentation des cibles proposé dans le cadre de cette thèse. Dans cette partie, nous détaillons le modèle proposé, la métrique utilisée pour le calcul de

la similarité à différents temps, ensuite, nous évaluons la consistance du modèle proposé par rapport au changement d'éclairage inter-caméra, inter-image et par rapport à la capacité discriminatoire inter-objet.

3.1 Introduction

La performance d'un système de vision artificielle ; d'un algorithme de suivi ou d'un processus de reconnaissance dépend de la qualité de l'information qu'on peut extraire à partir de l'image ou la vidéo. L'extraction des caractéristiques visuelles est souvent référée par des chercheurs comme une étape clé dans un processus de reconnaissance. Elle consiste en un ensemble de transformations mathématiques calculées sur les pixels d'une image ou des séquences d'images. Cependant, l'extraction des indices visuels significatifs à partir des images ou des vidéos est un problème très complexe dans un système de vision. Les indices extraits à partir d'une région d'intérêt doivent être invariants aux conditions d'éclairage et discriminants par rapport aux autres régions codant d'autres contenus sémantiques. Comme nous l'avons déjà détaillé dans le premier chapitre, la validité et la robustesse des opérations de suivi et de ré-identification d'objets inter-caméra dans un réseau multi-caméra dépendent de la qualité représentative des indices visuels extraits à partir des régions d'intérêts (objets mobiles). Ces indices doivent être invariants aux conditions d'éclairage inter-caméra et inter-image et discriminants inter-objet. Dans la littérature plusieurs indices visuels ont été proposés pour mieux représenter les régions d'intérêts dans un système de suivi ; allant des indices de bas niveau comme la couleur et le mouvement ; vers des indices d'un niveau d'abstraction plus élevé qui encode des caractéristiques basées sur la modélisation d'objets, comme les modèles articulatoire pour les objets humains. Dans la section.3.2.1, nous présentons une classification possible des techniques proposées jusqu'à présent pour l'extraction d'indices visuels, et qui sont utilisées dans les systèmes de suivi moderne pour la représentation des cibles mobiles. Dans la deuxième partie de notre thèse, nos efforts sont mis dans le test des indices visuels proposés dans la littérature pour aboutir à une description visuelle invariante et discriminante des objets inter-caméra. Les tests que nous avons effectués nous ont mené à proposer une nouvelle représentation des cibles basée sur une mixture ad-hoc des indices de texture et des moments statistiques, afin de créer un modèle d'apparence invariant inter-caméra et discriminant inter-objet. Le modèle proposé est brièvement détaillé dans la section.3.5, de ce chapitre.

3.2 Etat de l'art sur l'extraction des caractéristiques

3.2.1 Taxonomie

Dans la littérature plusieurs techniques ont été proposées afin d'extraire des indices visuels, riches et significatifs, allant de la couleur avec ces variantes d'espaces, du mouvement et de la texture ; vers les caractéristiques (descripteurs) de plus en plus compliquées basées objets. Vis-à-vis de la multitude des techniques proposées dans la littérature pour l'extraction des caractéristiques, classer et hiérarchiser ces techniques n'est pas une tâche facile. Dans cette partie, nous proposons une taxonomie possible des techniques utilisées pour l'extraction des caractéristiques à partir des vidéos dans les systèmes de suivi modernes. Ces caractéristiques sont classées par rapport à son niveau d'abstraction, où nous distinguons trois grands groupes des techniques : (1) les techniques de bas niveau, englobant les techniques basées sur des transformations mathématiques de bas niveau dans les images ou les vidéos. (2) Les techniques d'un niveau intermédiaire, se basant sur les points et les régions d'intérêts. (3) Les techniques de haut niveau, se basant sur la modélisation d'objets d'intérêts, ces techniques sont souvent classées comme des techniques de représentation des cibles dans les systèmes de suivi. Dans ce travail nous distinguons entre le processus d'extraction des caractéristiques et le processus de représentation des cibles. La figure suivante illustre une taxonomie possible des techniques utilisées jusqu'à présent pour l'extraction des caractéristiques à partir des images ou des vidéos dans les systèmes de suivi visuel moderne.

3.2.2 Caractéristiques de Bas-Niveau

Les caractéristiques de bas niveaux, extraites à partir d'une image ou d'une séquence d'images consistent en des transformations mathématiques de bas niveau calculées sur les pixels. Dans cette section, nous détaillons les caractéristiques les plus utilisées jusqu'à présent dans le contexte du suivi et la ré-identification d'objets inter et intra-caméra. Ensuite, nous présentons les caractéristiques utilisées dans le cadre de cette thèse pour une représentation invariante et discriminante des cibles.

3.2.2.1 Couleur

La couleur est l'une des caractéristiques les plus utilisées jusqu'à présent pour le suivi et la ré-identification d'objets dans les systèmes de suivi. Citons comme exemple : la calibration de la couleur inter-caméra [Chen 2011], la re-

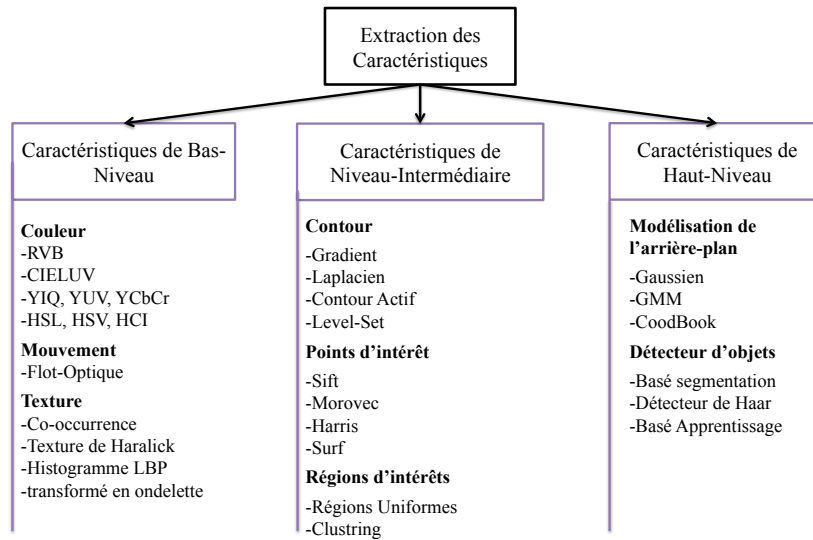


Fig. 3.1: Taxonomie Possible des techniques utilisées jusqu'à présent pour l'extraction des caractéristiques dans les systèmes de suivi visuel

présentation des cibles dans les algorithmes de suivi comme dans le MeanShift [Bousetouane 2013b], le filtre particulaire [Bousetouane 2012], [Bousetouane 2013a], etc. La couleur est une caractéristique très importante pour la description de des signatures visuelle d'objets.. Un espace colorimétrique est une représentation mathématique de nos perceptions visuelles qui nous permet d'analyser et de gérer les couleurs. Dans la littérature plusieurs espaces de représentation de la couleur ont été utilisées dans les systèmes de suivi. Dans cette partie, nous détaillons les espaces de couleur les plus utilisés :

1. RVB

Rouge, Vert, Bleu : est un modèle non-linéaire de l'espace de couleur qui utilise un système de coordonnées cartésiennes, avec trois axes correspondant aux rouge, vert, bleu. Cet espace de couleur a été utilisé dans plusieurs systèmes de suivi et de ré-identification d'objets inter et intra-caméra. Cependant, cet espace reste très sensible aux variations de la couleur, à la déformation locale et au mouvement rapide, ce qui rend la description des cibles d'intérêts avec cet espace insuffisante dans des conditions réelles d'application.

2. TSL

Teinte, saturation, lumière : représente un ensemble d'espace colorimétrique comme : *HSI* (*I* pour intensity), *HSV* (*V* pour Value), *HCI* (*CI* pour chroma colourfulness, intensity), etc. Parmi ces variantes de *HSL*, *HSV* est l'espace de couleur le plus utilisé dans la communauté qui travaille sur le suivi visuel, par exemple cet espace a été utilisé pour la description des cibles dans l'algorithme de suivi CamShift et dans la ré-identification inter-caméra [Javed 2008a]. Cet espace de couleur consiste en une transformation non-linéaire de l'espace *RGB*. La valeur *Hue* « *H* » est mesurée par l'angle autour de l'axe vertical, avec un rouge de 0, vert à 120 et l'autre couleur est calculée autour de l'angle 180. Dans la littérature plusieurs techniques ont été proposées pour la conversion entre l'espace *HSV* et l'espace *RGB*, l'algorithme proposé dans [Hanbury 2003] est le plus utilisé. Malgré que la couleur avec ces variantes d'espace soit l'indice visuel le plus important et même le plus utilisé dans les systèmes de suivi, mais il reste insuffisant pour la description des cibles. La déviation de la couleur est causée par plusieurs facteurs tel que : la déformation locale, la variation temporaire, l'éclairage, le mouvement rapide, etc. Ces facteurs engendrent une variation remarquable dans la signature d'apparence des cibles et qui mène à une divergence dans le suivi, la localisation des cibles et dans l'identification/ré-identification d'objets inter et intra-caméra.

3.2.2.2 Moments Statistiques

Les moments statistiques définis dans [Loza 2007] comme étant l'espérance statistique de certaines fonctions de puissance d'une variable aléatoire. En traitement d'image, les moments statistiques sont des mesures calculées autour d'un pixel donnent des informations pertinentes sur le voisinage comme : la tendance centrale, l'orientation, la nature de la relation entre les pixels, etc. L'utilisation de la distribution globale dans les systèmes de suivi et la reconnaissance (comme l'histogramme de couleur) est une solution couteuse en temps et en espace mémoire surtout pour les applications à temps réels. La résolution de ce problème nous mène à calculer uniquement les dominantes caractéristiques de la couleur dans les régions d'intérêts (comme la variance, la moyenne, le 3eme et le 4eme moment, etc.) au lieu de calculer la distribution complète. Dans [Chan 2004], les auteurs ont prouvés que les méthodes de reconnaissance qui se basent sur des moments statistiques marchent plus vite avec des résultats meilleurs que les méthodes qui se basent sur les histogrammes. Dans le contexte de suivi visuel d'objets intra-caméra, les moments statistiques sont utilisés pour la description des cibles pour le suivi des yeux, des visages, etc. Les moments statistique de 1er ordre sont calculés a partir de l'histogramme de couleur.

Dans cette thèse, nous avons choisi l'utilisation de quatre moments statistiques prouvés comme invariants par rapport au changement d'échelle, orientation et à la déformation locale. Les moments sélectionnés sont : la moyenne, la variance, skewness et Kurtosis.

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. Les moments statistiques [Bob 2002] 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 m_1 . 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 :

$$m_1 = E[f(i)] = \sum_{i=0}^{G-1} if(i) \quad (3.1)$$

2. Variance : c'est le deuxième moment m_2 . 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.

$$m_2 = \sigma^2 = \sum_{i=0}^{G-1} (i - m_1)^2 f(i) \quad (3.2)$$

3. Coefficient de dissymétrie (Skewness) : c'est le moment standardiser d'ordre trois m_3 . Il permet de caractériser l'asymétrie dans la distribution locale de l'intensité autour du centre de la région d'intérêt. Ce moment est souvent utilisé pour une description discriminante inter-tissu dans des applications médicale [Rahman 2007].

$$m_3 = \mu_3 = \sigma^{-3} = \sum_{i=0}^{G-1} (i - m_1)^3 f(i) \quad (3.3)$$

4. Coefficient d'aplatissement (Kurtosis) : c'est le moment standardiser d'ordre quatre m_4 . Ce moment converge vers une grande valeur si un grand nombre de pixels se rapproche de la valeur moyenne dans un voisinage.

Ce moment est souvent combiné avec le moment Skewness pour une analyse discriminante des images : segmentation, reconnaissance, indexation, etc. La formulation mathématique du Kurtosis est donné par :

$$m_4 = \mu_4 = \sigma^{-4} = \sum_{i=0}^{G-1} (i - m_1)^4 f(i) - 3 \quad (3.4)$$

3.2.2.3 Texture et la Cooccurrence

L'analyse de la texture est l'une des techniques les plus utilisées pour la description du contexte spatial et la relation entre les pixels dans une région d'intérêt à différents niveaux. La texture est définie comme une fonction de la variation spatiale de l'intensité des pixels. Elle est souvent combinée avec la couleur pour améliorer les résultats de la segmentation ou la reconnaissance. Dans la littérature plusieurs techniques ont été proposées pour l'analyse de la texture à partir des images, allant des statistiques du premier ordre, de la transformée en ondelette, des histogrammes LBP ; vers l'exploitation de l'auto-corrélation et la relation entre les pixels, comme la cooccurrence et les indices de Haralick. Parmi toutes ces techniques proposées pour l'analyse de la texture à partir des images, nous avons choisi l'utilisation des indices de Haralick pour la description des cibles par rapport à l'invariance et la capacité discriminatoire de ces indices prouvées dans plusieurs travaux de recherche. Les indices de Haralick ont été utilisés dans plusieurs domaines : classification d'image satellitaire [Bekkari 2011], indexation d'image [Chitrakala 2009] et la reconnaissance des structures cérébrales dans les images IRM-3D [Bousetouane 2010]. Néanmoins, les indices de Haralick n'ont jamais été utilisés dans le contexte de suivi et l'analyse de mouvement. Les indices de texture de Haralick représentent un ensemble de métriques calculées à partir de la matrice de cooccurrence pour inférer des informations pertinentes : le degré de désordre, l'orientation de la variation, le degré de contraste, le degré d'organisation de la texture, etc.

3.2.2.4 Matrice de cooccurrence

Matrice de cooccurrence des niveaux de gris (GLCM, Gray Level Co-occurrence Matrix) ou la matrice de l'espace moyen d'ordre 2, a été proposée par Haralick et son groupe en 1970s [Haralick 1973]. La matrice de cooccurrence est une matrice carrée où chaque élément M_{ij} dans la direction de la ligne (i) et la colonne (j) enregistre une fréquence d'occurrence relative à une paire de pixel. Cette matrice reflète l'information exhaustive sur la direction, l'intervalle de l'adjacence, le changement d'amplitude, etc. Pour un voisinage de (d_x, d_y) , la définition mathématique de la matrice de cooccurrence est donnée par la

formule suivante :

$$M(d_x, d_y)(i, j) = \sum_{p=1}^n \sum_{q=1}^m \begin{cases} 1 & \text{if } I(p, q) = i \wedge I(p+d_x, q+d_y) \\ 0 & \text{Otherwise} \end{cases} \quad (3.5)$$

(n, m) la dimension de la région d'intérêt (dans notre cas c'est la cible), (i, j) est l'échelle du niveau de gris de la cible, I_t est l'image en niveau de gris de la cible capturée au temps t . Cette matrice est définie sur un pair (D, O°) , où D est la distance entre deux pixels et O° est l'orientation du vecteur construit par deux pixels $O^\circ = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$. Dans ce travail nous avons choisi deux orientations 0° et 45° avec une distance entre deux pixels $D = 1$ par rapport au temps de calcul. La figure suivante (Fig.3.2) illustre deux matrices de cooccurrences calculées sur les deux régions d'intérêts extraites en utilisant la chaîne de traitement précédemment présentée dans le chapitre.2, section.2.4. Les séquences sont capturées par la camera E-CAM-1 de notre réseau multi-caméra.

La matrice de cooccurrence contient une masse d'informations très importante mais difficile à manipuler, ce qui la rend non directement utilisable mais plutôt à travers les indices de texture de Haralick [Iftene 2004]. Pour l'exploitation de cette matrice, Haralick et son groupe ont proposés quatorze indices correspondant à la description globale de la texture . Parmi ces indices, nous avons choisi six prouvés discriminants dans [Haralick 1973] et [Bousetouane 2013b], et qui sont les suivants : l'homogénéité, le contraste, l'entropie, la corrélation, la dis-similarité et le second moment angulaire.

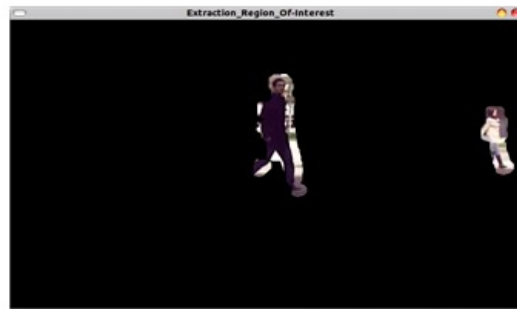
1. HOMOGENEITE : Plus on retrouve le même couple de pixels, plus la valeur de cet indice sera élevée. Par exemple : image uniforme ou texture périodique dans le sens de la translation :

$$HOM = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{P_{i,j}}{(1 + (i - j))} \quad (3.6)$$

où $P_{i,j}$ est le contenu de la matrice de cooccurrence, (i, j) sont les indices de la matrice de cooccurrence, N_g est la taille de la matrice. **2. CONTRASTE** : Une forte valeur traduit de grandes variations locales (frontières) :

$$CONT = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{P_{i,j}}{(i - j)^2} \quad (3.7)$$

3. ENTROPIE : Mesure le degré d'organisation de la texture (varie de



Extraction d'objets d'intérêts après la génération de la carte de soustraction étiquetée : Obj1 à gauche, Obj2 à Droite.

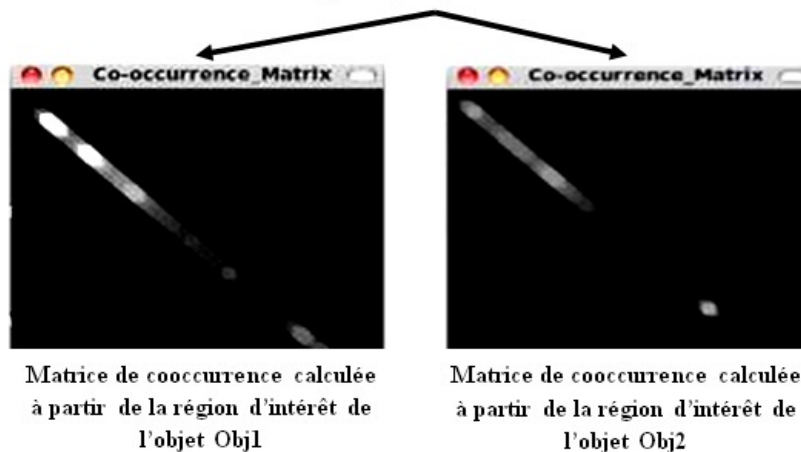


Fig. 3.2: Deux matrices de cooccurrence calculées à partir des régions d'intérêts (objets en mouvement) après la génération des cartes de soustraction binarisées et étiquetées.

façon opposé au second moment angulaire) :

$$ENT = - \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{i,j} * \log(P_{i,j}) \quad (3.8)$$

4. CORRELATION : 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 :

$$CORR = \frac{- \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - \mu) * (j - \mu) * P_{i,j}}{\sigma_x \sigma_y} \quad (3.9)$$

où, $\mu = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{i,j} * i$ et $\sigma = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{i,j} * (i - \mu)^2$
 μ est la moyenne, σ est l'écart type.

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

$$DISS = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{i,j} * (i - j) \quad (3.10)$$

6. SECONDE MOMENT ANGULAIRE (énergie) : 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 :

$$ASM = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P_{i,j}^2 \quad (3.11)$$

N_g est le niveau de gris de l'image I_t et $P_{i,j}$ sont les valeurs de la matrice de co-occurrence.

D'après nos connaissances, Les indices de Haralick ont été utilisés pour la première fois dans le domaine de suivi d'objets mobiles dans le cadre de cette thèse. Dans [Bousetouane 2011], nous avons proposés l'utilisation des indices de texture de Haralick pour la description des cibles afin d'améliorer la mise en correspondance inter-image basée régions pour le suivi d'objets. Dans [Bousetouane 2013b], nous avons prouvé que les indices de Haralick permettant d'inférer des informations invariantes inter-image sur la nature de la relation spatiale entre les pixels dans une régions d'intérêt (cible). Dans le même travail, les indices de haralick ont été combinés avec un modèle de distribution de la couleur afin d'améliorer l'algorithme de suivi Mean-shift (estimateur à noyau). La convergence de la nouvelle version du Mean-Shift est très remarquable et meilleur par rapport aux autres algorithmes.

3.2.3 Caractéristiques de Niveau Intermédiaire

L'extraction des caractéristiques de bas niveau peut ne pas être suffisante suffisantes pour arriver à une synthèse sur le contenu de l'image ou la vidéo, ce qui réduit la capacité des méthodes de suivi et de ré-identification inter et intra-caméra. Face à ce problème une solution largement utilisée dans la communauté consiste à analyser des sous-ensembles des pixels qui encodent des structures pertinentes (eg. les contours, les points d'intérêts ou les régions uniformes), où les pixels partagent la même couleur ou d'autres caractéristiques. Dans cette partie nous détaillons les caractéristiques (descripteurs) d'un niveau intermédiaire les plus utilisées dans les systèmes de suivi visuel modernes.

3.2.3.1 Les contours

Les détecteurs des contours produisent généralement des cartes binaires qui contiennent le/les contours présents dans une image. Les contours dans une image sont souvent associés aux réponses importantes de la dérivée-premier de l'image. La forme la plus simple de la détection des contours consiste en une simple opération de seuillage du gradient de l'image, la carte du contour $E(x, y)$ est calculée comme suit :

$$E(x, y) = \begin{cases} 1 & \text{Si } |\nabla I(x, y)| > T \\ 0 & \text{Dans un autre cas} \end{cases} \quad (3.12)$$

T est un seuil prédéterminé et $I(x, y)$ est l'image. Dans les systèmes de suivi, les contours sont largement utilisés surtout pour le codage des cibles humaines inter et intra-caméra dans les réseaux multi-caméra comme dans [Farenzena 2010]. 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.

3.2.3.2 Les points d'intérêts

Les détecteurs de points d'intérêts dans une image ou une séquence d'image visent à sélectionner des caractéristiques pertinentes d'une image. Ces caractéristiques peuvent être invariantes dans les autres séquences d'images où les conditions d'éclairage et de capture changent remarquablement. La plupart des détecteurs de points d'intérêts dans des images ont tendance à sélectionner des motifs locaux (local pattern) très distinctifs tels que les coins, les contours proéminents et les régions avec de texture distinctifs. Les détecteurs de points les plus utilisés dans les systèmes de suivi sont : détecteur de coins de Moravec [Deng 2007], détecteur de coins de Harris [Berg 2005], Les SIFT

[Serre 2007] et récemment les SURF [Bay 2006]. Malgré la large utilisation de ce type de caractéristiques dans le suivi et la ré-identification d'objets inter-caméra, ces caractéristiques restent couteuses dans des applications à temps réel. La mise en correspondance inter-image dans des séquences vidéo engendre une explosion combinatoire. Si nous modélisant un objet d'intérêt par 50 points d'intérêts dans une image à l'instant t , alors la mise en correspondance entre cet objet et le même objet à $t+1$ nécessite $50!$ Opérations de tests ($2,43 * 10^{18} \text{testes}$). Dans le cas où nous voulons suivre plusieurs objets alors la complexité de ce problème devienne exponentielle. Suite à l'exemple donné, nous constatons que l'utilisation des détecteurs de points pour la caractérisation des cibles n'est pas une solution valide pour le problème traité dans le cadre de cette thèse, où les conditions d'éclairages et de capture inter-caméra sont variantes.

3.2.4 Caractéristiques de Haut Niveau

L'extraction des caractéristiques de haut-niveau consiste en des opérations sur une image ou une séquence d'images afin de définir les objets d'intérêts. Ces opérations peuvent être un ensemble des techniques utilisées dans l'extraction des caractéristiques de niveau intermédiaire comme les points et les régions d'intérêts ou des détecteurs d'objets automatique. La détection d'objet d'intérêt peut être aussi effectuée par la modélisation de l'arrière-plan et la classification arrière-plan/avant-plan (voir le chapitre.2), ou par l'apprentissage d'un ensemble de classifieurs pour modéliser les informations représentatives d'une classe des cibles. Dans la littérature plusieurs approches ont été proposées afin d'extraire un ensemble d'objets d'intérêts à partir d'une image ou des séquences d'images. Dans cette section nous présentons les approches les plus utilisées.

3.2.4.1 Modélisation d'objet

La modélisation d'objets d'intérêts dans les systèmes de suivi est référée par plusieurs chercheurs comme étant une opération de suivi offline, où des détecteurs d'objets modélisant l'apparence d'une classe des cibles prédéfinies sont utilisés. Ces approches détectent un objet dans une scène si et seulement si l'apparence d'objet d'intérêt a été déjà appris off-line, ce qui rend ces approches limités dans des scénarios réels d'applications. Face à cette limite, des détecteurs d'objets basent sur l'apprentissage incrémentale ont été proposés dans [Wu 2007], où les chercheurs proposent la mise à jour de la base d'apprentissage au fil du temps, ce qui est complexe en traitement et en calcul. Dans les réseaux multi-caméra à champs disjoints ce type de détecteurs est lar-

gement utilisé comme dans [Meden 2011], où l'apparence des cibles circulent dans le réseau a été appris offline, ce qui crée un réseau fermé.

3.2.4.2 Modélisation de l'arrière-plan

La modélisation de l'arrière-plan est l'une des techniques les plus utilisées dans les systèmes de suivi pour la détection de mouvement et l'extraction d'objets d'intérêts. Cette technique a été brièvement détaillée dans le chapitre.2, section.2.2.3.

3.3 Etat de l'art sur la représentation des cibles

3.3.1 Taxonomie

La représentation des cibles dans les systèmes de suivi consiste en des modèles d'agrégation des caractéristiques pertinentes d'objets d'intérêts. Ces caractéristiques se basent sur l'apparence ou des indices spatio-temporelle. Les modèles de représentation des cibles contiennent des informations pertinentes sur la forme, l'apparence, et la temporalité des cibles. La qualité d'un modèle de représentation des cibles dépend de sa robustesse dans les conditions complexes telles que : le changement d'éclairage inter/intra-caméra, l'occlusion, l'encombrement et le désordre. Dans la littérature plusieurs modèles de représentation des cibles ont été proposés ; allant d'un simple histogramme de couleur ; vers des modèles plus élaborés comme les modèles articuloire et déformable. Vis-à-vis la multitude de ces modèles, classer et hiérarchiser ces modèles n'est pas une tâche facile. Dans cette partie, nous présentons une taxonomie possible des modèles proposés jusqu'à présent pour la représentation des cibles d'intérêts dans les systèmes de suivi visuel modernes.

3.3.2 La Forme

Comme illustré dans la taxonomie (fig.3.3), il existe plusieurs façons pour représenter la forme d'un objet d'intérêt. Nous divisons ces techniques en trois grands groupes : représentation basique, représentation articuloire et représentation déformable.

1. *Les modèles Basiques :*

Ces modèles se basent essentiellement sur des primitives géométriques simple comme les rectangles, les ellipses, cercle, points, etc. La forme la plus utilisée dans le suivi visuel est le rectangle qui sert à encadrer l'objet d'intérêt en respectant ses dimensions.

2. *Les modèles articuloires :*

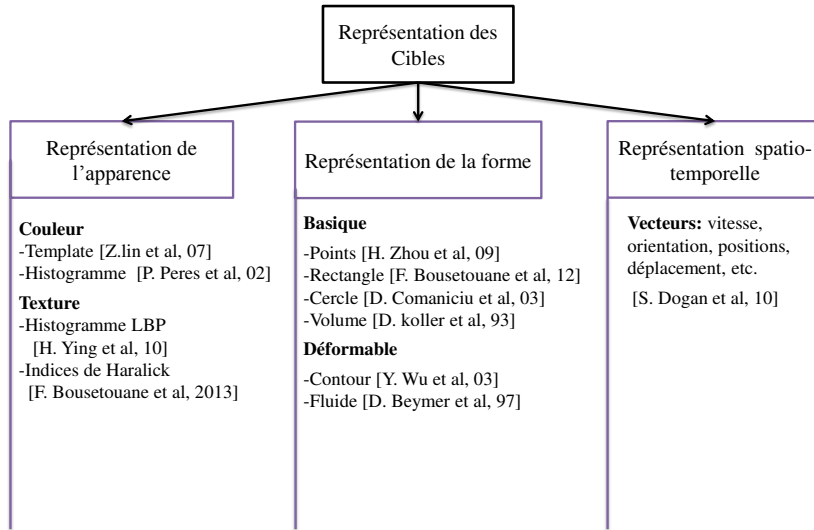


Fig. 3.3: Taxonomie possible des techniques utilisées jusqu'à présent pour la représentation des cibles

Ces modèles approximent la forme d'un objet par la combinaison d'un ensemble de modèle non-rigide (eg. points, région ou volume), et qui se basent sur des connexions topologiques et des contraintes de mouvement. Ces modèles sont largement utilisés pour le suivi et l'analyse du comportement humain.

3. Les 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 de fluides sont utilisés. Ces modèles se basent essentiellement sur des détecteurs de contours dans les images (voir la section.3.2.3.1).

3.3.3 L'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 inter-caméra 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 (en anglais, Histogram, Template).

3.3.3.1 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 [Pérez 2002] 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 $H_{RVB}(c_t)$ est calculé, la densité de couleur $q_u = u = 1..n$ qui estime la distribution de la couleur au temps t est approximée par :

$$q_u = w \sum_{d \in H_{rgb}(x_t)} K(x_d) \delta [b_t(x_d) - u] \quad (3.13)$$

δ est la fonction de Kronecker delta, x_d est la position de chaque pixel dans $H_{rgb}(c_t)$, w est une constante, K est un noyau d'Epanechnikov [Pérez 2002].

3.4 Mesures de similarités

Dans les systèmes de suivi, la mise en correspondance inter-objet aux différents temps dans des séquences d'images est une opération très importante. Cette opération permet l'estimation des positions d'objets et la recherche d'objets dans une autre caméra d'un réseau multi-caméra à champs disjoints. L'opération de la mise en correspondance inter-objet est assurée par des mesures de similarités qui dépendent non-seulement des types de données mais aussi du modèle de représentation des cibles. Comme nous l'avons déjà détaillé, pour représenter les cibles dans un système de suivi plusieurs modèles ont été proposés et pour mesurer la similarité entre les instances de ces modèles aux différents temps plusieurs métriques ont été utilisées. Ces métriques passent des plus simples, calculant la similarité entre deux simple vecteurs (comme la distance de Minkowski et euclidienne), jusqu'aux plus compliquées, calculant la similarité entre deux vecteurs multi-valué (comme la distance de Bhattacharya et la divergence de Kullback-Leibler). Ces distances sont souvent utilisées pour le calcul de la similarité entre deux histogrammes de couleur à deux moments consécutifs dans les systèmes de suivi.

3.4.1 Distance de Bhattacharya

La distance de Bhattacharya [Bousetouane 2013b] permet de mesurer la similarité entre deux distributions de probabilité discrète. Plus précisément,

elle calcule le cosinus entre deux distributions. Cette distance est directement reliée au coefficient de Bhattacharya, qui est une mesure statistique du recouvrement de deux ensembles d'échantillons [Julien 2007]. Dans le suivi et la ré-identification d'objets inter et intra-caméra, cette mesure est la plus utilisée pour la mise en correspondance entre deux observations basées sur l'histogramme de couleur. Soit deux histogrammes :

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

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

La similarité entre ces deux histogrammes calculée par la distance de Bhattacharya est donnée par :

$$d(y) = \sqrt{1 - \rho[p(y), q]} \tag{3.14}$$

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

3.4.2 Divergence de Kullback-leibler

La divergence de Kullback-leibler ou l'entropie relative est une métrique inventée par deux cryptanalyses [Runnalls 2007], qui permet de mesurer la dissimilarité entre deux distributions de probabilités.

La dissimilarité entre les deux histogrammes q_u et p_u calculée par la distance Kullback-leibler est donnée par :

$$K_d(q_y, p_u) = \int q_u \log\left(\frac{q_u}{p_u}\right) \tag{3.15}$$

Malgré l'utilisation étendue de ces deux distances pour le suivi et la ré-identification d'objets intra et inter-caméra, certains chercheurs [Hero III 2002] ont prouvé la faiblesse de ces métriques par rapport à leurs capacité discriminatoire. Cette insuffisance est causée par l'exploitation directe de l'intensité des pixels et l'absence du contexte et des relations entre les pixels (spatiales ou temporelles).

3.5 Modèle de représentation des cibles proposé

Dans la littérature et comme nous l'avons déjà détaillé dans le premier et le troisième chapitre, les caractéristiques utilisées jusqu'à présent pour la représentation des cibles restent limitées dans les conditions réelles d'application. Ces limites sont essentiellement causées par l'utilisation des modèles de représentation des cibles calculés à partir des pixels isolées comme la couleur.

L'utilisation directe des informations sur l'apparence des cibles calculées à partir des pixels isolées rend ces caractéristiques insuffisantes et souvent invalides dans des conditions réelles d'application, généralement en présence de changement d'éclairage inter-camera et inter-image comme : les déformations locales, changement d'échelle, encombrement, occultations complexes, etc. Face à ces problèmes, nos travaux sont focalisés à la découverte d'une description visuelle invariante des cibles inter-caméra et inter-image et discriminante inter-objet. Pour arriver à cet objectif nous devons augmenter le niveau de la description des cibles en exploitant la relation ou la dépendance spatiale entre les pixels composant la même cible. Cette manière de description des cibles permet d'inférer des informations pertinentes, comme : la variation locale, le degré d'organisation de la texture d'une cible, la direction des contours, le contexte de la couleur, etc. L'opération d'exploitation des relations entre les pixels pour la description des cibles d'intérêts est nommée par plusieurs chercheurs *Information Contextuelle de bas niveau*, "en anglais, Low-level Visual Contexte ". Les informations contextuelles de bas niveau ont été largement utilisées dans plusieurs domaines pour la description d'objets d'intérêts tel que : la reconnaissance d'objets [Oliva 2007], la classification d'images satellitaire [Qi 2010], l'estimation de mouvement [Wu 2009] et récemment et dans le cadre de cette thèse ce genre d'indices ont été utilisés pour le suivi [Bousetouane 2013b] et la mise en correspondance inter-image par [Bousetouane 2011]. Dans cette thèse nous proposons un modèle de description des cibles invariant inter-image et inter-caméra et discriminant inter-objet, ce modèle se base sur une combinaison des caractéristiques de haut niveau et des caractéristiques de bas niveau.

- Premièrement, les caractéristiques de haut niveau sont utilisées pour l'extraction d'objets d'intérêts où la chaîne de traitement proposée dans le deuxième chapitre est utilisée pour la détection d'objets en mouvement.

- En suite et après la génération de la carte de soustraction binarisée et étiquetée, nous utilisons des caractéristiques de bas niveau pour la description des régions d'intérêts.

Le modèle proposé pour la description de l'apparence des cibles se base sur une mixture ad-hoc des informations contextuelles calculées à travers des indices de texture de Haralick et quatre moments statistiques. Ce nouveau modèle est décrit comme suit :

$$M(e_t) = (Hom(p_i, t), Corr(p_i, t), Diss(p_i, t), Ent^2(p_i, t), Sma(p_i, t), Cont(p_i, t), Ent^1(R_{obj(x,y,t)}, t), Var(R_{obj(x,y,t)}, t), Sk(R_{obj(x,y,t)}, t), Ku(R_{obj(x,y,t)}, t)) \quad (3.16)$$

$$\begin{aligned}
 M(e_{t+1}) = & (Hom(p_i, t + 1), Corr(p_i, t + 1), Diss(p_i, t + 1), Ent^2(p_i, t + 1), \\
 & Sma(p_i, t + 1), Cont(p_i, t + 1), Ent^1(R_{obj(x,y,t+1)}, t + 1), Var(R_{obj(x,y,t+1)}, t + 1), \\
 & Sk(R_{obj(x,y,t+1)}, t + 1), Ku(R_{obj(x,y,t+1)}, t + 1))
 \end{aligned} \tag{3.17}$$

Tel que :

-*Obj*, l'objet d'intérêt qui va être suivi inter et intra-caméra dans des séquences d'images I_t où $0 < I_t < \infty$.

- $R_{obj}(x, y)$ est la région d'objet d'intérêt *Obj* extraite automatiquement par la génération de la carte de soustraction binarisée et étiquetée (voir chapitre deux).

- $M(e_t)$ et $M(e_{t+1})$ sont deux instances du modèle d'apparences proposé calculées dans la région $R_{obj}(x, y)$ à partir de deux images consécutives capturées au temps t et au temps $t + 1$.

La mise en correspondance entre l'instance du modèle proposé $M(e_t)$ calculé au temps t , et l'instance du modèle proposé du même objet $M(e_{t+1})$ calculée au temps $t + 1$ pour le suivi et la ré-identification d'objet inter et intra-camera, nécessite la mesure de la similarité entre les deux instances. Pour aboutir à cet objectif nous avons choisi l'utilisation de la distance de Mahalanobis [Bousetouane 2013b], qui est définie comme suit :

$$\delta_2(M(e_t), M(e_{t+1})) = \sqrt{(M(e_t) - M(e_{t+1}))^t S (M(e_t) - M(e_{t+1}))} \tag{3.18}$$

$S = \frac{1}{n} \sum_{i=1}^n ((M(e_t)_i - u)((M(e_{t+1}) - u)^t$, $u = \frac{1}{n} \sum_{i=1}^n M(e_t)_i$, u est le vecteur moyen, S est la matrice de covariance de taille n .

3.6 Evaluation

L'évaluation des modèles de représentation des cibles est une étape très importante, cela afin de tester la capacité représentative et discriminatoire de ces modèles en présence des conditions complexes tel que : le changement d'échelle, variation d'éclairage inter et intra-caméra, encombrement, occultations, les ombres, déformation locale, etc. L'évaluation de la performance du modèle de représentation de l'apparence d'objets d'intérêts (cibles) proposé dans cette thèse, a nécessité le test dans différentes conditions et scenarios. La figure (Fig.3.4) illustre un exemple des séquences d'images utilisées dans l'évaluation.

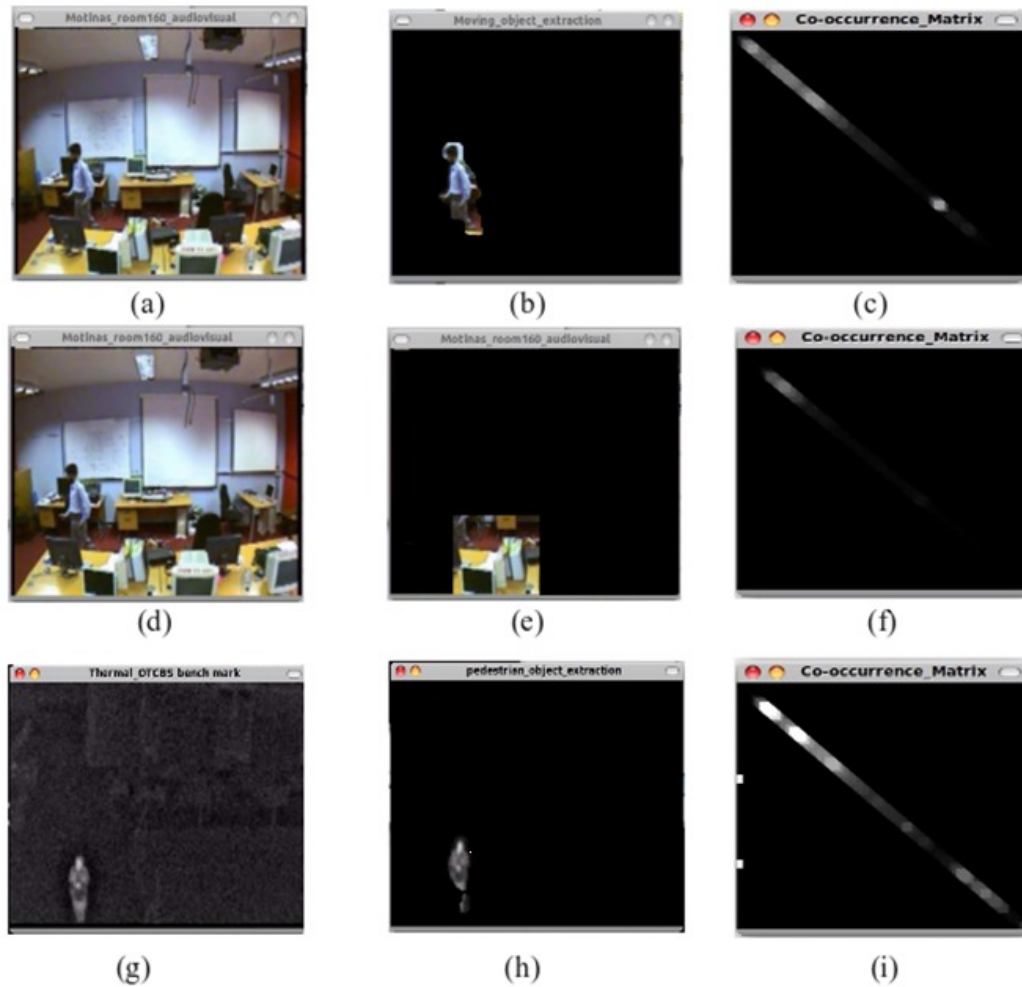


Fig. 3.4: Exemple des séquences d'images utilisées dans l'évaluation. (a),(b),(c) sont respectivement l'image numéro 200 de la séquence video du corpus de test SPEVI, extraction d'objet d'intérêt et la matrice de cooccurrence de l'objet d'intérêt. (d),(e),(f) sont respectivement l'image numéro 100 de la séquence video du corpus de test SPEVI, extraction d'objet non-important et la matrice de cooccurrence de cet objet. (g),(h),(i) sont respectivement l'image numéro 80 de la séquence video du corpus de test OSU capturée pas des capteur infrarouge, extraction d'objet d'intérêt et la matrice de cooccurrence de l'objet d'intérêt.

3.6.1 L'invariance inter-image et intra-caméra

Pour évaluer l'invariance du modèle proposé inter-image et intra-camera, nous avons calculé cinq instances du modèle proposé à partir de cinq images consécutives d'une région d'intérêt (cible). La figure (Fig.3.4 (a)) illustre la scène utilisée dans cette évaluation. Le tableau suivant (Tab.3.1) illustre les valeurs des cinq instances consécutives du modèle de représentation des cibles proposé et l'erreurs dans la similarité entre les instances calculées à partir des images de la même cible à différents temps.

A partir de ce tableau (Tab.3.1), nous remarquons que : Les valeurs des différentes instances du modèle proposé pour la représentation d'objets d'intérêts à différent temps ($t = 1, \dots, t = 5$) sont très proches, où l'erreur dans la similarité entre deux instances consécutives ne dépasse pas la valeur "0,27", ce qui prouve l'invariance du modèle proposé inter-image et intra-camera.

3.6.2 Capacité discriminatoire du modèle proposé inter-Objet

Pour évaluer la capacité discriminatoire inter-objet du modèle proposé, des instances du modèle d'apparence proposé ont été calculées à partir de deux régions d'intérêts contenant deux objets différents (Fig.3.4) à partir de quatre images consécutives. Le tableau (Tab.3.2) illustre les valeurs des instances du modèle proposé et la similarité entre ces instances dans quatre images consécutives.

A partir de ce tableau (Tab.3.2), nous remarquons que : Les valeurs de l'instance du modèle proposé calculées dans la région d'intérêt de l'objet Obj_1 (Fig.3.4(b)) à différents temps ($t = 1, , t = t + 2$) sont très différentes par rapport aux valeurs de l'autre objet Obj_2 (Fig.3.4(c)), ce qui prouve la capacité discriminatoire du modèle proposé inter-objet.

3.6.3 L'invariance aux conditions d'éclairages inter-Camera

Pour évaluer l'invariance du modèle proposé aux changements des conditions d'éclairages inter-caméra, plusieurs instances du modèle proposé ont été calculées à partir de quatre régions d'intérêts contenant le même objet dans quatre séquences d'images différentes capturées par quatre caméras différentes. Le tableau (Tab.3.3) illustre les valeurs des instances du modèle proposé dans quatre vues différentes et l'erreur dans la similarité entre les quatre instances du même objet dans quatre vues différentes. A partir de ce tableau (Tab.3.3), nous remarquons que : Les valeurs des différentes instances du modèle proposé calculées dans quatre emplacements différents sont très proches. L'erreur dans

la similarité entre ces instances est très petite, ce qui prouve l'invariance du modèle proposé aux conditions d'éclairage inter-caméra.

3.6.4 L'invariance aux autres modalités de captures

Pour évaluer l'invariance du modèle proposé par rapport aux autres modalités de captures (eg. séquences d'images infrarouge Fig.3.4(g)), plusieurs instances du modèle d'apparence proposé ont été calculées à partir de cinq images infrarouges consécutives d'une région d'intérêt (cible) (Fig.3.4(h)). Le tableau (Tab.3.4) illustre les valeurs de ces instances et les erreurs dans la similarité calculées entre les instances.

TABLE 3.1: Erreurs dans la similarité calculées entre cinq instances du modèle proposé a partir de 5 images consécutives (12 FPS) d'un seul objet (les valeurs sont normalisées).

Temps	HOM	CONT	CORR	DISS	ASM	ENT	E_1	VAR	SK	KU	% Erreurs (Similarité)
t	3.436	2.868	0.839	1.064	0.095	4.048	2.327	3.313	0.035	2.963	0
t+1	3.689	3.043	0.591	1.138	0.087	4.232	2.435	3.458	0.043	2.876	0.091
t+2	3.235	2.120	0.683	1.452	0.075	4.595	2.672	4.022	0.023	3.120	0.122
t+3	4.213	2.782	0.912	1.230	0.051	4.432	2.293	4.324	0.015	3.252	0.278
t+4	4.087	2.211	0.794	1.471	0.068	4.623	2.972	4.126	0.058	3.086	0.181

TABLE 3.2: Erreurs dans la similarité calculées entre quatre instances du modèle proposé a partir des images de deux objets différents (12 FPS) d'un seul objet (les valeurs sont normalisées)

Temps	Objs	HOM	CONT	CORR	DISS	ASM	ENT	E_1	VAR	SK	KU	% Erreurs (Similarité)
t	<i>Obj₁</i>	3.436	2.868	0.839	1.064	0.095	4.048	2.327	3.313	0.035	2.963	0
t	<i>Obj₂</i>	2.542	1.122	1.532	2.346	0.192	6.343	3.769	4.767	0.263	4.862	0.786
t+1	<i>Obj₁</i>	3.237	2.731	0.682	1.454	0.075	4.596	2.672	4.023	0.023	3.120	-
t+1	<i>Obj₂</i>	2.145	1.097	2.092	2.591	0.286	6.765	4.192	5.133	0.184	5.212	0.860
t+2	<i>Obj₁</i>	4.087	2.213	0.799	1.475	0.068	4.621	2.973	4.121	0.058	3.086	-
t+2	<i>Obj₂</i>	2.322	1.373	1.892	2.412	0.345	5.879	3.936	5.473	0.322	4.689	0.630

TABLE 3.3: Valeurs des instances du modèle de représentation des cibles proposé, calculées a partir des images de quatre vues différentes d'un seul objet (les valeurs sont normalisées)

Temps	Caméras	HOM	CONT	CORR	DISS	ASM	ENT	E ₁	VAR	SK	KU	% Erreurs(Similarité)
t	<i>Cam₁</i>	2.870	2.867	0.734	1.537	0.142	3.672	2.623	4.326	0.391	3.542	0
t+100	<i>Cam₂</i>	3.232	2.733	0.864	2.172	0.209	4.098	2.671	4.025	0.223	3.872	0.203
t+500	<i>Cam₃</i>	3.690	3.122	0.941	2.125	0.301	4.141	2.836	4.193	0.452	3.124	0.178
t+897	<i>Cam₄</i>	3.737	2.953	0.621	2.433	0.285	3.820	2.719	3.924	0.371	3.340	0.119

TABLE 3.4: Valeurs des instances du modèle de représentation des cibles proposé, calculées a partir des images infrarouge consécutives d'un seul objet (les valeurs sont normalisées)

Temps	HOM	CONT	CORR	DISS	ASM	ENT	E ₁	VAR	SK	KU	% Erreurs (Similarité)
t	4.612	2.868	0.437	1.493	0.113	3.984	2.121	4.093	0.072	3.092	0
t+1	4.584	3.043	0.594	1.331	0.087	4.131	2.540	3.854	0.043	2.976	0.108
t+2	3.988	2.125	0.686	1.457	0.075	4.495	2.472	4.021	0.023	3.129	0.150
t+3	4.210	2.786	0.813	1.233	0.095	4.232	2.393	4.223	0.015	3.153	0.089
t+4	4.642	2.213	0.794	1.470	0.075	4.123	2.171	4.127	0.048	3.086	0.208

A partir de ce tableau (Tab.3.4), nous remarquons que : Les valeurs des différentes instances du modèle d'apparence proposé pour la représentation d'objets d'intérêts à différent temps ($t = 1, \dots t = 5$) sont très proches. L'erreur dans la similarité entre deux instances consécutives ne dépassent pas la valeur de 0,2, ce qui prouve l'invariance de notre modèle dans d'autres modalités de capture.

A travers ces expérimentations nous avons prouvé que le modèle proposé pour la représentation de l'apparence d'objets d'intérêts (cibles) est invariant inter-image, intra-caméra, inter-caméra, dans d'autres modalités de capture et discriminant inter-objet. Vu que ce modèle permet d'exploiter des propriétés internes spécifiques pour chaque objet d'intérêt (tel que le degré de désordre de la texture, variation locale, la relation entre les pixels, degré d'homogénéité, les moments statistiques, etc), cela prouve la capacité de notre modèle à mieux caractériser les cibles d'intérêts que la couleur et d'autres descripteurs cités dans l'état de l'art. Dans le chapitre 5 nous présentons l'utilisation du modèle proposé basé sur l'apparence pour la description des cibles dans le processus de la ré-identification d'objets inter-caméra.

3.7 Conclusion

Dans ce chapitre nous avons préalablement présenté un état de l'art sur les méthodes proposées jusqu'à présent pour l'extraction des caractéristiques dans les systèmes de suivi visuels modernes. Ensuite nous avons illustré une taxonomie possible des modèles utilisés pour la représentation des cibles en mouvement. Malgré l'utilisation ré pondue des histogrammes de couleurs et la temporalité pour la représentation des cibles, ces modèles restent limités dans les conditions réelles d'applications, telles que le changement d'éclairages intra et inter-caméra. Les limites des modèles classiques nous mène à proposer un nouveau modèle de représentation de l'apparence d'objets d'intérêts (cibles). La troisième partie de ce chapitre a été réservée à la présentation du modèle que nous avons proposé pour la représentation des cibles et qui se base sur une mixture des moments statistiques et des indices de textures de Haralick. Le modèle proposé a été évalué dans des conditions réelles d'applications et les résultats obtenus prouvent sa capacité discriminatoire inter-objet et sont invariance inter et intra-caméra. Dans le chapitre suivant nous présentons l'intégration du modèle d'apparence proposé dans un filtre particulière et cela afin d'améliorer l'estimation des positions d'objets intra et inter-caméra dans des conditions complexes.

Suivi d'objets Mobiles intra-caméra

Sommaire

4.1	Introduction	82
4.2	Etat de l'art sur le suivi d'objets mobiles	83
4.2.1	Taxonomie	83
4.2.2	Méthodes déterministes d'estimation des positions	83
4.2.3	Méthodes stochastiques : Filtrage récursif d'états	88
4.3	Filtre particulière adaptatif proposé	93
4.3.1	Indices visuels et modèle de Vraisemblance	95
4.3.2	Suivi Multi-objet	99
4.4	Résultats et Evaluation	101
4.4.1	Résultats de suivi Mono/Multi-Objet	101
4.4.2	Evaluation	107
4.5	Conclusion	108

DANS ce chapitre, nous présentons tout d'abord quelques définitions sur le problème de suivi visuel d'objets mobiles dans des séquences d'images. Ensuite, nous présentons un état de l'art sur les techniques proposées jusqu'à présent pour le suivi et l'estimation des positions d'objets mobiles dans des systèmes mono-caméra. Nous introduisons ensuite l'approche proposée dans le cadre de cette thèse pour le suivi et l'estimation des trajectoires d'objets d'intérêts dans des séquences d'images. L'approche proposée repose sur l'intégration du modèle de représentation des cibles proposé (voir chapitre.3, section.3.5) dans un filtre particulière afin d'améliorer le suivi inter/intra-caméra dans des conditions complexes. Enfin, nous consacrons une section aux résultats d'application de notre approche sur des séquences vidéo capturées par des caméras composant notre réseau multi-caméra (Chapitre.1 (Fig.1.3)).

4.1 Introduction

Aujourd'hui, la capture des vidéos devient une tâche très aisée avec la maîtrise des techniques d'acquisitions et l'évolution des capteurs. Les machines qui ont la capacité de voir et de comprendre partiellement leurs environnements existent, et leurs développements sont très accélérés dernièrement avec les avancées technologiques remarquables dans les capteurs nano/micro-électronique et dans les algorithmes d'analyse des vidéos. Avec tous ces progrès, actuellement plusieurs opportunités sont ouvertes pour le développement des applications riches dans différents domaines tel que : la surveillance, la navigation robotiques visuelle, l'interaction homme machine, etc. Un processus fondamental et essentiel pour que les machines puissent voir et analyser leurs environnements est le suivi des objets mobiles. Le processus d'estimation des positions d'objets d'intérêts au fil du temps utilisant une caméra est appelé le suivi ou le tracking d'objets. Ce processus est la brique de base de toutes les applications liées au suivi visuel (suivi distribué, suivi dans un réseau multi-caméra, etc.). Le suivi visuel multi-objet intra-caméra (monoculaire) est un problème très complexe, où jusqu'à présent aucun algorithme de suivi n'est valide dans toutes les conditions réelles d'application. Cette complexité est causée par :

- 1.** La variation de l'apparence d'objet inter-image où la signature visuelle de l'objet d'intérêt change au fil du temps.
- 2** La ressemblance (couleur, forme, vitesse, etc.) entre un objet d'intérêt et un autre objet mobile ou entre l'objet d'intérêt et un autre composant de la scène, ce qui engendre une interférence avec l'observation de l'objet d'intérêt. Ce phénomène est nommé "encombrement ou désordre".
- 3.** Le changement de la pose de capture, par exemple dans un cas de rotation.
- 4.** L'éclairage ambiant, où la signature visuelle d'un objet change avec le changement de l'environnement de capture, par exemple : la couleur d'un objet dans une salle est totalement différente de la couleur du même objet à l'extérieur de cette salle (indoor and outdoor capture).
- 5.** Le bruit, ce phénomène est en relation avec le type et la qualité du capteur.
- 6.** Occultations, l'observation d'un objet peut être cachée par un autre objet ou un autre composant de la scène. Deux types d'occultations dans une scène sont possibles :

(a) Occultation partielle, quand un objet déplace devant un mur, une table et d'autre objet statiques. Ce type d'occultation affecte uniquement une petite partie de l'observation de l'objet d'intérêt.

(b) Occultation complète affecte toute la région de l'objet d'intérêt où l'observation de l'objet disparaît complètement certain temps.

7. Changement d'échelle d'objets d'intérêts et mouvement non-rigide, etc.
8. Manque d'information sur les paramètres intrinsèque et extrinsèque du capteur.

En tenant compte de ces complications, le suivi visuel d'objets mobiles intra-caméra dans des conditions réelles d'applications reste un grand défi et au cœur de la recherche en vision par ordinateur. Dans la littérature plusieurs algorithmes ont été proposés pour le suivi mono/multi-objet à partir des séquences d'images. Dans la section suivante nous présentons une taxonomie possible des méthodes proposées jusqu'à présent pour le suivi et l'estimation des trajectoires d'objets d'intérêts mobiles.

4.2 Etat de l'art sur le suivi d'objets mobiles

4.2.1 Taxonomie

Depuis les années 80s, plusieurs efforts ont été réalisés et plusieurs méthodes ont été publiées pour l'estimation des trajectoires d'objets d'intérêts dans des scènes complexes ; allant des estimateurs à noyaux et de la mise en correspondance inter-régions ; vers le filtrage récursif d'états et le suivi par détection. Vis-à-vis la multitude des méthodes proposées dans la littérature pour le suivi visuel intra-caméra d'objets en mouvement, classer ces méthodes n'est pas une tâche facile. Les états de l'art proposés dans [Yilmaz 2006], [Leibe 2007], nous permettent d'établir une taxonomie (classification) possible des méthodes proposées jusqu'à présent pour le suivi d'objets en mouvement (mobiles). Nous distinguons deux grandes familles des méthodes ; les méthodes déterministes et les méthodes stochastiques. La figure suivante (Fig.4.1) illustre une taxonomie possible des méthodes essentielles proposées jusqu'à présent pour le suivi d'objets en mouvement.

4.2.2 Méthodes déterministes d'estimation des positions

Ces méthodes sont souvent considérées par les chercheurs comme des méthodes basées sur le gradient (Gradient-based trackers). Ce type des méthodes utilisent les caractéristiques de l'image pour converger vers les meilleures positions des cibles au fil du temps dans des séquences d'images. Étant donné les caractéristiques extraites à partir de l'image courante et une fonction d'évaluation de l'état courant, ces méthodes raffinent itérativement l'état estimé à travers la convergence vers le maximum global. Les algorithmes les plus connus

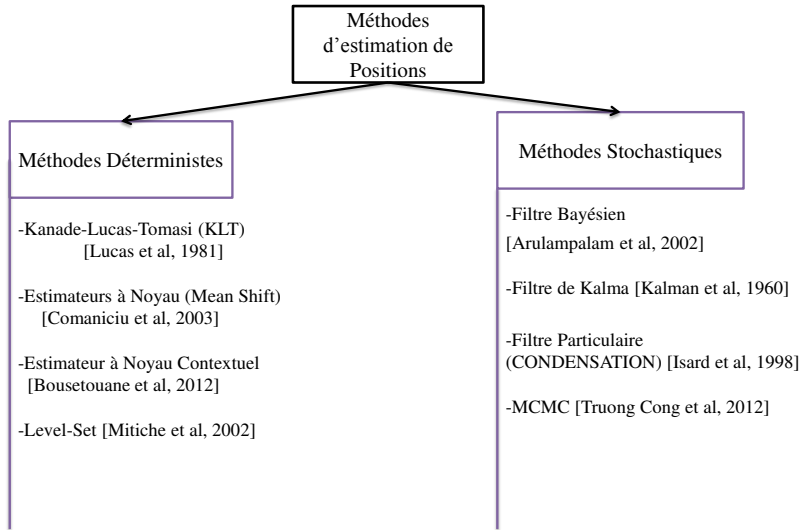


Fig. 4.1: Taxonomie possible des techniques proposées jusqu'à présent pour le suivi et l'estimation des positions d'objets mobiles.

dans les méthodes déterministes sont : KLT, l'estimateur à noyau (Mean-Shift) et Level-Set, dans cette partie nous détaillons ces méthodes.

4.2.2.1 Kanade–Lucas–Tomasi

Cette méthode (*KLT*) [Takacs 2010], 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. Le *KLT* utilise la notion de la régression linéaire multiple pour l'estimation des paramètres du vecteur de déplacement temporelle, calculé par le flot optique (voir chapitre.2, section.2.2.2.2). La zone de la cible d'intérêt est représentée par une fenêtre carrée de taille $N=(2w-1)\times(2w-1)$, où w est la largeur du template. Le problème de suivi peut être défini comme l'estimation de mouvement de la transition de l'objet d'un état à un autre, où l'état est défini comme suit :

$$x_t = (u_t, v_t) \quad (4.1)$$

(u, v) est le vecteur de déplacement. Supposant que les coordonnées du Template (patron) $I_t(\cdot)$ dans le plan image sont 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 :

$$X_t = x_{(0)t} + \Delta x_t \quad (4.2)$$

Δx_t est le petit déplacement de la cible ajouté au déplacement précédent, en outre imposant la contrainte de l'éclairage constante [Takacs 2010], la différence de l'apparence entre le Template (patron) et la fenêtre centrée en x_k est donnée comme suit :

$$I_k(w) = I_t(w - x_k) + n_k(w) = I_t(w - x_{(0)t} - \Delta x_t) + n_k(w) \quad (4.3)$$

w est la position d'un pixel dans l'image et $n_k(w)$ est le modèle de bruit additif dans les valeurs des pixels. Le problème de suivi utilisant cette méthode se focalise sur la recherche d'un déplacement minuscule δx_t qui minimise l'erreur entre la zone de la meilleur position estimée x_k et le Template :

$$\varepsilon(\Delta x_t) = \sum_{|w-x_k|<W} [(I_t(w - x_k^0) - \Delta x_t) - I_t(w)]^2 \quad (4.4)$$

Pour un petit déplacement de la cible δx_t , nous pouvons approximer le template $I_t(\cdot)$ avec son série de Taylor centré autour de x_{t-1} par :

$$I_t(w - \hat{x}_k^{(0)} - \Delta x_t) \approx I_t(w - x_t^{(0)}) + b' \Delta x_t \quad (4.5)$$

b' est le transposé du gradient du template, donné par :

$$b = \frac{\partial I_T(w - \hat{x}_k^{(0)})}{\partial w} \quad (4.6)$$

La complexité de cette méthode dépend du type de la transformation temporelle de la cible ou explicitement du modèle de déplacement. Si le déplacement est linéaire et la cible rigide, l'estimation de la meilleur position de la cible au temps t revient à minimiser l'équation quadratique. S'il s'agit d'une cible non-rigide (cible humaine) le modèle de déplacement est non-linéaire, alors d'autres facteurs doivent être rajoutés à cette méthode comme dans [Živković 2004]. La contrainte de la non-rigidité des cibles complique la convergence de cette méthode. Malgré la large utilisation de cette approche dans l'état de l'art [Benfold 2011] pour l'estimation de mouvement, cette dernière souffre de sérieux problèmes comme : la variation de la couleur, la non-rigidité des cibles, le mouvement saccadé, etc. Vis-à-vis les insuffisances d'application de cette méthode dans des conditions réelles d'application, d'autres chercheurs [Bousetouane 2013b] proposent la description des cibles par des modèles non-paramétriques (histogrammes de couleur) afin de résoudre le problème de la non-rigidité des cibles et le déplacement non-linéaire. Cette méthode est connue sous le nom « estimateur a noyau, en anglais, Kernel Density Estimation » ou encore « Mean-Shift ».

4.2.2.2 Estimateur à Noyau "MeanShift"

Le Mean-Shift est une technique d'optimisation utilisée dans plusieurs applications. Le Mean-Shift estime le taux de déplacement du noyau dans les séquences d'images. Le noyau est la primitive géométrique qui entoure la cible (objet d'intérêt), il peut être sous forme rectangulaire, elliptique, articulaire, etc. L'algorithme de suivi par Mean-shift [Comaniciu 2000] se base essentiellement sur deux étapes : (1) la description des cibles par un modèle de couleur non-paramétrique (Histogramme).(2) la convergence du Mean-shift, où une recherche itérative de la nouvelle position de l'objet dans l'image courante est effectuée à travers la maximisation d'une métrique de similarité entre deux histogrammes de couleurs (Histogramme candidat calculé au temps t dans les positions temporaires et l'histogramme de référence calculé au temps $t - 1$). La métrique la plus utilisée est la distance de Bhattacharya, (equation.3.4.1, chapitre.3) .

L'algorithme de suivi "Mean-Shift" se décompose en six étapes séquentielles [Bousetouane 2013b] :

1. Initialisation de l'algorithme par la position initiale de l'objet d'intérêt (cible), cette étape pourra être réalisée manuellement par l'utilisateur ou automatiquement utilisant un détecteur d'objet (voir le chapitre.2).

2. Le calcul de la distribution de la couleur (histogramme) dans cette région d'intérêt utilisant l'équation.3.13.

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

a. Pour chaque pixel, le poids est calculé par l'équation suivante :

$$\begin{aligned} w_i &= \sum_{u=1}^M \frac{q_u}{\sqrt{p_u(y_0)}} g[b(x_i) - u] \\ g(x) &= -k'(x). \end{aligned} \quad (4.7)$$

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. Calculer la nouvelle position temporaire de la cible dans l'image courante :

$$y_1 = \frac{\sum_{i=1}^N x_i w_i g\left(\left\|\frac{y_0 - w_i}{h}\right\|^2\right)}{\sum_{i=1}^N w_i g\left(\left\|\frac{y_0 - w_i}{h}\right\|^2\right)} \quad (4.8)$$

où h est la taille de la fenetre, x_i est la position temporaire de la cible.

c. Calculer l'histogramme de la couleur dans la nouvelle position temporaire utilisant l'équation.3.13.

4. Calculer la similarité entre l'histogramme de couleur de référence et l'histogramme de couleur candidat utilisant la distance de Bhattacharya (équation.3.4.1, chapitre.3).

Une approximation est possible de cette distance utilisant la série de Taylor autour de l'histogramme de couleur candidat, cette approximation est donnée par :

$$\rho[p(y), q] \approx \frac{1}{2} \sum_{u=1}^N \sqrt{p_u(y_0)q_u} + \frac{1}{2} \sum_{u=1}^N p_u(y) \sqrt{\frac{q_u}{p_u(y_0)}} \quad (4.9)$$

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 [Leichter 2010], [Bousetouane 2013b], [Ju 2010], affirment que la combinaison de plusieurs indices visuels peuvent 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. Dans cette thèse, nous proposons une nouvelle idée de la description des cibles par des indices visuels contextuels [Bousetouane 2013b], où une combinaison entre la distribution de couleur et les relations spatiales invariantes de bas-niveau entre les pixels a été proposée afin d'améliorer la convergence de cet algorithme. Ceci nous mène à créer une nouvelle version d'estimateur à noyau nommée Estimateur à Noyau Contextuel, « Contextual Kernel density estimation ». Les résultats obtenus par notre approche sont plus précis que ceux du MeanShift Classique, voir notre publication [Bousetouane 2013b]. La même mixture des indices visuels que nous avons proposé dans [Bousetouane 2013b] (voir chapitre.3, section.3.5) est intégrée dans la nouvelle version du Mean-Shift. La figure suivante (Fig.4.2) illustre quelques résultats d'application de notre algorithme pour le suivi d'objet en mouvement utilisant une simple webcam. Cette expérimentation a été réalisée au niveau de département d'informatique de l'université d'Annaba en présence de plusieurs obstacles, chan-

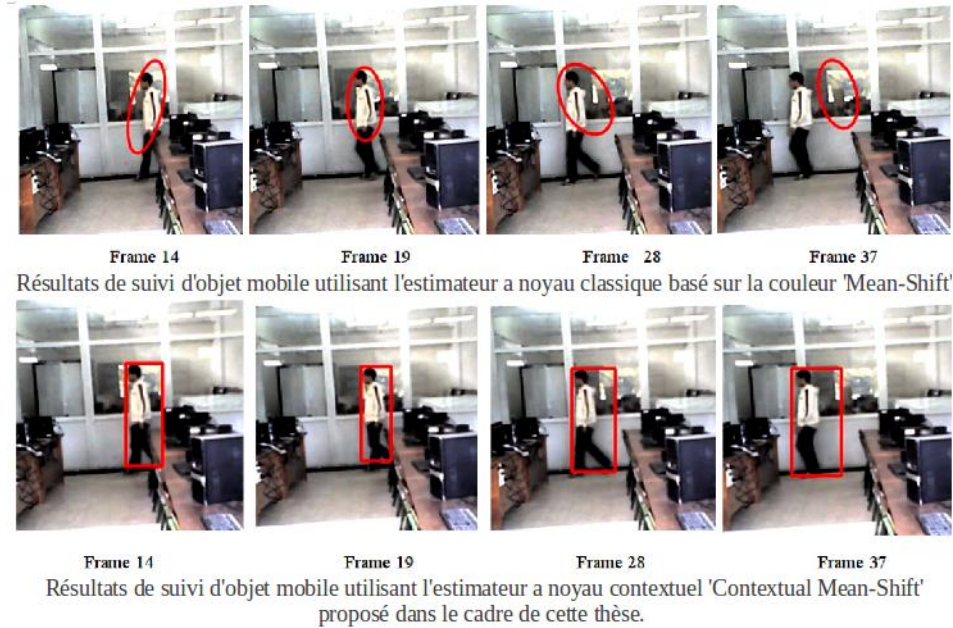


Fig. 4.2: Résultats de suivi d'objet mobile utilisant le mean-shift proposé et le mean-shift classique basé sur la couleur

gement de la lumière, objet non-rigide, rotation, etc. L'estimateur à noyau contextuel proposé dans le cadre de cette est brièvement détaillé dans notre publication [Bousetouane 2013b].

4.2.3 Méthodes stochastiques : Filtrage récursif d'états

Les méthodes stochastiques ou le filtrage récursif d'états pour le suivi d'objets en mouvement ont récemment attirées l'attention de plusieurs chercheurs par rapport à leur rapidité et leur extensibilité dans la description des cibles. Ces méthodes se basent essentiellement sur l'inférence bayésien, où l'estimation de la position courante d'objet d'intérêt revient à estimer la probabilité à postériori que l'objet dans la position x au temps t connaissant l'indice visuel de cet objet au temps $t - 1$. Dans cette partie, nous détaillons les principes de base des méthodes stochastique. Premièrement, nous décrivons le filtrage bayésien récursif et son implémentation dans un contexte linéaire et gaussien utilisant le filtre de KALMAN. En suite, nous présentons une des méthodes les plus répandues qui est le filtrage particulaire ou Monte-Carlo séquentielle.

4.2.3.1 Filtre Bayésien récursif

Le filtre bayésien récursif [Arulampalam 2004], [Bousetouane 2013a] est un algorithme de filtrage d'état pour l'estimation de la dynamique des systèmes.

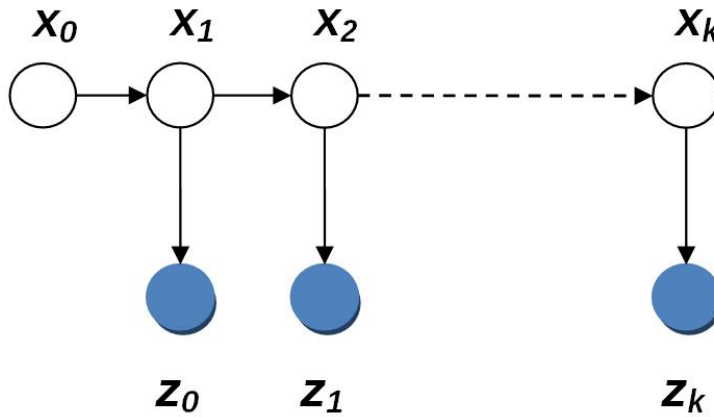


Fig. 4.3: Schéma graphique de la dépendance en suivi visuel sous l'hypothèse Markovienne. Chaque cercle représente un vecteur aléatoire ; Les flèches montrent la dépendance entre les vecteurs aléatoires (vecteur d'états x_t et le vecteur d'observation z_t) ; L'observation courante z_t dépend uniquement de l'état courant. L'état courant x_t dépend uniquement de l'état précédent x_{t-1}

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 (equation.4.10). (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 ou la forme de l'objet d'intérêt (equation.4.11). Sous l'hypothèse markovienne : étant donné 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 .

$$x_t = f_t(x_{t-1}, v_{t-1}) \quad (4.10)$$

$$z_t = h_t(x_t, \varphi_t) \quad (4.11)$$

avec φ_t et v_{t-1} 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, Update) [Arulampalam 2004].

La prédiction : Dans cette étape un modèle de prédiction d'états sui-

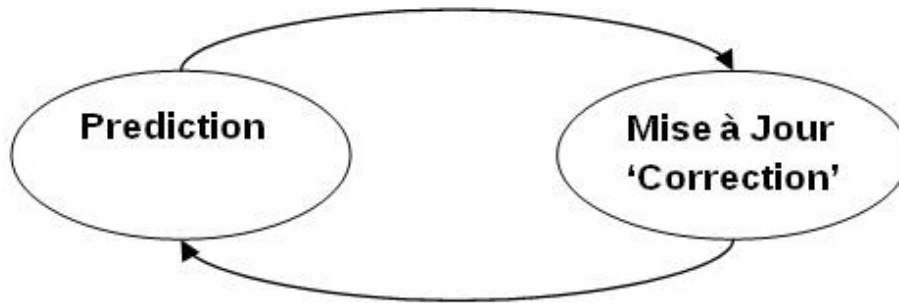


Fig. 4.4: Étapes cycliques pour l'estimation de la distribution de probabilité à posteriori utilisant le filtre de Kalman.

vant modélisant la dynamique d'objets est utilisé, comme le modèle de Chapman–Kolmogorov [Pantrigo 2005] ou encore le modèle autorégressif basé sur les séries temporelles [Agarwal 2006].

La correction (Mise à jour) : après la prédiction de l'état temporaire au temps t , l'observation z_t devient disponible. La loi bayésienne est utilisée afin d'estimer la probabilité à posteriori (équation.4.12) [Agarwal 2006].

$$P(x_t|Z_t) = \frac{p(z_t|x_t)p(x_t|z_{t-1})}{\int p(z_t|x_t)p(x_t|z_{t-1})dx_t} \quad (4.12)$$

$p(z_t|x_t)$ est le modèle de vraisemblance (Likelihood), se base sur une métrique de similarité entre deux observations (modèles de représentation des cibles) consécutives.

En pratique la détermination analytique de l'inférence bayésien (équation.4.12) reste très compliquée voir impossible. Afin de faciliter l'estimation de la probabilité à posteriori, d'autres algorithmes ont été proposés tel que le filtre de Kalman et le filtre particulaire.

4.2.3.2 Filtre de Kalman

Le filtre de Kalman [Weng 2006] est un algorithme récursif linéaire qui permet d'estimer la probabilité a posteriori $P(x_t|Z_t)$ en se basant sur les deux hypothèses suivantes : (1) la linéarité de l'équation d'évolution (équation.4.10). (2) la gaussianité de la distribution préalable et les deux processus de bruits. Le filtre de Kalman se base sur les même deux étapes du filtre bayésien récursive, la prédiction et la mise à jour. Pour chaque nouvelle position à estimer, le filtre de Kalman se déroule en deux étapes cycliques (Fig.4.5) :

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 [Bellotto 2007], des voitures [Ponsa 2005], le suivi distribué [Spanos 2005], 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 [Stenger 2001], mais cette solution reste insuffisante surtout pour des applications intégrant des aspects temps réels.

4.2.3.3 Filtre Particulaire

Contrairement au filtre de Kalman où les équations (équation.4.10 et équation.4.11) sont supposées linéaires et gaussiennes, le filtre particulaire [Isard 1998] (les filtres à particules ou encore CONDENSATION) est utilisé pour des systèmes non-linéaire avec des modèles d'observation non-gaussiens. Ce filtre se base principalement sur l'algorithme de simulation Monte-Carlo séquentielle [Arulampalam 2004]. 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 postérieure $p(x_t|z_t)$ 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^i : $\{x_{1:k}^i, \hat{w}_k^i\}_{i=1}^N$. La probabilité à postérieure [Isard 1998] est approximée par :

$$p(x_t|z_t) \approx \sum_{i=1}^N \hat{w}_t^i \delta(x_t - x_t^i) \quad (4.13)$$

Le poids des particules est défini comme suit :

$$\hat{w}_t^i = \hat{w}_{t-1}^i \frac{P(z_t|x_t^i)P(x_t^i|x_{t-1}^i)}{q(x_t^i|x_{t-1}^i, Z_t)} \quad (4.14)$$

$P(z_t|x_t^i)$, est le modèle de vraisemblance [Isard 1998] (en anglais, likelihood). 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.3.4.1, chapitre.3), la vraisemblance est estimée par l'équation suivante :

$$L_{H_{RGB}}(z_{H_{RGB},t}|x_t) \propto e^{\frac{-d_c^2(H_{t-1},H_t)}{2\sigma^2}} \quad (4.15)$$

d_c est la distance de Bhattacharya [Bousetouane 2013b], [Bousetouane 2013a]. L'algorithme du filtre particulaire basé sur la couleur pour l'estimation des trajectoires des cibles dans des séquences d'images se décompose en 3 étapes principales : (1) La prédiction (en anglais, Prediction), (2) la mesure et (3) le ré-échantillonnage [Isard 1998] :

1. La Prédiction : dans cette étape un modèle de prédiction des nouvelles positions des particules dans l'image courant t est utilisé, par exemple : le modèle autorégressif de 2^{eme} degré (en anglais, second order autoregressive model) [Bousetouane 2012].

2. Mesure : après la prédiction des nouvelles positions temporaires des particules dans l'image courante t , une mesure de similarité est utilisée (équation.3.4.1) afin de calculer les poids des particules pour la correction des positions prédites.

3. Ré-échantillonnage : Après le calcul des poids des particules dans l'étape précédente, cette étape sert à ordonnancer et sélectionner les particules selon leurs poids. Les particules avec un poids important seront multipliées et les particules avec des petits poids seront éliminées. D'autres techniques d'estimation de la probabilité à posteriori sont possibles comme le calcul de l'espérance entre les poids des particules :

$$E[x_t|z_{1:t}] \approx \frac{1}{N_t} \sum_{i=1}^{N_t} \hat{w}_t^{(i)} x_t^{(i)} \quad (4.16)$$

Une description graphique de l'algorithme filtre particulaire est illustrée dans la figure suivante :

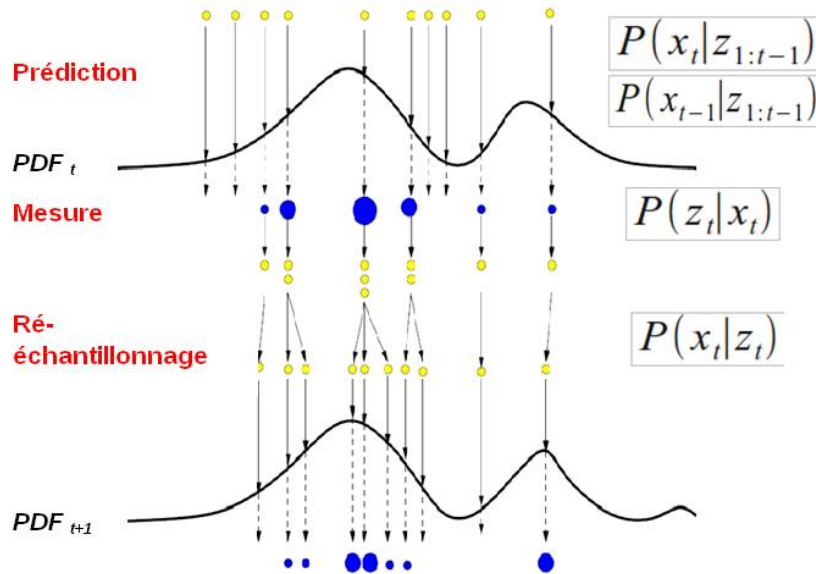


Fig. 4.5: Représentation graphique des 3 étapes d'estimation numérique de la probabilité a postérieure utilisant l'échantillonnage de Monte-Carlo et l'équation récursive bayésienne (Filtre Particulaire) [Isard 1998].

4.3 Filtre particulaire adaptatif proposé

Comme nous l'avons détaillé dans les sections précédentes, le filtre particulaire basé sur la couleur est l'un des algorithmes les plus populaires dans le suivi visuel, par rapport à sa capacité de gérer la multi-modalité et la non-linéarité. Le filtre particulaire basé sur la couleur a été utilisé dans plusieurs sous-domaines du suivi visuel : en suivi des personnes, suivi des mains, suivi inter-caméra, suivi distribué, etc. Cet algorithme devient alors une brique de base de toutes les applications liées au suivi visuel. Toutefois, l'utilisation des modèles de couleur pour la description des cibles reste insuffisante en pratique, surtout dans le suivi inter-caméra où la couleur d'une cible dans une caméra peut être relativement différente de la couleur de la même cible dans une autre caméra du même réseau (les défaillances de la couleur ont été détaillées dans le chapitre.2). Un des objectifs principaux de cette thèse est l'amélioration de l'algorithme de suivi filtre particulaire basé sur la couleur utilisé dans un système mono-caméra afin de le rendre invariant et robuste pour le suivi inter-caméra dans les systèmes multi-caméra. Récemment, plusieurs travaux remarquables ont été publiés afin d'améliorer cet algorithme. Pour ce but, plusieurs chercheurs affirment que : la combinaison de plusieurs indices visuels peut rendre le filtre particulaire basé couleur robuste et invariant dans des

conditions complexes. Le choix des indices visuels invariants et significatifs et la manière de combiner ces indices restent des problèmes au cœur de la recherche. Nous citons à titre d'exemples les travaux suivant : [Erdem 2010], proposent une nouvelle approche basée sur la fusion de plusieurs indices visuels dans l'algorithme filtre particulaire afin d'améliorer le suivi dans des conditions complexes. Cette approche reste complexe pour des applications temps réel, parce que les auteurs proposent l'attribution à chaque particule un indice visuel spécifique ce qui nécessite beaucoup de calcul. [Khan 2011], proposent un schéma de fusion de plusieurs indices visuels où l'estimateur à noyau (Mean-Shift) a été embarqué dans un filtre particulaire afin d'améliorer la convergence de chaque particule. Bien que c'est une bonne idée pour améliorer la capacité de chaque particule afin d'éviter les minimums locaux, mais cette approche reste aussi complexe car la convergence du filtre particulaire proposé est strictement conditionné par la convergence de chaque MeanShift à chaque cycle d'estimation. [Ying 2010], proposent un filtre particulaire basé sur des indices visuels multiples (multi-indice visuels), où une mixture de texture basée sur l'histogramme LBP et la distribution de la couleur a été utilisée comme modèle d'observation. [Fazli 2009], proposent un filtre particulaire amélioré par le détecteur des points (SIFT). [Nummiaro 2002], proposent un filtre particulaire adaptatif basé sur une mixture des descripteurs de contour et la distribution de la couleur. Un schéma démocratique de fusion de plusieurs indices visuels a été proposé par [ZHANG 2010]. Ce filtre est basé sur la distribution de la couleur et les transformés en ondelettes discrète. Malgré toutes les tentatives employées par plusieurs chercheurs afin d'améliorer l'algorithme de suivi « filtre particulaire basé couleur », les conditions complexes du monde réel restent un grand défi. A ce jour, les approches proposées pour la combinaison de plusieurs indices visuels dans un filtre particulaire restent limitées autour des limites des indices visuels utilisés où la combinaison de plusieurs indices visuels non-significatifs reste pauvre en terme de description des cibles. Nous croyons que pour l'amélioration de l'algorithme de suivi "filtre particulaire" et afin de le rendre robuste intra-image et inter-caméra, il faut combiner des indices visuels significatifs (i.e. invariants inter-image, inter-caméra et discriminants inter-objet). Dans le cadre de cette thèse nous proposons un filtre particulaire adaptatif basé sur une combinaison de plusieurs indices visuels [Bousetouane 2012], où le modèle de description de l'apparence des cibles proposé (voir chapitre.3) a été intégré dans le filtre particulaire afin d'améliorer la convergence des particules. Une adaptation de l'espace de recherche de chaque particule est aussi proposée en utilisant la dimension réelle de la cible calculée à partir de la carte de soustraction (voir chapitre 2) afin de rendre cet algorithme robuste aux changements d'échelles et aux mouvements non-rigide.

4.3.1 Indices visuels et modèle de Vraisemblance

La performance d'un algorithme de suivi dépend de la qualité représentative des informations visuelles extraites à partir des séquences d'image pour la description des cibles (voir chapitre.3). Dans ce travail nous avons choisi l'utilisation de trois types d'indices visuels : Forme, Couleur et Texture (information contextuelle), [Bousetouane 2013b], [Bousetouane 2013a].

Indice de Forme : la forme de la cible est modélisée par un rectangle flexible pour encadrer l'objet d'intérêt et adapter l'espace de recherche de chaque particule avec la dimension réelle de la cible. L'état de chaque particule est modélisé comme suit : $X = (x, y, H, W)^T$, où x et y sont les coordonnées du centre d'un rectangle, H et W sont respectivement la largeur et la hauteur du rectangle qui encadre l'objet d'intérêt (la cible). Dans notre algorithme, la largeur et la hauteur du rectangle sont mises à jour à chaque image de la séquence vidéo en utilisant la dimension réelle de la cible sur le plan image, calculée à partir de la carte de soustraction binarisée et étiquetée (voir chapitre.2). Cette technique permet d'améliorer la capacité d'adaptation du filtre particulaire aux changements d'échelle et aux mouvements non-rigides. Le processus d'adaptation de l'espace de recherche de chaque particule avec la dimension réelle de la cible est illustré dans la figure suivante (Fig.4.6).

indice de la Couleur : le modèle de la couleur intégré dans le filtre particulaire proposé se base sur l'histogramme de couleur (voir chapitre.3). Le modèle de vraisemblance de la couleur utilisé dans notre algorithme se base sur la distance de Bhattacharya entre l'histogramme de couleur calculé au temps t et l'histogramme de couleur calculé au temps $t + 1$, equation.4.15.

Informations contextuelles de bas niveau : basées sur l'analyse de texture et la relation spatiale entre les pixels composant l'image de la cible d'intérêt, calculées à travers la matrice de cooccurrence et les indices de Haralick. Ces indices ont été prouvés dans le cadre de cette thèse comme invariant inter-image et inter-caméra et discriminant inter-objets (voir chapitre.3). Le modèle de vraisemblance proposé dans l'espace de texture pour calculer la similarité entre le modèle de texture calculé au temps $t - 1$ et le modèle de texture calculé au temps t , se base sur la distance de Mahalanobis [Mao 2011]. Ce modèle de vraisemblance est défini de la même façon que la vraisemblance de la couleur :

$$L_{Context}(z_{Context,t}|x_t) \propto e^{\frac{-\delta_2(V(x,t-1),V(x,t))}{2\sigma^2}} \quad (4.17)$$

$V(x, t - 1)$ et $V(x, t)$, représentent le modèle de référence et le modèle candidat respectivement de description des cibles basés sur l'information contex-

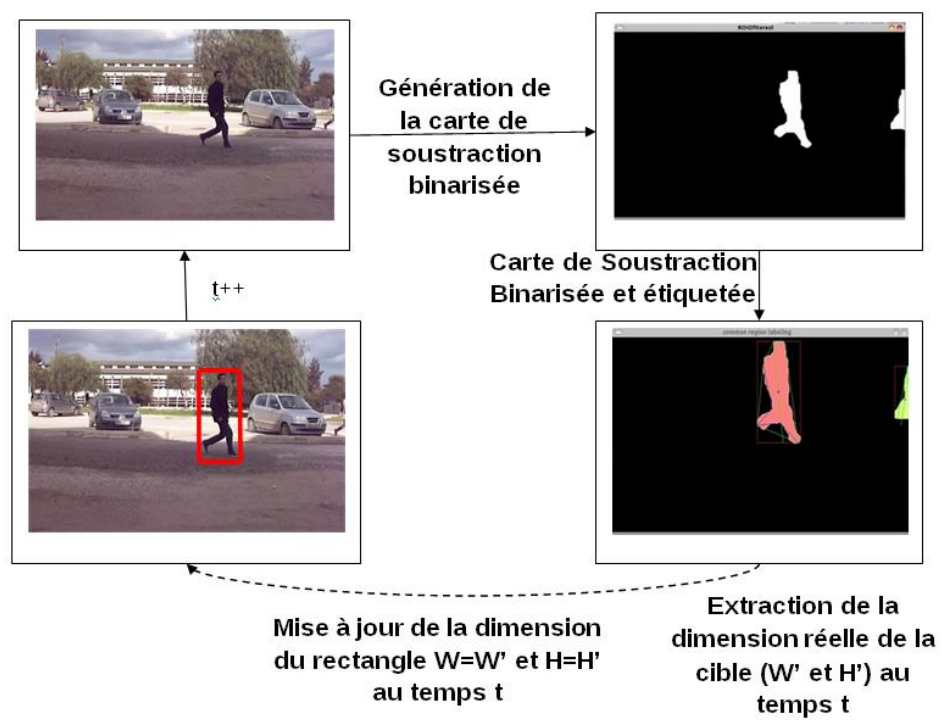


Fig. 4.6: illustration graphique de la technique proposée pour l'adaptation de l'espace de recherche de chaque particule avec la dimension réelle de la cible.

tuelle de bas niveau (texture). Sous l'hypothèse que les indices visuels de description des cibles sont indépendants, alors le modèle global de vraisemblance que nous proposons se base sur un produit pondéré des modèles de vraisemblance de chaque indice (couleur et information contextuelle de bas niveau), voir notre publication [Bousetouane 2013a]. Le modèle de vraisemblance global proposé pour calculer le poids des particules est défini comme suit :

$$L(z_t|x_t) = \alpha L_{H_{RGB}}(z_{H_{RGB},t}|x_t) * \beta L_{Context}(z_{Context,t}|x_t) \quad (4.18)$$

α et β sont des poids de pondération prédéterminés de l'importance des indices visuels.

Le filtre particulaire adaptatif basé sur l'intégration de multi-indice visuels "APF-MC" que nous proposons se décompose en sept étapes successives :

1. Dans la première image de la séquence vidéo :
 - a. Extraction automatique de l'objet d'intérêt (cible) utilisant la chaîne de traitement proposée pour la détection d'objet en mouvement (voir chapitre.2).
 - b. Le calcul du modèle de texture proposé et la distribution de la couleur à partir de l'image de cette région (objet d'intérêt) (chapitre.3 équation.3.13, équation.3.17).
2. Initialisation du filtre particulaire proposé par les paramètres de l'état et du poids (x , $bruit$, , , Nombre de particule N).
3. La prédiction des positions temporaires des particules dans l'image suivante utilisant le modèle autorégressive de 2^{me} degré [Pérez 2002], ce modèle est défini comme suit :

$$X_t = A_2(X_{t-2}) + A_1(X_{t-1}) + B_0\varpi_0.$$

(4.19)

X est la description spatiale de la localisation d'objets d'intérêt, A_2, A_1, B_0 des paramètres prédéterminés du modèle, ϖ_0 un bruit gaussien.

4. Calculer le poids de chaque particule $\hat{w}_i(i = 1, \dots, N)$ utilisant l'équation.4.18.
5. Calculer la probabilité a postériori moyenne courante :

$$\hat{X} = \sum_{i=1}^N \hat{w}_t^{(i)} x_t^{(i)} \quad (4.20)$$

6. Ré-échantillonnage des particules et l'adaptation de l'espace de recherche :

a. Suppression et multiplication des particules selon le poids, les particules avec un poids important seront multipliées et les particules avec un poids non-important seront supprimées (le même principe de la sélection naturelle utilisée dans la méta-heuristique).

b. Adaptation de l'espace de recherche de chaque particule sélectionnée avec la dimension réelle de la cible d'intérêt utilisant la technique proposée dans la section.4.3.1.

7. Retourner vers l'étape 3 jusqu'à la non-mobilité de la cible.

La forme algorithmique de notre filtre particulaire est illustrée dans (Algo.1) :

Algorithm.1. Contextual Adaptive Particle Filter Object
Tracking Algorithm APF-MC

INPUT : Image sequence I_0, I_1, \dots, I_n and the initial interest object region.

OUTPUT : Target locations c_0, c_1, \dots, c_n at tracking over times, in frames I_0, I_1, \dots, I_n .

Begin

Initialization

1. $k = 0$, for $i = 1, \dots, N$, generate samples w_0^i from the initial distribution $P_0(x)$.

Prediction Step

2. For $k = 1, 2, \dots; i = 1, \dots, N$ sample $x_k^{(i)} \sim P(x_k | x_{k-1}^{(i)})$, dynamic system model according to equation (14).

Measurement Update

Evaluate the importance weights of each particle in the current frame.

3. Computing the weights according to equation.4.18 and Normalize the weights.

$$\hat{w}_k^{(i)} \propto l_{Context+HRGB}(z_k | x_k^{(i)}) \quad (4.21)$$

4. Computing the approximate posterior distribution, according to equation (5).

5. Computing the current posterior mean.

$$\hat{X} = \sum_{i=1}^N \hat{w}_k^{(i)} x_k^{(i)} \quad (4.22)$$

Re-Sampling

6. Multiply/suppress samples with high or low importance weights, adapting the size of the rectangle box of each particle with the size of the extracted mask of motion in the current frame k .
7. $k = k + 1$ and return to step 2, until the target is in stationary state.

End

Le filtre particulaire adaptatif proposé est capable d'estimer les positions d'un objet d'intérêt dans des conditions complexes. Pour assurer le suivi multi-objet, nous proposons l'utilisation d'un filtre par objet d'intérêt. L'approche proposée pour le suivi multi-objet intégrant la gestion de l'association de données est détaillée dans la section suivante.

4.3.2 Suivi Multi-objet

Récemment, plusieurs approches ont été proposées pour l'estimation des positions de plusieurs objets dans une scène (en anglais, Multi-Target tracking). Ces approches peuvent être décomposées en deux grandes familles : approches centralisées et approches décentralisées ou distribuées.

Les approches centralisées : ces approches introduisent une représentation conjointe de l'espace d'état, où un méta-état est créé à travers l'agrégation de tous les états des cibles (objets d'intérêts). L'association de données conjointe est inférée à travers la caractérisation de toutes les associations possibles entre les objets et les observations candidates. Malgré l'évolution remarquable des approches centralisées, comme les MCMC [Khan 2005], RJMCMC [Smith 2007] et JPDA [Musicki 2004], elles restent complexes pour des applications temps réelles. En effet, l'association de données entre le méta-état qui encode plusieurs objets et les observations candidates crée un problème de nature combinatoire, ce qui est complexe non seulement en résolution théorique mais également lors de son implémentation logicielle.

Les approches décentralisées : Ces approches se basent essentiellement sur l'utilisation, soit de plusieurs trackeurs simultanément pour l'estimation des positions d'objets séparément, soit d'un trackeur par objets afin d'éviter l'état conjoint. Comme résultats, le suivi multi-objet est réalisé avec une complexité linéaire en termes d'objets. Malgré la large utilisation de ces approches [Qu 2007], elles restent limitées surtout en présence d'occultation complexe. En effet, en présence d'occultations complexes, la résolution de l'association

de données et la gestion d'identités deviennent des tâches très compliquées. D'après nos expériences nous croyons que la résolution de l'association de données est strictement conditionnée par la capacité discriminatoire inter-objet du modèle d'observation utilisé pour la description des cibles.

Dans ce travail de thèse nous optons sur les approches décentralisées par rapport à sa rapidité, mais l'efficacité de ces approches est strictement conditionnée par l'efficacité de l'algorithme de suivi mono-objet utilisé. L'approche proposée pour le suivi multi-objet dans le cadre de cette thèse se base sur l'utilisation de plusieurs instances du filtre particulaire adaptatif proposé "APF-MC", où chaque instance opère sur une cible (objet d'intérêt). Dans cette approche, chaque filtre possède ses propres particules et chaque particule peut converger vers la position d'objet d'intérêt la plus appropriée. L'association de données est implicitement gérée dans notre approche à travers le modèle de vraisemblance intégré dans chaque filtre. Ce dernier est prouvé dans (chapitre 3) comme invariant inter-image et inter-caméra et discriminant inter-objet. La probabilité à posteriori $p(x_k|Z^t)$ est étendue fréquemment dans le suivi multi-objet et mise à jour d'une manière récursive à l'état conjoint des S objets avec l'état $x_k = (x_1, x_2, \dots, x_S, k)$, S est le nombre d'objets d'intérêts (cibles). Sous l'hypothèse que les observations (équation.4.18) des différents objets d'intérêts sont indépendantes, alors la distribution bayésienne multi-objet dans un contexte distribué est définie comme suit :

$$p(x_t|Z^t) = \frac{\prod_{s=1}^S p_s(z_t|x_{s,t})p_s(x_{s,t}|Z^{t-1})}{\prod_{s=1}^S p_s(z_t)} \quad (4.23)$$

Sous l'hypothèse de l'indépendance entre les observations d'objets, l'équation (équation.4.23) surmonte le problème de la dimensionnalité par le calcul des distributions marginales de la distribution $p(x_k|Z^t)$. Cette distribution est estimée à travers l'estimation de la position de chaque objet indépendamment des autres objets utilisant des instances séparées du filtre particulaire adaptatif proposé "APF-MC". En utilisant l'équation.4.23 pour la modélisation du suivi multi-objet basé sur le filtre bayésien, la complexité du filtrage d'états multi-objet évolue linéairement en relation avec le nombre d'objet S , plutôt qu'une explosion exponentielle dans le cas des approches centralisées. Dans notre approche le nombre d'objets d'intérêts est initialement calculé utilisant la chaîne de traitement proposée dans (chapitre.3). Si un nouvel objet est détecté alors une nouvelle instance du filtre particulaire adaptatif proposé est attribuée à cet objet pour estimer sa trajectoire. Le grand souci de notre approche pour le suivi multi-objet est l'association de données après occultation complète, une technique basée sur « Birth and Death process » et la gestion "inter-blobs" [Sherrah 2009] sera intégrée dans nos travaux futurs. Les résultats d'applications de notre approche pour le suivi mono/multi-objet

dans des conditions complexes prouvent la validité et la puissance de notre approche. La section suivante illustre quelques résultats d'application de notre approche pour le suivi multi/mono-objet à partir des séquences vidéo capturées par des caméras de notre réseau multi-caméra monté dans le cadre de cette thèse (chapitre.1, section.1.2.1).

4.4 Résultats et Evaluation

Pour évaluer la performance de l'algorithme (APF-MC) proposé dans le cadre de cette thèse pour le suivi d'objets mobiles, nous avons utilisé plusieurs vidéos capturées par des caméras de notre réseau multi-caméra et des vidéos provenant des corpus de test de la communauté scientifique. Les vidéos de teste utilisées dans ce travail sont capturées en présence des conditions complexes, comme : la variation de la couleur intra-caméra, occultations complexes, plusieurs objets (jusqu'à 8 cibles), objets non-rigide, arrière-plan dynamique (mouvement saccadé causé par les arabes, les voitures), in/outdoor, etc. Dans ces conditions le suivi d'objets mobiles reste un grand défi et l'algorithme proposé est bien évalué. Dans cette partie, nous détaillons les résultats d'application de notre algorithme dans des conditions réelles, ainsi que nous évaluons quantitativement la consistance de notre algorithme par rapport aux autres travaux.

4.4.1 Résultats de suivi Mono/Multi-Objet

Pour tester notre algorithme, nous avons choisie 10 séquences d'images dont 4 proviennent des corpus de test en vidéo surveillance (Benchmark) et les autres 6 sont capturées par des caméras de notre réseaux multi-caméra. Les corpus de test en vidéo surveillance utilisés sont [CAVIAR 2004], [Pets 2000], [Pets 2009], [SPEVI 2007]. Les expérimentations sont réalisées en utilisant une machine sous Linux dotée d'un processeur Core 2 Duo, 2Go de RAM. L'algorithme a été implémenté avec le langage C++ utilisant les bibliothèques de traitement d'image OpenCv et GSL. Dans toutes les expérimentations 45 particules ont été créées par filtre, $N = 45$.

Séquence.1 : Cette séquence a été extraite à partir du corpus de test en vidéo surveillance [CAVIAR 2004]. Cette vidéo contient un objet non-rigide humain qui circule en présence des conditions complexes : changement d'échelle, observation non-homogène, arrière-plan texturé, mouvement non-rigide, déformation locale, etc. Le suivi d'objet dans ces conditions est très compliqué. La figure (Fig.4.7) illustre les résultats d'estimation des positions de cet objet

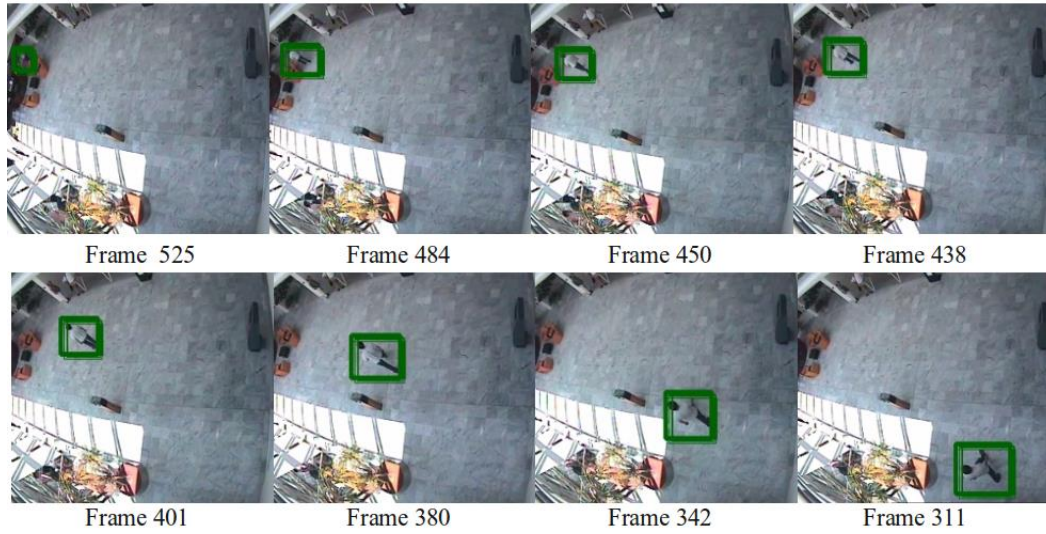


Fig. 4.7: Résultats d'application de l'algorithme proposé APF-MC sur les séquences [CAVIAR 2004] pour l'estimation des positions d'objet non-rigide. Nombre des particules , $N=45$.

au fil du temps en utilisant l'algorithme proposé (APF-MC, filtre particulaire adaptatif basé multi-indice visuels).

A partir des résultats d'application de notre algorithme sur cette séquence (Fig.4.7), nous pouvons visuellement remarquer que : les états des particules (les rectangles vert) convergent vers la meilleur position de la cible d'intérêt. Cette convergence est explicitement justifiée par la capacité descriptive (l'invariance intra/inter-caméra et discriminant inter-objet) du modèle de vraisemblance proposé et la capacité d'adaptation de l'algorithme proposé au changement d'échelle et au mouvement non-rigide.

Séquence.2 : Cette séquence a été extraite à partir du corpus de test en vidéo surveillance [SPEVI 2007], et qui consiste en un objet non-rigide humain circulant en intérieur (indoor) en présence des conditions complexes : observation non-homogène, arrière-plan texturé, mouvement non-rigide, occultation complexe, variation de couleur inter-image et résolution très faible des images. La figure (Fig.4.8) illustre les résultats d'estimation des positions de cet objet au fil du temps en utilisant notre algorithme "APF-MC".

A partir des résultats d'estimation des positions d'objet d'intérêt obtenus utilisant notre algorithme APF-MC (Fig.4.8), nous pouvons clairement remarquer que les particules du filtre (rectangles rouge) convergent vers les meilleures positions de la cible au fil du temps, même quand la cible en occultation complexe (frame 778) et après occultation (frame 786). Ces résultats



Fig. 4.8: Résultats d'application de notre algorithme APF-MC sur les séquences [SPEVI 2007] pour l'estimation des positions d'objet non-rigide. Nombre de particules $N=45$.



(a) résultats de suivi utilisant notre filtre particulaire APF-MC



(b) résultats de suivi utilisant le filtre particulaire basé couleur classique

Fig. 4.9: (a) Résultats d'estimation des positions d'objet d'intérêt au fil du temps utilisant notre algorithme APF-MC. (b) Résultats d'application du filtre particulaire classique basé couleur pour le suivi d'objet d'intérêt.

prouvent la capacité de notre algorithme à gérer la multi-modalité à travers la convergence vers les optimums globaux où les positions *fausses positives* ont été évitées.

Séquence.3 : Cette séquence illustre une comparaison des résultats de suivi entre notre algorithme APF-MC et le filtre particulaire classique basé sur la couleur, en utilisant les séquences d'images [CAVIAR 2004].

A partir de cette comparaison (Fig.4.9), nous pouvons remarquer que : en utilisant notre algorithme "APF-MC" (Fig.4.9(a)), l'ensemble des particules convergent vers la meilleure position de la cible, en revanche une divagation remarquable (Fig.4.9(b)) dans la convergence des particules utilisant l'algorithme classique « filtre particulaire basé-couleur ». Cette comparaison prouve que le modèle de vraisemblance proposé (section.4.3.1), et qui se base sur une mixture d'indices visuels est plus significatif que le modèle utilisé dans le filtre particulaire classique basé sur la distribution de la couleur. Cette expérience

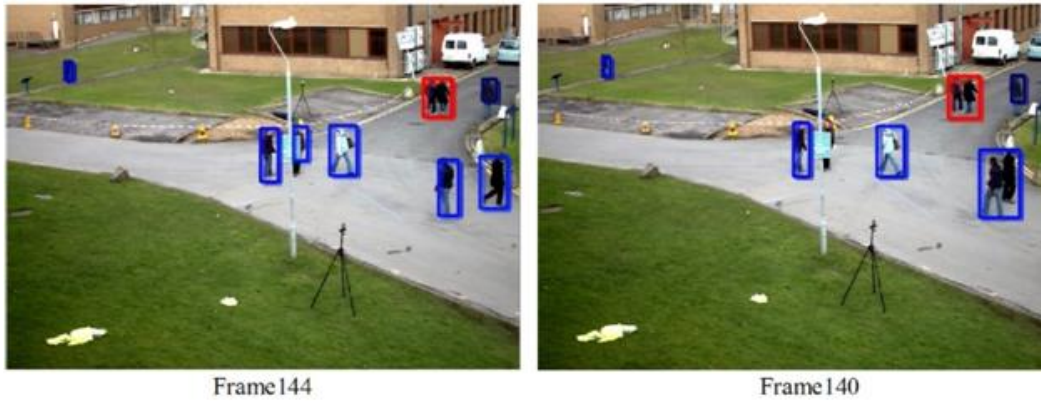


Fig. 4.10: Résultats de suivi multi-objet en utilisant notre approche basée sur plusieurs instances du filtre particulaire proposé APF-MC sur les séquences *PETS2006* [Pets 2009].

tation affirme nos croyances (section.4.3) que la capacité et la robustesse d'un algorithme de suivi dans des conditions et des configurations complexes dépend explicitement de la capacité représentative du modèle de vraisemblance utilisé.

Séquence.4 :

Pour évaluer la capacité de l'approche proposée pour le suivi multi-objet dans un contexte distribué et qui se base sur des multiples instances du filtre particulaire proposé APF-MC, nous avons choisi une séquences d'images du corpus de test en vidéo surveillance [Pets 2009]. Cette séquence consiste en 8 objets humains non-rigides circulant dans un environnement spécifique "outdoor" en présence des occultations complexes, mouvement non-rigide, changement d'échelle, arrière-plan texturé, etc. La figure (Fig.4.10) illustre le résultat de suivi multi-objet utilisant notre approche :

A partir des résultats de suivi multi-objet obtenu (Fig.4.10), nous pouvons remarquer que : chaque particule converge vers les meilleures positions de son objet d'intérêt au fil du temps, même dans le cas d'occultation complexe (frame 140). Ces résultats prouvent la capacité de l'approche proposée à suivre plusieurs objets à travers l'attribution d'une instance du filtre particulaire proposé à chaque objet d'intérêt. Le problème de l'association de données est implicitement résolu à travers la capacité discriminatoire du modèle de vraisemblance proposé (équation.4.18) et intégré dans chaque instance du filtre particulaire proposé "APF-MC".

Séquence.5 : Cette séquence a été capturée par la caméra numéro 1 (E-Cam-1) de notre réseau multi-caméra monté dans le cadre de cette thèse



Fig. 4.11: Résultats de suivi multi-objet utilisant notre approche basée sur multiple instances du filtre particulaire proposé APF-MC, sur des séquences capturées par la caméra E-cam-1 de notre réseau multi-caméra.

(voir chapitre 1). Cette séquence consiste en trois objets d'intérêts non-rigides circulant dans le réseau en présence des conditions complexes : objets en rotation (frame 162), mouvement dans l'arrière-plan de la scène causé par le mouvement des arbres, des voitures et d'autres objets non-cible, changement de direction (frame 187) et arrière-plan textures. La figure (Fig.4.11) illustre les résultats de suivi multi-objet en utilisant notre approche.

A partir des résultats d'application de notre approche pour le suivi multi-objet (Fig.4.11) sur des séquences capturées par la caméra E-CAM-1 de notre réseau multi-caméra, nous pouvons remarquer que : malgré la présence des mouvements en arrière-plan, changement d'éclairage, les ombres, etc., notre approche estime parfaitement les trajectoires d'objets d'intérêts en temps réel. Cette expérimentation prouve encore la capacité et la robustesse de notre approche pour le suivi Mono/Multi-objet en présence des conditions complexes.

L'algorithme proposé "APF-MC" et l'approche proposée pour le suivi Multi-objet ont été sérieusement évalués sur plusieurs séquences d'images en présence des conditions complexes. Toutes les expérimentations réalisées prouvent la robustesse et la capacité de notre approche à suivre les objets mobiles. L'algorithme classique du filtre particulaire basé-couleur a été remarquablement amélioré dans ce travail à travers le modèle de vraisemblance proposé (section.4.3.1) et l'adaptation de l'espace de recherche de chaque particule avec la dimension réelle de la cible. Ces améliorations permettent à notre approche d'éviter les positions erronées (fausses positives ou encore les minimums locaux) et gérer le changement dans la signature visuelle au fil du temps. Nous avons présenté dans cette section plusieurs résultats très encourageants de notre approche sur 5 séquences différentes. La section suivante

illustre une comparaison quantitative entre les résultats d'estimation des positions de notre approche et d'autres approches de l'état de l'art.

4.4.2 Evaluation

La capacité d'un algorithme de suivi d'objets mobiles dépend de la tâche à accomplir et la complexité des séquences vidéo utilisées. Aucun algorithme de suivi n'est universellement adapté pour toutes les applications. En effet, la création d'un algorithme de suivi générique est une tâche très complexe. Dans la littérature, plusieurs méthodes ont été proposées pour l'évaluation des algorithmes de suivi d'objets mobiles. Ces méthodes mettent l'accent sur un jugement de l'algorithme de suivi lui-même sur une base algorithme ou directement à partir des résultats de suivi. Ces méthodes peuvent être décomposées en deux grandes familles [Wu 2010] :

1. Les Méthodes analytiques : Ces méthodes évaluent les algorithmes de suivi d'une manière purement algorithmique, prennent en considération la complexité en temps et en espace et la méthode de la mise en correspondance inter-image intégrée dans l'algorithme (Object Matching, Points Matching, Contextual Matching, etc.).

2. Les Méthodes empiriques : Ces méthodes évaluent les algorithmes de suivi d'une manière indirecte à travers les résultats de suivi.

La deuxième méthode pour l'évaluation des algorithmes de suivi nommée méthode de divergence [Schlogl 2004] (en anglais, discrepancy method), permet de mesurer l'erreur d'estimation des positions à travers une comparaison entre les résultats de suivi d'un algorithme et la vérité du terrain d'un corpus de test (en anglais, Ground truth). Une comparaison dans l'estimation des positions au fil du temps entre la vérité de terrain proposée par le corpus de test [SPEVI 2007], la séquence «motinas-Room160» et les résultats de suivi de notre algorithme APF-MC est illustrée dans la figure (Fig.4.12).

Dans ce travail, nous utilisons la deuxième méthode d'évaluation. L'erreur dans l'estimation des positions au fil du temps est utilisée comme un critère quantitatif. Cette erreur est estimée par la distance euclidienne entre la trajectoire estimée par notre algorithme et les positions réelles de la cible dans l'image (vérité du terrain obtenue de manière manuelle). La cinquième séquence est utilisée dans cette évaluation. La figure suivante illustre une comparaison dans l'erreur d'estimation des positions de l'objet d'intérêt avec l'identité "1" (Fig.4.12) entre l'algorithme proposé APF-MC, le filtre particulaire classique basé sur la couleur, l'estimateur à noyau basé sur la couleur et l'estimateur à noyau contextuel. La figure (Fig.4.12) illustre l'erreur d'estima-

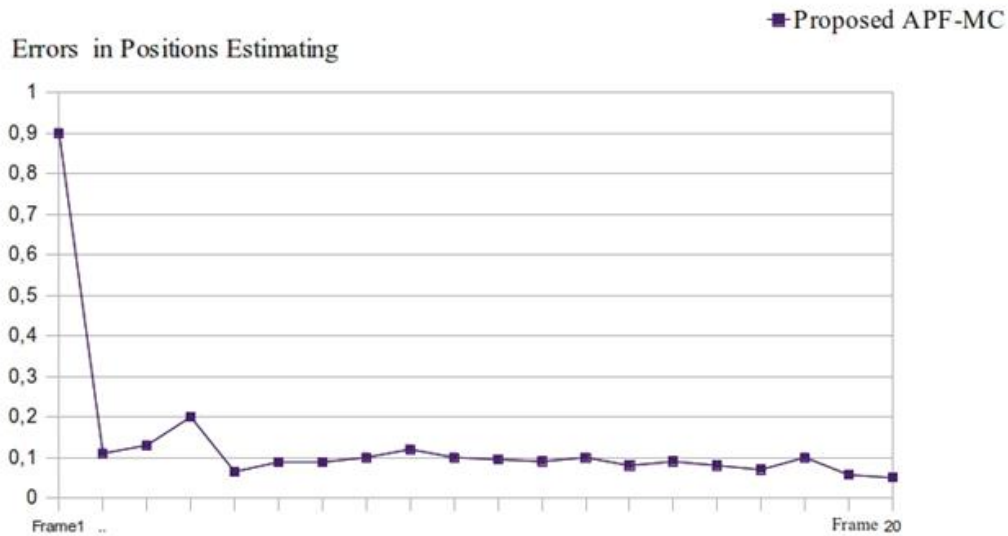


Fig. 4.12: Pourcentage d'erreurs dans l'estimation des positions de notre algorithme APC-MC

tion des positions des 3 objets d'intérêt présent dans la séquence.5.

A partir des résultats d'évaluation de notre algorithme APF-MC, nous remarquons que l'erreur dans l'estimation des positions est remarquablement minimisée comparant avec d'autres algorithmes (Fig.4.13). Ces résultats prouvent encore l'efficacité de l'algorithme proposé APF-MC pour le suivi d'objets mobiles. Le modèle de vraisemblance proposé et l'adaptation de l'espace de recherche de chaque particule au fil du temps rendent notre algorithme APF-MC plus robuste et capable de gérer la variation de la couleur inter/intra-caméra et les minimums locaux (positions fausses positives). La gestion d'identités et l'association de données sont aussi résolues implicitement dans un contexte distribué à travers le modèle de vraisemblance proposé (chapitre.3).

4.5 Conclusion

Dans ce chapitre nous avons présenté préalablement un état de l'art sur les techniques utilisées jusqu'à présent pour le suivi d'objets mobiles à partir d'une configuration mono-caméra. Après une présentation des insuffisances des techniques classiques pour le suivi d'objets mobiles, nous avons proposé un nouveau filtre particulière adaptatif basé sur une combinaison de plusieurs indices visuels « APF-MC ». L'algorithme proposé dans le cadre de cette thèse se base sur deux opérations essentielles : (1) l'adaptation de l'espace de recherche de chaque particule avec la dimension réelle de la cible afin de gérer le

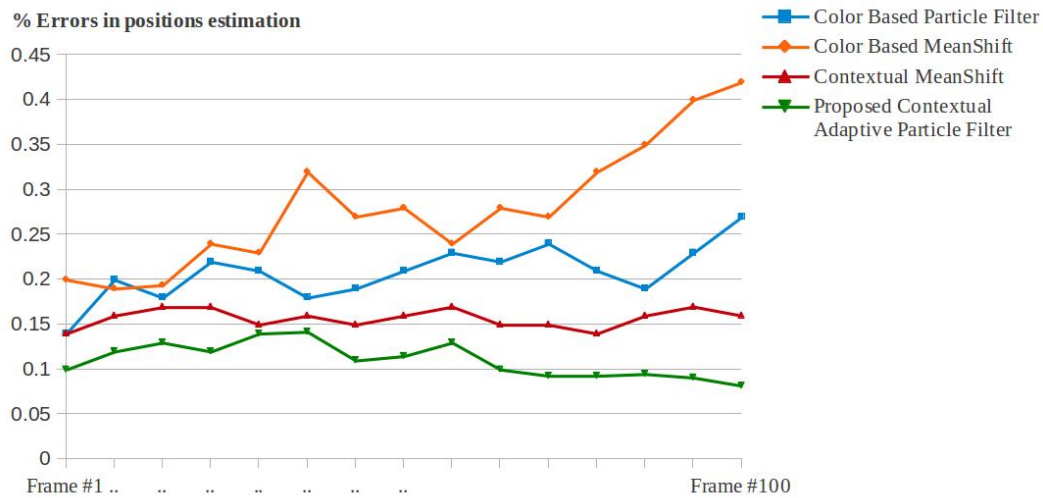


Fig. 4.13: Evaluation de l'algorithme proposé APF-MC

changement d'échelle et le mouvement non-rigide. (2) la mixture des indices visuels en un modèle de vraisemblance général pour améliorer la convergence des particules afin d'éviter les positions fausses positives (minimums locaux). Les expérimentations réalisées prouvent la validité et la puissance temps réel de l'algorithme proposé "APF-MC" pour le suivi mono-objet et multi-objet en utilisant multiple instances de notre algorithme dans un contexte distribué. Dans le chapitre suivant nous présenterons une nouvelle approche basée sur l'inférence distribué pour le suivi et la ré-identification d'objets mobiles inter-caméra dans les réseaux multi-caméra à champs disjoint sans fil. Nous détaillons aussi l'intégration de notre approche pour le suivi multi-objet "APF-MC" dans le suivi inter-caméra.

Suivi et ré-identification d'objets mobiles inter-caméra

Sommaire

5.1	Introduction	112
5.2	Description de la solution proposée	113
5.2.1	Architecture Générale	113
5.2.2	Détection d'objets mobiles	117
5.2.3	Adaptation de la Couleur inter-caméra	118
5.2.4	Extraction d'indices visuels et représentation des cibles	122
5.2.5	Ré-identification distribuée des cibles Inter-caméra	124
5.2.6	Suivi d'objets Mobiles Intra-Caméra	129
5.3	Expérimentations	130
5.3.1	Résultats et discussions	131
5.4	Evaluation	138
5.5	Conclusion	141

DANS ce chapitre, nous présentons premièrement un système décentralisé pour le suivi et la ré-identification d'objets mobiles inter-caméra dans un réseau multi-caméra à champs disjoints sans fil. Le système proposé repose sur le paradigme multi-agent. Les processus présentés dans les chapitres précédents sont intégrés dans l'architecture multi-agent proposée : (1) la détection d'objets mobiles et la génération de la carte de soustraction binarisée et étiquetée. (2) Extraction d'indices visuels et la représentation des cibles. (3) Le suivi d'objets mobiles intra-caméra. Ensuite, nous détaillons l'algorithme que nous proposons pour la ré-identification décentralisée des cibles inter-caméra et qui se base sur le système multi-agent proposé. Les caméras intelligentes (les instances d'agent cognitif proposé) collaborent entre elles afin de parvenir à un consensus sur les identités des cibles traversant le réseau multi-caméra. Enfin, nous consacrons une section aux résultats obtenus pour le suivi et la ré-identification d'objets mobiles d'une manière totalement décentralisée sur des séquences vidéo capturées par des caméras de notre réseau multi-caméra (Chapitre.1 (Fig.1.3)).

5.1 Introduction

Le domaine de la vidéosurveillance a connu une très forte expansion ces dernières années surtout avec la mise en place des réseaux multi-caméra surveillant des grandes espaces. La multiplication des caméras installées dans des espaces publics ou privés, rend l'exploitation des données visuelles distribuées ou réparties par des opérateurs humains une tâche très complexe. Pour cela, le développement des systèmes automatiques pour l'analyse des données visuelles distribuées dans les réseaux multi-caméra devienne une tâche indispensable. Cependant, l'agrégation, l'interprétation et l'analyse de l'information visuelle à partir de plusieurs flux vidéo dans des scénarios de la vie réelle sont des opérations très complexes. Néanmoins, ces opérations nécessitant le développement des nouveaux algorithmes et des méthodes d'inférence collaborative sophistiquées capables d'analyser l'information visuelle répartie ou décentralisée en temps réel. Dans un réseau multi-caméra trois types de configurations par rapport aux champs de captures sont possibles : caméras avec chevauchement dans les champs de captures, caméras sans chevauchement dans les champs de captures ou caméras à champs disjoints et caméras hybrides. Les contraintes matérielles et économiques limitent en général le nombre des caméras et empêchent une couverture totale de l'espace, ce qui engendre des discontinuités dans le champ de vision du réseau. Un des principaux défis des réseaux multi-caméra à champs disjoints est la ré-identification des cibles traversant le réseau : lorsqu'un objet apparaît dans le champ de couverture d'une caméra, il s'agit de déterminer si cet objet a déjà été observé et suivi par l'une des caméras du réseau. Dans la littérature plusieurs approches ont été proposées pour le suivi et la ré-identification d'objets mobiles inter-caméra dans les réseaux multi-caméra à champs disjoints. Ces approches restent limitées dans des conditions spécifiques d'application (voir l'état de l'art, chapitre.1, section.1.3.2) : infrastructures centralisées, les réseaux fermés, calibration des caméras, topologie spécifique, réseau multi-caméra filaire, etc. En présence de toutes ces restrictions le suivi et la ré-identification d'objets mobiles inter-caméra dans des scénarios de la vie réelle reste un grand défi. La nature ad-hoc et intrinsèquement distribuée des réseaux multi-caméra sans fil ont orienté les chercheurs de plus en plus vers les techniques d'inférences distribuées [Bousetouane 2013c], [Soto 2009] et la théorie des jeux [Li 2011] pour la réalisation des tâches de haut niveau. L'efficacité du paradigme multi-agent pour la modélisation des problèmes complexes de natures distribués prouvée dans plusieurs travaux de recherches [Soto 2009], [Hu 2010], [Quaritsch 2007] à attirer notre attention. Dans ce travail de thèse, le système distribué proposé pour le suivi et la ré-identification d'objets inter-caméra se base sur le paradigme multi-agent. Le système proposé est composé de plusieurs caméras

intelligentes (instances du modèle d'agent cognitif proposé) sans chevauchement dans les champs de capture. Un modèle générique d'agent cognitif est proposé pour la modélisation des caméras. Physiquement, chaque instance du modèle d'agent cognitif est utilisée par caméra. Les caméras intelligentes (les instances) collaborant afin de suivre et ré-identifier les cibles traversant le réseau multi-caméra.

5.2 Description de la solution proposée

Comme nous l'avons détaillé dans le chapitre.1(section.1.3.2), les approches proposées jusqu'à présent pour le suivi et la ré-identification d'objets mobiles inter-caméra dans un réseau multi-caméra à champs disjoints restent limitées dans des conditions spécifiques d'applications. La majorité des approches proposées jusqu'à présent, comme [Chen 2011], [Javed 2008a], se basant sur une infrastructure centralisée. La nécessité des solutions matériellement économique, invariantes à l'échelle, capable d'analyser l'information visuelle répartie en temps réel, rend les architectures centralisées des solutions insuffisantes. Ces insuffisances nécessitant le passage vers des architectures complètement décentralisées afin de favoriser le calcul distribué et l'inférence répartie. D'après nos connaissances, une solution totalement décentralisée pour le suivi et la ré-identification d'objets mobiles inter-caméra n'a pas été proposée auparavant. Le système que nous proposons dans le cadre de cette thèse pour le suivi et la ré-identification d'objets mobiles inter-caméra est totalement décentralisé. Dans ce système, le paradigme multi-agent a été adopté afin de modéliser le réseau multi-caméra et l'interaction entre les caméras en terme de collaboration. Dans la section suivante nous détaillons le système proposé.

5.2.1 Architecture Générale

Nous concéderons un réseau multi-caméra ad-hoc nommé $G = (V, E)$, V est le nombre des caméras où $V = \{1, \dots, Nc\}$ et $E \subset (V * V)$. Chaque caméra est connectée à une unité de calcul différente des autres caméras. Les caméras sont distribuées géographiquement sans chevauchement dans les champs de captures, ce que crée une discontinuité où une zone noir entre les caméras. Le support de communication entre les caméras est sans fil à bande passante limitée. Aucune calibration des caméras ni nécessaire. La figure (Chapitre.1 (Fig.1.3)) illustre le réseau multi-caméra à champs disjoints sans fil composé de sept caméras monté dans le cadre de cette thèse au niveau de notre département d'informatique. La nature distribuée des problèmes de

suivi et la ré-identification d'objets inter-caméra exige l'utilisation d'une architecture décentralisée, les caméras intelligentes composant le réseau collaborent entre elles afin de suivre et ré-identifié les cibles traversant le réseau. Cette nature nous mène à modéliser le problème par l'approche multi-agent, où un nouveau modèle d'agent cognitif est proposé. Chaque caméra intelligente (instance) est autonome dans sa décision locale et ses traitements. Les caméras intelligentes (instances du modèle d'agent cognitif) collaborent mutuellement afin de parvenir à un consensus sur les identités des cibles traversant le réseau multi-caméra. Le modèle générique d'agent cognitif proposé pour la modélisation des caméras incorpore cinq modules essentiels : (1) la détection d'objets en mouvement et la génération de la carte de soustraction binarisée et étiquetée (Chapitre.2). (2) Extraction des caractéristiques et représentation des cibles (Chapitre. 3). (3) Adaptation de la couleur inter-caméra basée sur la fonction de transfère de luminosité moyenne (CBTF).(4) La ré-identification collaborative d'objets inter-caméra basée sur le nouvel algorithme que nous proposons (Consensus). (5) Suivi d'objets mobiles intra-caméra (chapitre. 4). Cette structuration modulaire des différents traitements de chaque instance du modèle d'agent cognitif (caméra intelligente) permet de décentraliser l'expertise, ce qui assure l'amélioration de la performance de chaque instance en terme de temps de calcul et de gestion. Techniquement, une décomposition modulaire de l'expertise assure une fluidité pour une implémentation parallèle des différents modules sur des plateformes multiprocesseurs ou à processeur graphique (GPU).

La figure suivante (Fig.5.1) illustre l'architecture du système distribué proposé composé de plusieurs instances du modèle d'agent cognitif proposé, où une instance modélise une caméra. La figure (Fig.5.2)illustre l'architecture du système distribué proposé composé de plusieurs agents intelligent pour le suivi et la ré-identification des cibles inter-caméra dans un réseau multi-caméra à champs disjoints sans fil.

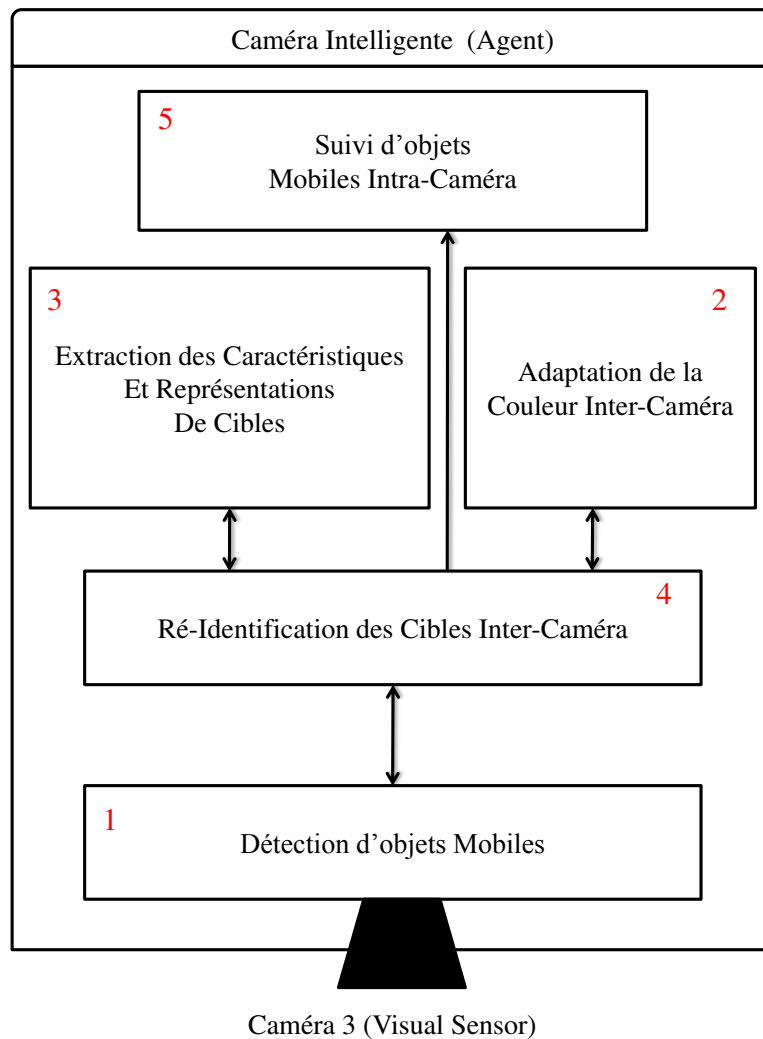


Fig. 5.1: Modèle générique d'agent proposé pour la modélisation des caméras (Caméra Intelligente)

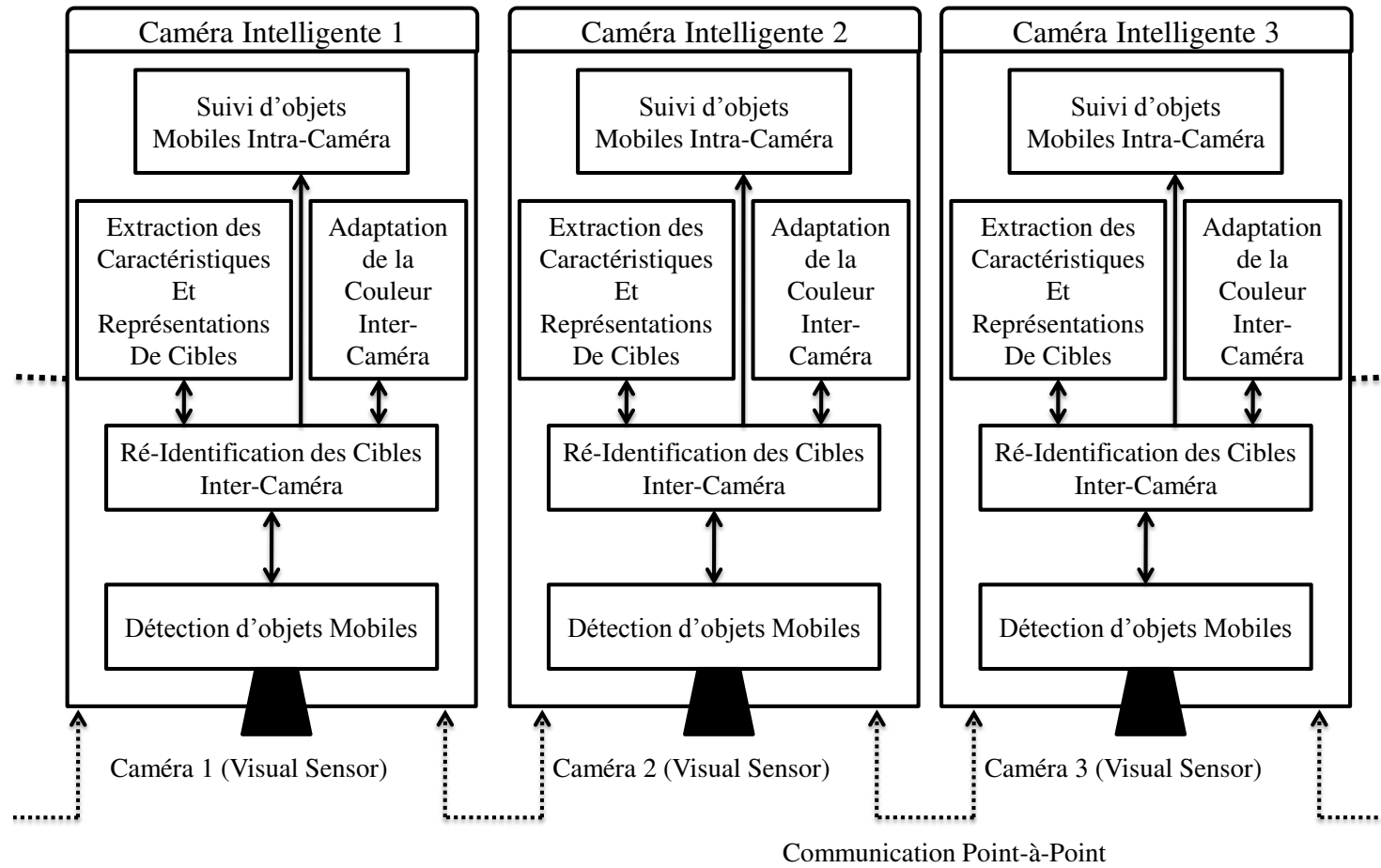


Fig. 5.2: Architecture générale du système distribué proposé basé sur le paradigme multi-agent pour la détection, la ré-identification et le suivi d'objets mobiles inter-caméra dans un réseau multi-caméra disjoints sans fil.

Dans le reste du manuscrit on utilise le terme « caméra intelligente » à la place de « une instance du modèle agent cognitif ».

En se basant sur l'architecture décentralisée proposée, le déroulement des opérations de suivi et la ré-identification des cibles inter-caméra est comme suit :

1. Si un nouveau objet Obj_{t_1} apparaitre dans le champ de couverture d'une caméra Cam_i , cet objet sera détecté et la carte de soustraction binarisée et étiquetée sera générer par le module.1 de la caméra intelligente Cam_i .

2. Une adaptation de la couleur inter-caméra de l'observation O_{obj_1} générer par l'objet détecté Obj_1 et capturée par la caméra Cam_i est assurée par le module.2. Cette adaptation est réalisée par la fonction de transfert de luminosité moyenne afin de diminuer la variation de la couleur inter-caméra, cet algorithme est détaillé dans la section suivante.

3. Une opération d'extraction des caractéristiques de cet objet Obj_{t_1} est effectuée après l'adaptation et le vecteur de représentation de l'apparence de l'observation O_{obj_1} basé sur une mixture d'indices visuels sera créé par le module.3.

4. Une opération de collaboration assurée par le module.4 est déclenchée entre les instances d'agent cognitif (cameras intelligentes) afin de parvenir à un consensus sur l'identité de l'objet Obj_{t_1} caractérisé par l'observation O_{obj_1} . Cet algorithme est détaillé dans la section suivante.

5. Une fois une identité valide est attribuée à l'objet d'intérêt Obj_{t_1} caractérisé par l'observation O_{obj_1} , alors le module de suivi intra-caméra (module.5) assure l'estimation des positions de l'objet Obj_1 à travers le filtre particulaire proposé "APF-MC" à été mixte.

Les modules intégrés dans le modèle générique d'agent cognitif proposé sont détaillés dans les sections suivantes.

Dans le reste du manuscrit on utilise le terme caméra intelligente à la place d'un agent intelligent.

5.2.2 Détection d'objets mobiles

La technique de détection d'objets mobiles à été détaillée dans le Chapitre.2, où une nouvelle chaine de traitement a été proposée afin de générer la carte de soustraction binarisée et étiquetée. La chaine de traitement proposée se base sur trois opérations successives : (1) Modélisation de l'arrière-plan par une mixture gaussienne (GMM) et la classification des pixels arrière-plan/premier-plan afin de générer une carte de soustraction binarisée. (2) filtrage de la carte par un post-traitement proposé dans ce travail, et qui se base sur une convolution de la carte de soustraction binarisée et érodée par un noyau gaussien afin d'éliminer la vibration, l'ombre et le mouvement saccadé [Bousetouane 2013b] . (3) Regroupement et étiquetage des zones connexes

afin d'extraire le masque d'objet en mouvement, pour faciliter le comptage des cibles. La détection d'objets mobiles est un module essentiel dans l'architecture proposé vu qu'il est le déclencheur de la collaboration entre les caméras intelligentes afin de parvenir à un consensus sur les identités des cibles traversant le réseau.

5.2.3 Adaptation de la Couleur inter-caméra

Les observations d'un objet dans un réseau multi-caméra à champs disjoints sont différentes d'une caméra à une autre, parce que ces observations sont capturées dans des conditions différentes, temps et espaces de captures différents, etc. La séparation en temps et en espace entre les observations capturées par des caméras différentes crée des variations remarquable inter-caméra entre les signatures visuelles du même objet. Ces variations influencent sur les opérations de la ré-identification et le suivi des cibles inter-caméra. L'adaptation de la couleur entre l'ensemble des caméras est une étape nécessaire afin de minimiser les variations dans les signatures visuelles des cibles d'une caméra à une autre. La figure suivante (Fig.5.3) illustre des images capturées par des caméras différentes de notre réseau multi-caméra à champs disjoints. Nous remarquons clairement que l'histogramme de couleur de la cible (Fig.5.3, *Objet₁*) capturée par la caméra $E - CAM - 1$ est totalement différente des histogrammes de couleur de la même cible capturées par les autres caméras $E - CAM - 2$, $E - CAM - 3$, $E - CAM - 6$, et $E - CAM - 7$. Cette différence entre les histogrammes crée une variation incontrôlable dans la signature visuelle de la cible inter-caméra. Sous telle condition les opérations de suivi et la ré-identification des cibles inter-caméra sont un grand défi.

Plusieurs chercheurs [Javed 2005], [Javed 2008b],[Prosser 2008],[Chen 2008], soulignent l'importance de l'adaptation de la couleur inter-caméra dans un réseau multi-caméra à champs disjoints. Ces chercheurs, ont montré que les opérations de suivi et la ré-identification des cibles inter-caméra ont été considérablement améliorés dans les réseaux multi-caméra avec une adaptation de la couleur inter-caméra. La section suivante illustre un état de l'art sur les techniques utilisées jusqu'à présent pour l'adaptation de la couleur inter-caméra dans les réseaux multi-caméra à champs disjoints.

5.2.3.1 État de l'art sur l'adaptation de la couleur inter-caméra

Plusieurs méthodes ont été développées récemment pour l'adaptation de la couleur entre chaque couple des caméras composant un réseau multi-caméra, nous citons ici les plus importantes. [Javed 2005], proposent l'utilisation d'une fonction de transfert de la luminosité (BTF, Brightness Transfer Function),



Fig. 5.3: Histogrammes de couleur d'un objet d'intérêt calculés sur cinq images capturées par cinq caméras différentes de notre réseau ad-hoc multi-caméra à champs disjoints

sous l'hypothèse que la luminosité soit constante dans chaque camera. La fonction BTF permet de cartographier les valeurs de la luminosité de chaque canal de couleur entre chaque couple de caméras. Une analyse en composante principale stochastique a été utilisée dans [Javed 2008b] pour l'apprentissage d'un sous-espace de dimension basse afin d'adapter la couleur inter-caméra. Dans [Javed 2008a], les auteurs proposent une extension de la fonction BTF vers la mise en correspondance non seulement de la luminosité mais aussi les caractéristiques spatio-temporelle (temps de transfert, vitesse des objets, identités des caméras, etc.). [Prosser 2008], proposant une extension de la fonction de transfert de la luminosité vers une fonction de transfert de luminosité bidirectionnelle. Ces approches sont basées sur l'hypothèse que la luminosité reste constante dans toutes les caméras. D'autres chercheurs proposent des mécanismes d'adaptation de la couleur inter-caméra basant sur des techniques d'apprentissage afin de résoudre le problème de la variation de la couleur au fil du temps. Au lieu d'apprendre la fonction de transfert de la luminosité dans une étape hors ligne, [Chen 2008] proposent de mettre à jour la fonction BTF au fil du temps utilisant les dernières observations des cibles qui ont été suivies. Un algorithme de clustering (K-means) a été utilisé dans [Madden 2007] pour faciliter le calcul de la moyenne de la luminosité d'objets. Des techniques de constance de la couleur ont été utilisées dans [Gilbert 2006] pour compenser le changement de la luminosité, où des nouveaux espaces de couleur ont été développés. Un modèle d'apprentissage incrémental a été proposé dans [Chen 2011] pour la modélisation de la variation de la couleur et les liens spatio-temporelle (i.e les points d'entrées/sorties) entre l'ensemble des caméras. Parmi les approches présentées précédemment pour l'adaptation de la couleur entre chaque couple des caméras composant le réseau, la fonction de transfert de luminosité (BTF) avec ses variantes (MBTF, CBTF, ABTF, BBTF, etc.) est la plus utilisée jusqu'à présent. La fonction de transfert de luminosité BTF, assure une cartographie de la couleur entre chaque couple de caméras dans le réseau à travers la mise en correspondance entre l'histogramme de couleur d'une observation dans une camera avec l'histogramme de couleur de la même observation dans une autre caméra. Dans les travaux de [Prosser 2008], une comparaison a été réalisée entre deux variantes de la fonction de transfert de luminosité dans un réseau à deux caméras indoor : la fonction de transfert de luminosité moyenne (MBTF) et la fonction de transfert de luminosité cumulative (CBTF). D'après les expérimentations réalisées, les chercheurs constatent que la fonction de transfert de luminosité moyenne est mieux adaptée dans la cartographie de la couleur en présence d'une variation importante par rapport à la fonction de transfert de luminosité cumulative. Dans ce travail, la fonction de transfert de luminosité moyenne (MBTF) a été choisie pour l'adaptation et la cartographie de la couleur inter-caméra dans notre réseau multi-caméra

à champs disjoint sans fil. Cette opération d'adaptation permet de compenser les variations de la luminosité dans les signatures visuelles des cibles inter-caméra.

5.2.3.2 Fonction de transfert de luminosité

Soit l'observation O_{Cam_i} capturée par la caméra Cam_i et O_{Cam_j} une observation capturée par la camera Cam_j de la même personne. La fonction de transfert de luminosité inter-caméra est notée par $F_{i,j}$, cette fonction permet de cartographier une valeur de luminosité quelconque B_u de O_{Cam_i} vers une valeur qui lui correspond B_v de O_{Cam_j} . Cette mise en correspondance est réalisée au niveau des pixels.

$$B_v = F_{i,j}(B_u) \quad (5.1)$$

La fonction de transfert de luminosité entre les deux observations O_{Cam_i} et O_{Cam_j} est calculer comme suit :

Soit H_i l'histogramme normalisé calculé à partir de l'observation O_{Cam_i} et H_j l'histogramme normalisé calculé à partir de l'observation O_{Cam_j} .

Supposons que le pourcentage des points dans l'observation O_{Cam_i} avec une luminosité inférieur ou égale à B_u est égale au pourcentage des points de l'image dans l'observation O_{Cam_j} avec une luminosité inférieur ou égale à B_v . Cette hypothèse [Javed 2008b] permet la création de l'équation suivante :

$$H_i(B_u) = H_j(B_v) = H_j(F_{i,j}(B_u)) \quad (5.2)$$

Par le calcul de l'histogramme inverse nous pouvons déduire l'équation de la fonction de transfert de la luminosité :

$$F_{i,j}(B_u) = H_j^{-1}(H_i(B_u)) \quad (5.3)$$

Cette fonction $F_{i,j}$ est calculée entre chaque couple des caméras composant le réseau afin de cartographier la couleur inter-caméra.

Dans ce travail, la fonction de transfert de luminosité moyenne *MBTF* a été utilisée pour adapter la luminosité entre chaque couple de caméras dans le réseau, cette fonction moyenne (MBTF) est calculée comme suit :

Soit O l'ensemble d'observations générer par plusieurs objets, soit $M = F_1, F_2, F_3, \dots, F_n$ les fonctions de transfert de luminosité calculées à partir de ces observations. La fonction moyenne de transfert de luminosité *MBTF* est calculée par l'équation suivante :

$$MBTF = \frac{\sum f_i}{n} \quad (5.4)$$

Cette fonction est calculée pour chaque paire de caméras dans le réseau. La figure suivante (Fig.5.4) illustre des images d'objets d'intérêts que nous avons choisies pour calculer la fonction de transfert de luminosité moyenne (MBTF) dans l'étape d'apprentissage. La première ligne présente les objets capturés par la caméra $E - CAM - 1$, la deuxième et la troisième lignes sont respectivement les mêmes objets capturés par la caméra $E - CAM - 2$ et la caméra $E - CAM - 4$. La figure (Fig.5.5) illustre graphiquement les fonctions de transfert de luminosité moyenne MBTFs calculées entre la caméra $E - CAM - 1$ et la caméra $E - CAM - 2$ à partir des observations des trois objets d'intérêts. En se basant sur la fonction de transfert de luminosité moyenne (MBTF) calculée entre chaque couple des caméras composant le réseau, nous pouvons cartographier la couleur. Cette opération nous permet de compenser la variation de la couleur inter-caméra à n'importe quel moment afin d'effectuer des traitements comme l'opération de la ré-identification et le suivi des cibles inter-caméra.

A partir des histogrammes de couleurs illustrés dans la figure (Fig.5.6) calculés après l'adaptation de la couleur inter-caméra, nous remarquons que la distribution de la couleur de l'objet d'intérêt $Objet_1$ capturé par la caméra $E - CAM - 1$ est très proche de la distribution de la couleur du même objet capturé par la caméra $E - CAM - 2$. A partir de cette expérimentation nous constatons que : malgré la séparation spatio-temporelle des deux observations du même objet, la fonction de transfert de luminosité moyenne (MBTF) a compensé cette variation à travers la cartographie de la couleur inter-caméra. Le module d'adaptation de la couleur inter-caméra intégré dans chaque agent intelligent (caméra intelligente) de notre architecture est essentiel pour l'opération de la ré-identification et le suivi des cibles. Ce module permet de minimiser les variations dans les signatures visuelles des cibles inter-caméra.

5.2.4 Extraction d'indices visuels et représentation des cibles

Après l'adaptation de la couleur entre chaque couple des caméras composant le réseau afin de minimiser la variation de la couleur inter-caméra, l'opération d'extraction d'indices visuels est nécessaire afin de coder les signatures visuelles des cibles. Pour la représentation des cibles nous avons proposé dans le chapitre.3 un modèle vectoriel basé sur une mixture d'informations contextuelles et des moments statistiques de différents ordres. Les informations contextuelles de bas niveau proposées sont calculées à partir de la dépendance spatiale et la cooccurrence entre les pixels composant l'image de la cible, en utilisant les indices de texture de Haralick [Bousetouane 2013b]. Le modèle proposé dans le cadre de cette a été prouvé comme discriminant



Fig. 5.4: Observations d'objets d'intérêts capturées par trois caméras de notre réseau multi-caméra, E-CAM-1 (Sony), E-CAM-2 (Sony-G) et E-CAM-4 (Sony-G-lens).

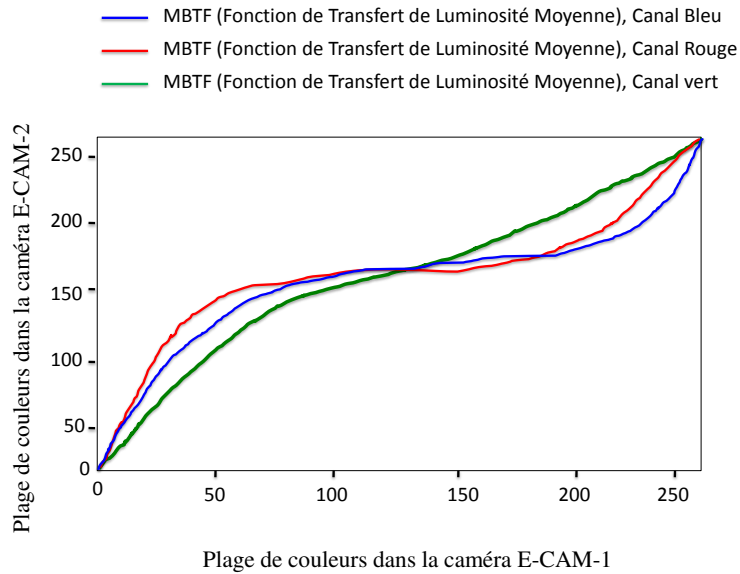


Fig. 5.5: Illustration graphique des fonctions de transfert de luminosité moyenne calculées inter-caméra ($E - CAM - 1$, $E - CAM - 2$) pour chaque canal de couleur (RVB)

inter-objet, invariant intra-caméra et inter-caméra (section.3.6), ce qui assure un taux de ré-identification important. Les opérations d'extraction d'indices visuels et la représentation des cibles ont été détaillées dans le chapitre.3.

5.2.5 Ré-identification distribuée des cibles Inter-caméra

L'opération d'observation d'un objet dans le champ de couverture d'une caméra et sa reconnaissance par la même caméra ou par une autre caméra est souvent appelée opération de ré-identification d'objet inter-caméra. Cette opération a une grande importance pour la sécurité des larges espaces comme : les aéroports, les stations de bus, les centres commerciaux, etc. Dans cette section, nous présentons le nouvel algorithme que nous proposons pour la ré-identification d'objets mobiles d'une manière totalement décentralisée dans un réseau multi-caméra à champs disjoints sans fil. Basant sur l'architecture multi-agent proposée dans (la section.5.2.1), l'algorithme proposé assure une collaboration entre les caméras intelligentes (les instances du modèle d'agent cognitif proposé) afin de parvenir à un consensus sur les identités des cibles traversant le réseau.

Dans la terminologie multi-agent [Soto 2009], [Olfati-Saber 2007], un consen-

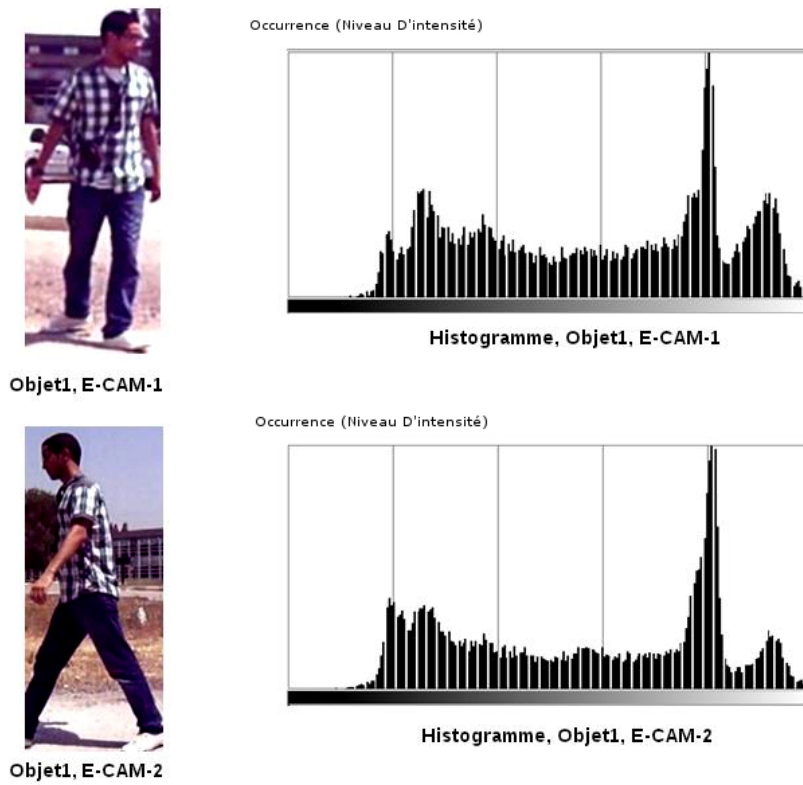


Fig. 5.6: Histogrammes de couleur des deux observations d'un même objet capturées par les caméras E-CAM-1 et E-CAM-2 respectivement après l'adaptation de la couleur inter-caméra (E-CAM-1 , E-CAM-2)

sus est défini comme un moyen (Algorithme, Protocole) pour parvenir à un accord concernant certaines quantités d'intérêts qui dépendent de l'état interne de l'agent ou d'autres informations capturés par les systèmes perceptifs des agents dans un environnement multi-agent. Dans notre cas, un consensus est défini comme un algorithme qui permet de définir les règles d'interaction pour l'échange d'information et les connaissances entre une caméra intelligente et ses voisins dans le réseau. Toutes les caméras intelligentes de notre réseau multi-caméra collaborent afin de parvenir à un consensus sur l'identité d'un ou plusieurs objets traversant le réseau. Pour plus de détails sur la notion de consensus, voir notre publication [Bousetouane 2013c].

5.2.5.1 Algorithme proposé basé sur le Consensus

La topologie d'interaction entre les caméras intelligentes est représentée par le graphe $G(V, E)$ avec V l'ensemble de nœuds, $V = 1...N_c$ où N_c est le nombre totale des cameras dans le réseau. E les arrêtes entre les nœuds, où $E \subset (V * V)$. La topologie de réseau multi-caméra est totalement dynamique, aucune restriction n'est utilisée contrairement aux approches proposées dans l'état de l'art, [Javed 2008a], [Chen 2011]. Chaque caméra intelligente possède une identité et une table de données (DataSet) nommé T_a qui contient l'historique de la communication entre les caméras intelligentes. Soit CAM_x l'ensemble des cameras intelligentes dans le réseau. Nous définissons un sous-ensemble des cameras intelligentes $Sub_C \subset (CAM_x)$ où un objet en mouvement a été déjà détecté. O_{Cam_i} est l'observation générée par l'objet Obj_1 dans la camera Cam_i où $Cam_i \subset (Sub_C)$, donc toutes les caméras intelligentes qui ont déjà détectées un ou plusieurs objets mobiles sont supposées des voisins. $Ned_C \subset (CAM_x)$ est l'ensemble des caméras où aucun objet n'a été déjà détecté, $Sub_C \cup Ned_C = CAM_x$. Dans le système distribué proposé la camera intelligente qui a détecté une nouvelle observation sera l'initiatrice de la communication avec les agents (caméras intelligentes) voisins, afin de parvenir à un consensus sur l'identité de cette observation. La diffusion des messages entre les caméras intelligentes est utilisée dans l'architecture proposée. Supposons que les fonctions de transfert de la luminosité moyenne (MBTFs) (voir la section.5.2.3) entre l'ensemble de K paire de caméras ont été déjà calculées, où $k = n * (n - 1)/2$. Pour tout paire de caméras Cam_i et Cam_j dans le réseau (CAM_x, E) , $F_{i,j}$ est calculée tel que $i \neq j$ et $i, j = 1...N_c$.

Soit $e_{OCami} = e_1, e_2, ...e_n$, le vecteur d'indices visuels (voir chapitre.3) qui caractérise la signature visuelle de l'objet Obj_1 calculée à partir de l'observation O_{Cam_i} , où $Cam_i \subset (Sub_C)$. L'algorithme proposé (voir notre publication, [Bousetouane 2013c]) pour la ré-identification d'objets mobiles inter-caméra dans un réseau multi-caméra à champs disjoints est décomposé en quatre

étapes séquentielles :

1. Si une nouvelle observation O_{Cam_i} est détectée dans la caméra intelligente Cam_i , alors la collaboration entre cette caméra Cam_i et les autres caméras voisines $Cam_j \subset (Sub_C)$ est lancer afin de parvenir à un consensus sur l'identité de cette observation.

2 POUR chaque caméra intelligente $Cam_j \subset (Sub_C)$ **FAIRE**

a. Calculer les fonctions de transfert de luminosité moyenne MBTFs entre Cam_i et ses voisins $Cam_j \subset (Sub_C)$.

b. Crée le vecteur des caractéristiques visuelles e_{OCam_i} de l'observation O_{Cam_i} après l'adaptation de la couleur entre Cam_i et ses voisins.

3. Envoi d'un message à ses voisins $M_{OCam_i} = (e_{OCam_i}, Cam_i)$

FIN POUR

4.Réception d'un message de chaque caméra Cam_j voisine de Cam_i $M_j = (Id_j, Cam_j, Bol)$, $Cam_j \subset (Sub_C)$.

Si l'observation O_{Cam_i} de l'objet Obj caractérisée par le vecteur e_{OCam_i} a été déjà identifiée dans une ou plusieurs caméras intelligentes voisines $Cam_j \subset (Sub_C)$, alors la variable Bol prend la valeur *Vrai* et l'identité Id_j sera attribuée à l'objet Obj . **Sinon** l'objet Obj détecté par la caméra Cam_i et caractérisé par e_{OCam_i} est un nouveau objet dans le réseau où $Bol = Faux$, alors la caméra Cam_i où l'objet obj a été détecté attribue une identité globale a l'objet Obj . L'identité globale est calculée par le maximum entre les dernières identités attribuées par les caméras voisines e_{OCam_i} , $Id_i = Max(Id_j) + 1$. L'identité attribuée à l'objet Obj , le vecteur qui caractérise l'objet e_{OCam_i} et l'identité de la caméra qui à identifié l'objet Cam_i seront enregistrés dans la table de données T de chaque caméra intelligente pour garder une trace des cibles circulant dans le réseau. La forme algorithmique de l'approche proposée pour la ré-identification distribuée des cibles inter-caméra est illustre dans la figure suivante (Fig.5.7) :

La capacité discriminatoire du modèle de représentation des cibles inter-caméra proposé permet de résoudre l'ambiguïté dans la décision locale de la caméra intelligente $Cam_i \subset (Sub_C)$ afin d'attribuer une identité globale a l'objet d'intérêt obj . En réalité, une ambiguïté peut avoir lieu, surtout si deux objets portant les mêmes vêtements. Dans ce cas, d'autres techniques peuvent être intégrées dans notre algorithme pour la résolution de cette ambiguïtés, à titre d'exemple, l'utilisation des contraintes spatio-temporelle comme dans

Consensus-Based Algorithm for distributed inter-camera objects re-identification

Input : **Cami**, **OCami**, **Fij**, **T**, **SubC**,

Output: **Id_OCami**

Begin

For each Camera Camj \subseteq **SUBc do**

1. Compute the Color Transformation MBTFs, **Fij(Bu)=Hj-1(Hi(Bu))**
2. Compute $F_{O_{CAM^i}} = \{f_1, f_2, \dots, f_{10}\}$
 - a. Compute the Color histogram **H(OCami)**
 - b. Compute the Statistic Moments
 - c. Compute the proposed low-level contextual information (Haralick Features)
3. Send Message $M_{O_{CAM^i}} = (MBTF_S^{i,j}(F_{O_{CAM^i}}), CAMi)$

End For each

Receive messages $M_{O_{CAM^i}} = (ID^j(F_{O_{CAM^i}}), CAMj, Bool)$ from all cameras **Camj** \subseteq **SUBc**

If (Bol= true) then **Id_OCami = IDj**

Else **Id_OCami = Max(IDj)+1**

Store (**Cami**, **Id_OCami**, $F_{O_{CAM^i}}$) in **T**

End

Fig. 5.7: L'algorithme proposé pour la ré-identification des cibles mobiles inter-caméra basé sur le consensus

les travaux de [Motamed 2007], enrichir la décision globale par des croyances sur les caméras intelligentes (transferable belief model,[Smith 2006]), etc. Une fois les caméras intelligentes voisines $Cam_j \subset (Sub_C)$ reçoivent un message de la caméra intelligente initiatrice de la collaboration inter-caméra Cam_i , une opération d'identification intra-caméra est effectuée. Cette opération permet de vérifier l'existence de la signature visuelle de l'objet Obj caractérisé par e_{OCami} dans la table de données T de chaque caméra intelligente $Cam_j \subset (Sub_C)$. L'opération d'identification des cibles intra-caméra se base sur la distance de Mahalanobis (voir chapitre.3, section.3.5). Cette distance permet de calculer la similarité entre les caractéristiques e_{OCami} de l'objet d'intérêt envoyées par la caméra Cam_i et les vecteurs des caractéristiques e_{OObj} d'objets déjà identifiées par la caméra voisine $Cam_j \subset (Sub_C)$. Plusieurs expérimentations réalisées dans notre réseau multi-caméra à champs disjoints sans fil prouvent la robustesse de notre algorithme pour la ré-identification d'objets inter-caméra d'une manière totalement distribuée et en temps réelle.

La section.5.3 illustre plus de détails sur les expérimentations réalisées.

5.2.6 Suivi d'objets Mobiles Intra-Caméra

La technique de suivi d'objets mobiles intra-caméra utilisée dans ce module a été détaillé dans le chapitre.4, où un nouveau filtre particulière adaptatif «APF-MC » a été proposé, voir notre publication [Bousetouane 2013a]. Le filtre particulière proposé se base essentiellement sur deux principes : (1) Adaptation de l'espace de recherche de chaque particule avec la dimension réelle de la cible pour améliorer la convergence du filtre particulière classique dans le cas de mouvement non-rigide complexe et au changement d'échelle. (2) L'intégration d'une mixture d'indices visuels dans le filtre pour éviter les états faux positifs et garder la robustesse dans le cas de la multi-modalité. Le modèle de vraisemblance intégré dans le filtre particulière proposé a été prouvé comme invariant intra-caméra et discriminant inter-objets (chapitre.3). Plusieurs instances du filtre particulière proposé APF-MC sont utilisées dans un contexte distribué pour le suivi multi-objet où une instance par objet d'intérêt. L'algorithme proposé "APF-MC" a été détaillé dans le chapitre.4, section.4.3. Après l'identification d'objets mobiles inter-caméra utilisant l'algorithme proposé basé sur un consensus entre les caméras intelligentes, chaque filtre particulière doit être en mesure de suivre un objet spécifique où une gestion d'identités est nécessaire dans chaque filtre. Pour assurer une gestion d'identités dans le suivi nous proposons l'extension de l'état de chaque filtre particulière APF-MC. Chaque particule incorpore non seulement la position de l'objet au fil du temps mais aussi l'identité de l'objet. L'état étendu de

chaque particule est décrit comme suit :

$$X = (x_p, y_{id}^{(i)}) \quad (5.5)$$

$x_p = (x, y, H, W)^T$, $xety$ sont les coordonnées du centre d'un rectangle, H et W sont respectivement la largeur et la hauteur du rectangle qui encadre l'objet d'intérêt (la cible).

$y_{id}^{(i)}$ est la i^{me} identité de l'objet d'intérêt où $i = 1, \dots, N_{id}$, N_{id} est le nombre d'objets détectés dans le réseau et égale au nombre d'instances du filtre particulière.

La section suivante illustre des détails d'ordre applicatif de la solution proposée pour le suivi et la ré-identification d'objets mobiles inter-caméra dans un réseau multi-caméra à champs disjoints sans fil.

5.3 Expérimentations

Cette section illustre une série d'expérimentations que nous avons effectuées en utilisant l'approche proposée pour le suivi et la ré-identification distribuée des cibles inter-caméra dans un réseau multi-caméra à champs disjoints sans fil. Dans ces expérimentations, nous avons utilisé des séquences d'images capturées par sept caméras de notre réseau multi-caméra à champs disjoints sans fil (Voir, chapitre.1 (Fig.1.3)). Les vidéos ont été capturées dans quatre scénarios déférents où six objets d'intérêts circulant d'une manière aléatoire inter-caméra dans notre réseau, trois caméras intelligentes sont utilisées par scénario. Aucune information préalable n'a été utilisée sur les liens d'entrées/sorties, le temps de transfert, la durée d'apparition, la couleur d'objets, etc. Les fonctions de transfert de luminosités moyenne (MBTF) entre chaque K paire de caméras ont été calculées Hors-ligne. Toutes les expérimentations ont été réalisées en-ligne. Trois gammes des caméras sont utilisées dans notre réseau multi-caméra : Sony, Sony-G, Sony-G-lens, chaque caméra est connectée à une machine différente. Les configurations des trois machines utilisées sont :

Machine.1 : HP, CPU Intel CoreTM 2 Duo T2300 1.66 GHz, RAM 2 GB.

Machine.2 : Packard bell, notebook-one, CPU Intel atom n450 1.66 GHz, RAM 1GB.

Machine.3 : Macbook-Air, CPU Intel Core i5 2,3 GHz, RAM 2GB.

Toutes les machines sont dotées d'un système d'exploitation Linux-distribution-Ubuntu sauf le MacBook qui est doté d'un système Mac-os-x 10.6. Le système

multi-agent (multi-caméra intelligente) proposé a été implémenté en utilisant la bibliothèque YARP (Yet Another Robot Platform). Les modules intégrés dans chaque caméra intelligente et qui assurent des traitements de vidéos à différents niveaux ont été implémentés en utilisant la bibliothèque de la vision artificielle OpenCv-2.4, développée par la société Intel. Le langage C++ a été utilisé pour la programmation des algorithmes proposés dans le cadre de cette thèse. Les résultats de suivi et la ré-identification distribuée des cibles inter-caméra dans les différents scénarios sont détaillés dans la section suivante.

5.3.1 Résultats et discussions

Quatre scénarios différents ont été réalisés afin d'évaluer le système proposé pour le suivi et la ré-identification des cibles inter-caméra. Les scénarios consistent en six objets différents circulant dans notre réseau multi-caméra dans des chemins d'entrées/ sorties différents : (1) Le 1^{er} scénario consiste en trois objets entrant dans le champ de couverture de la caméra E-CAM-1, après ils sortent du champ de couverture de cette caméra et entrent dans le champ de couverture de la caméra E-CAM-2, en suite ils entrent dans le champs de couverture de la caméra E-CAM-3 (chemin rouge, Figure.5.8). (2) Le 2^{eme} scénario consiste en deux objets entrent et sortent dans le champ de couverture de la caméra E-CAM-4 et après entrent dans le champ de couverture de la caméra E-CAM-5, en suite ils sortent du champ de couverture de la caméra E-CAM-5 et entrent dans le champ de couverture de la caméra E-CAM-6 (chemin jaune, Figure.5.8). (3) Le 3^{eme} scénario de circulation d'objets consiste en deux objets déjà identifiées et suivi dans le 2^{eme} scénario entrent dans le champ de couverture de la caméra E-CAM-6. Après qu'ils sortent du champ de couverture de cette caméra ils entrent dans le champ de couverture de la caméra E-CAM-7, ils entrent ensuite dans le champ de couverture de la caméra E-CAM-5 (chemin marron, Figure.5.8). (4) Le 4^{eme} scénario consiste en une circulation aléatoire des objets dans le réseau où les deux objets d'intérêts entrent dans le champ de couverture de la caméra E-CAM-3, après ils entrent dans le champ de couverture de la caméra E-CAM-6 et en fin entrent dans le champ de couverture de la caméra E-CAM-1 (chemin Bleu, Figure.5.8). Dans tous ces scénarios, le système proposé pour le suivi et la ré-identification d'objets mobiles inter-caméra est bien évalué, ici nous ne détaillons que les trois premiers scénarios. les expérimentations effectuées et les résultats obtenus sont publiés dans ma page web personnelle, <http://bousfouad.webatu.com/>.

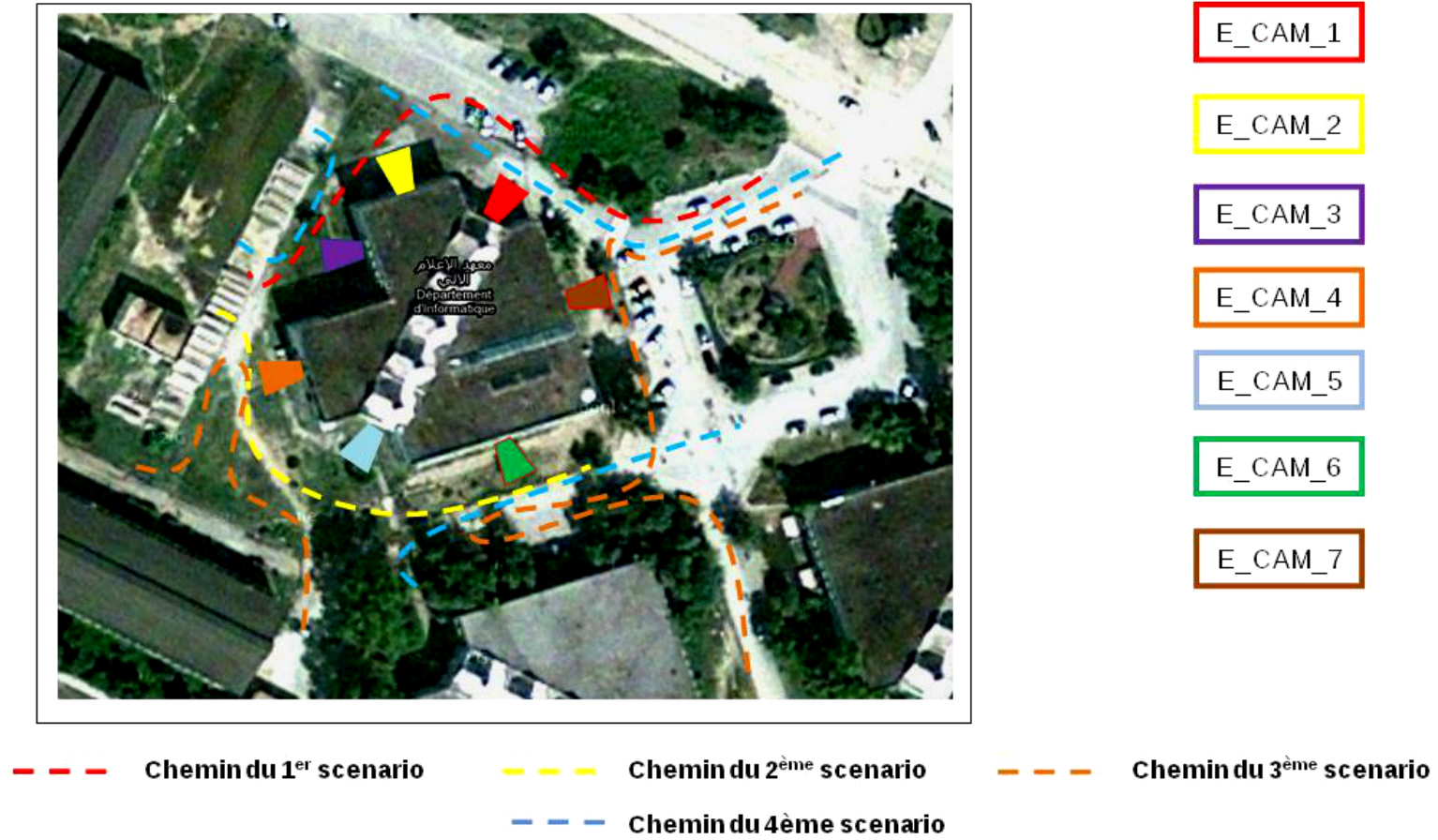


Fig. 5.8: Chemins des scénarios utilisés dans les expérimentations

5.3.1.1 Résultat.1

Le 1^{er} chemin a été utilisé dans cette expérimentation où trois vidéos ont été capturées par les caméras E-CAM-1, E-CAM-2 et E-CAM-3. Ces vidéos consistent en trois objets mobiles circulant dans les trois caméras en présence des conditions complexes : occultation complète, mouvement non-rigide, changement d'échelle, variation de la couleur inter-caméra, mouvement saccadé en arrière-plan causé par les arbres et les voitures, etc. Dans toutes ces conditions le suivi et la ré-identification d'objets inter-caméra est un grand défi. La figure (Fig.5.9) illustre les résultats d'application de notre approche basée sur le paradigme multi-agent sur le 1^{er} scénario pour le suivi et la ré-identification des trois objets inter-caméra. Les fenêtres englobant les objets d'intérêts représentent la localisation spatiale d'objets, le centre de chaque fenêtre contient l'identité de l'objet d'intérêt. A partir des résultats de suivi et la ré-identification des trois objets d'intérêts illustrés dans la figure (Fig.5.9), nous remarquons clairement que : dès l'apparition d'objets dans le champ de couverture de la caméra E-CAM-1, des identités ont été attribuées à ces objets en temps réel (*Objet*₁ avec *Id* = 1, *Objet*₂ avec *Id* = 2 et *Objet*₃ avec *ID* = 3). Une fois les objets sortent du champ de couverture de la caméra E-CAM-1 et entrent dans le champ de couverture de la caméra E-CAM-2, les mêmes identités ont été attribuées à ces objets par cette caméra. Enfin, dès l'apparition de ces objets dans le champ de couverture de la caméra E-CAM-3, les mêmes identités attribuées à ces objets par les autres caméras intelligentes ont été attribuées par la caméra E-CAM-3. Cette expérimentation prouve que la ré-identification réalisée dans le 1^{er} scénario est totalement valide où les identités attribués aux objets d'intérêts par la caméra E-CAM-1 sont les mêmes attribuées par les autres caméras. Cette re-identification prouve la robustesse de l'algorithme proposé basé sur la collaboration entre les caméras intelligentes, malgré la présence d'occultation partielle dans l'image (frame 1169, E-CAM-3) et une occultation complète de l'objet *Objet*₁ dans l'image (frame 1130, E-CAM-3). Bien que la détection et la ré-identification d'objets ont été parfaitement réalisées dans cette expérimentation, l'estimation des positions d'objets d'intérêts a été aussi correctement effectuée où les trois instances de l'algorithme proposé "APF-MC" à état mixte convergent vers les meilleurs positions d'objets d'intérêts.

5.3.1.2 Résultat.2

Le 2^{me} chemin a été utilisé dans cette expérimentation où trois vidéos ont été capturées par les caméras E-CAM-4, E-CAM-5 et E-CAM-6. Ces vidéos consistent en deux objets mobiles circulant dans les trois caméras en présence des conditions complexes. La figure (Fig.5.10) illustre les résultats

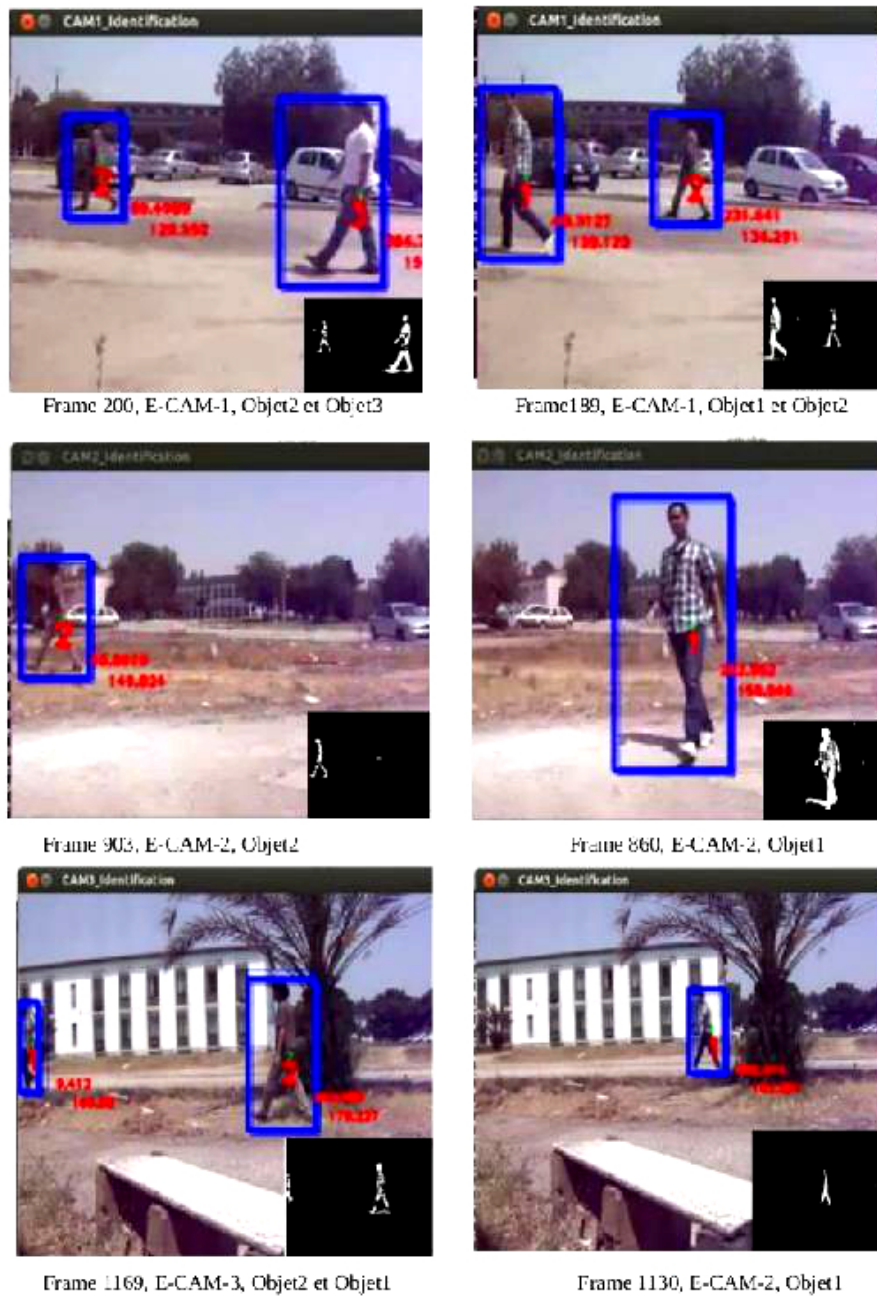


Fig. 5.9: Suivi et Ré-identification des trois objets d'intérêts ($Objet_1$, $Objet_2$, $Objet_3$) inter-caméra en utilisant le chemin du 1^{er} scénario dans notre réseau multi-caméra à champs disjoints sans fil.

d'application de notre approche sur le 2^{me} scénario pour le suivi et la ré-identification des deux objets inter-caméra. Les fenêtres englobant les objets d'intérêts représentent la localisation spatiale d'objets, le centre de chaque fenêtre contient l'identité de l'objet d'intérêt.

A partir des résultats de suivi et la ré-identification des deux objets d'intérêts illustrés dans la figure (Fig.5.10), nous remarquons que : les identités attribuées aux objets d'intérêts par la caméra E-CAM-4 sont identiques aux identités attribuées par les autres caméras (E-CAM-5 et E-CAM-6) à ces objets. Cette re-identification, prouve la robustesse de l'algorithme proposé malgré la présence d'occultation partielle dans l'image (frame 1602, E-CAM-4) et une rotation de l'objet *Objet₁* dans l'image (frame 2723, E-CAM-6). Après la ré-identification d'objets d'intérêts inter-caméra, l'estimation des positions de ces d'objets a été aussi correctement effectuée où les fenêtres en blue englobent les objets d'intérêts (*Objet₁* et *Objet₂*) convergent vers les meilleurs positions.

5.3.1.3 Résultat.3

Le 3^{me} chemin a été utilisé dans cette expérimentation où trois vidéos ont été capturées par les caméras E-CAM-6, E-CAM-7 et E-CAM-5. Ces vidéos consistent en deux objets mobiles circulant d'une manière aléatoire dans les trois caméras en présence d'une variation de luminosité importante inter-caméra. La figure (Fig.5.11) illustre le résultat d'application de notre approche sur le 3^{me} scénario pour le suivi et la ré-identification des deux objets inter-caméra. A partir des résultats de suivi et la ré-identification des deux objets d'intérêts illustrés dans la figure (Fig.5.11), nous remarquons que : les identités attribuées aux objets d'intérêts par la caméra E-CAM-6 sont identiques aux identités attribuées par les autres caméras (E-CAM-7 et E-CAM-5). Cette expérimentation prouve encore la robustesse et la puissance décisionnelle de l'algorithme proposé pour la ré-identification et le suivi temp-réel d'objets mobiles inter-caméra.

Les expérimentations réalisées en temps réel, en ligne et en présence des conditions complexes prouvent la robustesse et la capacité du système proposé pour le suivi et la ré-identification d'objets mobiles dans un réseau multi-caméra à champs disjoints sans fil. L'algorithme proposé pour la ré-identification d'objets mobiles inter-caméra a été bien testé dans ces expérimentations et les résultats obtenus (Fig.5.9, Fig.5.10, Fig.5.11) prouvent la validité et la robustesse de notre algorithme. Le système multi-agent proposé est invariant à l'échelle où l'ajout ou la suppression des instances d'agent (caméras intelligentes) n'influence pas sur la capacité de suivi et la ré-identification d'objets mobiles inter-caméra. La section suivante illustre une évaluation de

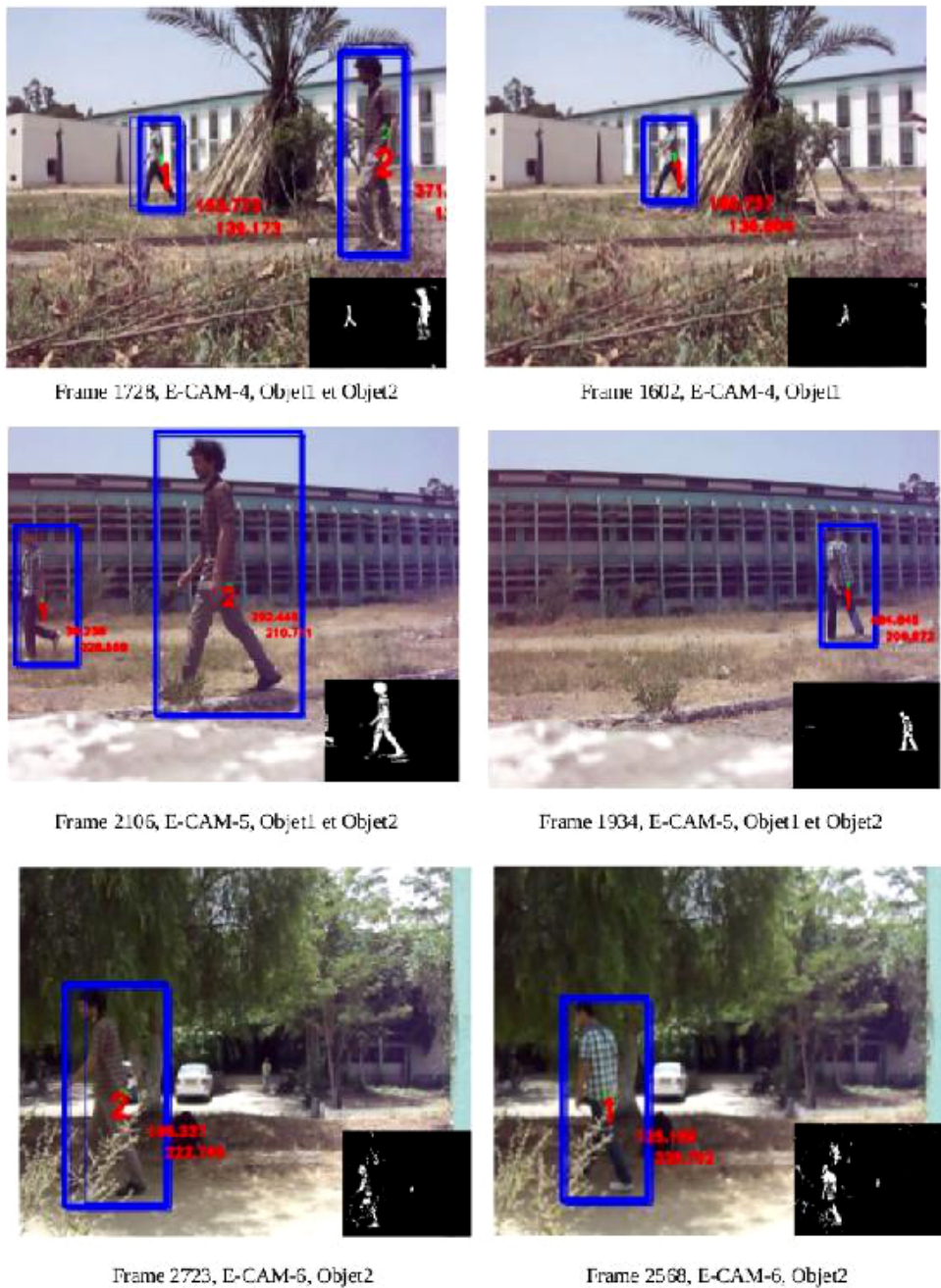


Fig. 5.10: Suivi et Ré-identification des deux objets d'intérêts (*Objet₁*, *Objet₂*) inter-caméra en utilisant le chemin du 2^{me} scénario dans notre réseau multi-caméra à champs disjoints sans fil.

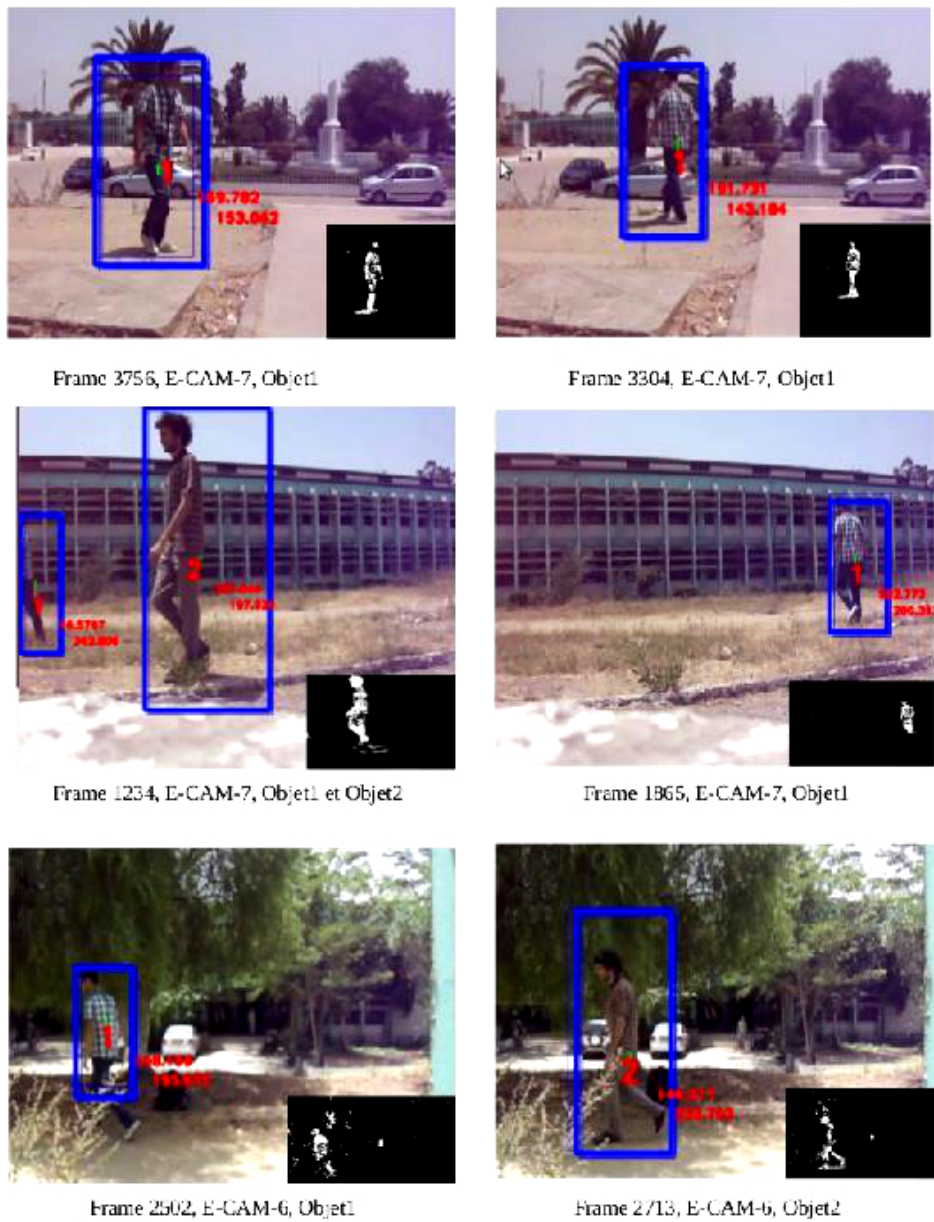


Fig. 5.11: Suivi et Ré-identification d'objets d'intérêts ($Objet_1$, $Objet_2$) inter-caméra en utilisant le chemin du 2^{me} scénario dans notre réseau multi-caméra à champs disjoints sans fil.

l'algorithme proposé basé sur la notion du consensus pour la ré-identification d'objets inter-caméra. Le suivi intra-caméra et qui se base sur le filtre particulaire proposé APF-MC a été déjà évalué dans le chapitre.4.

5.4 Evaluation

Pour évaluer la performance de l'algorithme proposé pour la ré-identification distribuée d'objets mobiles inter-caméra, nous avons choisie l'utilisation des deux espaces d'évaluation : l'espace *ROC* (Receiver Operating Characteristic) et l'espace *PR* (Precision Recall). Ces espaces d'évaluations représentés visuellement par des courbes sont très utilisés dans plusieurs domaines pour l'évaluation des algorithmes de classification en vision artificielle, fouille de données, intelligence artificielle, etc. Dans ce travail nous avons calculé les courbes *ROC* et *PR* à partir des résultats de la ré-identification de notre algorithme sur les quatre scénarios détaillés précédemment. La précision (Precision), Rappel (Recall), taux des vrais positifs et taux des faux positifs sont calculés comme suit :

$$Rappel = \frac{VP}{VP + FN} \quad (5.6)$$

$$Prcision = \frac{VP}{VP + FP} \quad (5.7)$$

$$TauxdeVraisPositifs = TVP = \frac{VP}{VP + FN} \quad (5.8)$$

$$TauxdeFauxpositifs = TFP = \frac{FP}{FP + VN} \quad (5.9)$$

Les quantités *VP*, *FN*, *FP*, *Prcision*, *Rappel* et *VN* sont définies comme suit :

Vrais Positifs (VP) : Si une ancienne identité est correctement attribuée par notre algorithme à un objet d'intérêt spécifique, où bien une nouvelle identité est correctement attribuée à un nouveau objet, alors un vrai positif *VP* est détecté.

Faux Positifs (FP) : Si une ancienne identité est attribuée par notre algorithme aux faux objets, alors cette attribution est considérée comme *FP*.

Faux Négatifs (FN) : Si une nouvelle identité est créée par notre algorithme et attribuée à un objet déjà identifié, alors cette opération est considérée comme *FN*.

Vrais Négatifs (VN) : Si une ancienne identité est attribuée par notre algorithme à un nouvel objet, alors cette attribution est considérée comme *VN*.

Dans ce travail, les deux espaces d'évaluation ont été calculés par une variation temporelle t du seuil de la re-identification inter-caméra : par exemple dans le 1^{er} scénario et après l'application de notre algorithme pour la ré-identification d'objets, trois séquences d'images d'une durée de 400 images consécutives par caméra ont été utilisées. Les taux (VP, FP, FN, VN) sont calculés périodiquement dans chaque 10 images par scénario.

Cette opération a été réalisée pour les autres scénarios, la figure (Fig.5.12) illustre les courbes ROC et la figure (Fig.5.13) illustre les courbes Rappel-Précision.

Précision : la précision indique le nombre d'objets ré-identifiés correctement par notre algorithme sans la provocation des faux positifs.

Rappel : Le rappel indique la possibilité des réponses correctes de la ré-identification inter-caméra. L'utilisation de la courbe ROC et la courbe rappel-précision nous permettent de comprendre la performance de notre algorithme pour la ré-identification distribuée d'objet inter-caméra, par exemple : l'augmentation des fausses ré-identifications d'objets influence explicitement sur la performance de notre algorithme. Le taux de la ré-identification est égal au taux des vrais positifs moyens des quatre scénarios illustrés dans les courbes ROC (Fig.5.12) :

$$TRM = (TVP_1 + TVP_2 + TVP_3 + TVP_4)/4 \quad (5.10)$$

Le pourcentage de ré-identification inter-caméra calculé à partir des quatre courbes *ROC* en utilisant l'équation.5.10 est de 90,11%. A partir des courbes *ROC* illustrées dans la figure (Fig.5.12), nous remarquons que : les taux des vrais positifs (ré-identifications correctes) dans les quatre scénarios utilisés dans notre expérimentation sont plus important que les taux des faux positifs. Les taux des vrais positifs prouvent la validité de notre algorithme pour la re-identification d'objets d'une manière totalement distribuée inter-caméra et en présence des conditions complexes. Les courbes Rappel-Précision illustrées dans la figure(Fig.5.13), illustrent aussi un taux de précisions important presque pour tous les niveaux du rappel, ce qui prouve encore la puissance et la validité de l'algorithme proposé. Ces expérimentations prouvent la robustesse de l'algorithme proposé pour la ré-identification d'objets mobiles inter-caméra en présence des conditions complexes.

Malgré la capacité remarquable de notre algorithme pour la ré-identification d'objets inter-caméra, notre algorithme souffre de quelques problèmes, surtout quand les signatures visuelles de deux ou plusieurs objets d'intérêts sont les mêmes comme dans les centres commerciaux où les travailleurs portent des vêtements identiques. Nous pourrions résoudre ce problème à travers l'insertion des signaux biométriques des visages dans le modèle de représentation

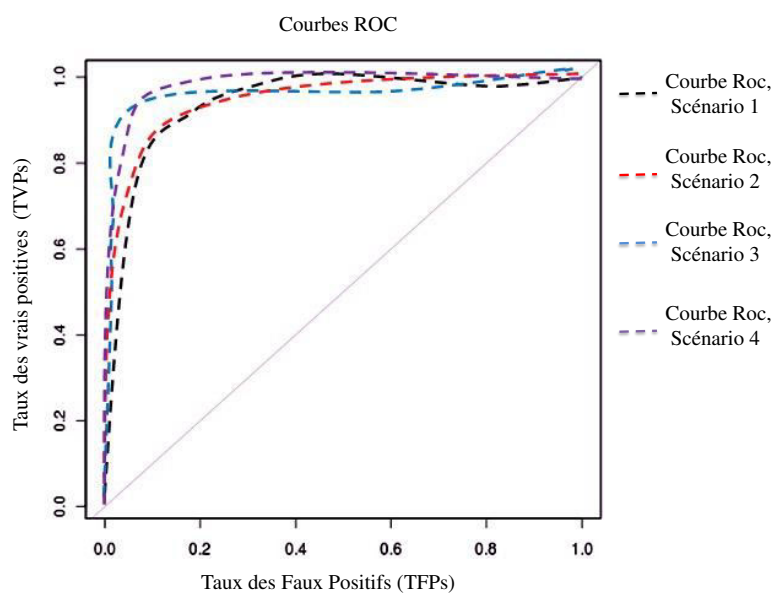


Fig. 5.12: Ré-identification d'objets inter-caméra, les Courbes *ROC* entre les *TFP* et les *TVP* des 4 scénarios.

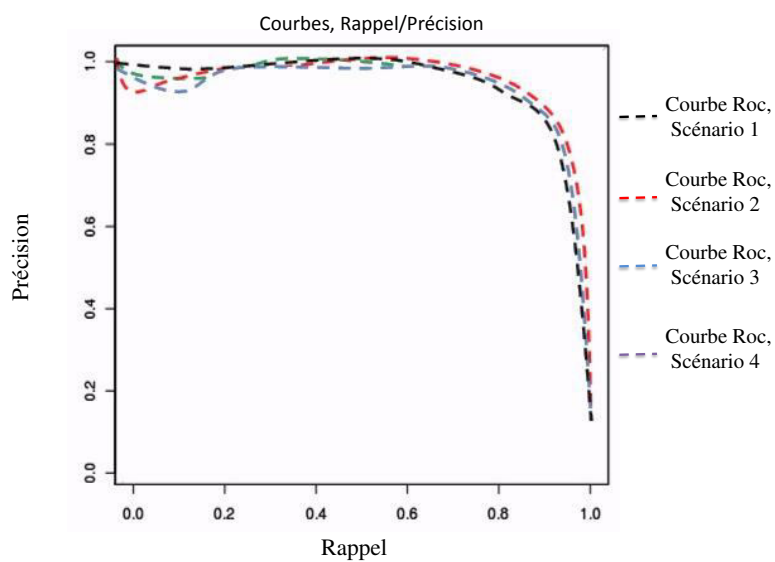


Fig. 5.13: Ré-identification d'objets inter-caméra, les Courbes Rappel-Précision des 4 scénarios.



Fig. 5.14: Un Faux positif (fausse alerte) généré par le module de re-identification du système proposé d'une durée de 2 Frames.

des cibles proposé (Module.3). Le mouvement des caméras peut aussi engendrer des fausses alertes (faux positifs) dans l'opération de la re-identification. Exemple d'un faux positif généré pendant l'opération de re-identification causé par le mouvement de la caméra est illustré dans la figure (Fig.5.14). La méthode utilisée dans le cadre de cette thèse pour calculer le taux des faux positifs est brièvement détaillée dans la section.5.4. Le développement d'un détecteur de mouvement robuste aux mouvements des caméras est l'un des perspectives de nos travaux de recherche.

5.5 Conclusion

Dans ce chapitre, nous avons préalablement proposé une architecture totalement distribuée basée sur le paradigme multi-agent pour le suivi et la ré-identification d'objets inter-caméra. Un modèle générique d'agent cognitif a été aussi proposé où une instance de ce modèle est utilisée par caméra. En deuxième temps et pour assurer une collaboration entre les caméras intelligentes, nous avons proposé un algorithme qui incorpore un ensemble d'instructions. Ces instructions permettant aux caméras intelligentes de collaborer mutuellement afin de parvenir à un consensus sur les identités des cibles traversant notre réseau multi-caméra à champs disjoints sans fil. Les expérimentations réalisées prouvent la robustesse en temps réel et la validité de notre algorithme pour la ré-identification d'objets mobiles inter-caméra. La ré-identification distribuée et les résultats spectaculaires de notre algorithme,

offre plusieurs opportunités aux chercheurs travaillant dans ce domaine pour explorer l'inférence distribuée afin de monter des réseaux intelligents capables d'analyser l'information visuelle ou multimodale répartie en temps réel.

Conclusion et perspectives

DANS cette thèse, notre intérêt s'est porté sur le suivi et la ré-identification distribuée d'objets inter-caméra dans un réseau multi-caméra à champs disjoints sans fil. La mise en place des techniques rebustes pour le suivi et la ré-identification d'objets mobiles dans ce genre de réseau multi-caméra est un grand défi. Les méthodes utilisées jusqu'à présent pour le suivi et la détection d'objets mobiles doivent être améliorées. La première partie de la thèse a été réservée à l'amélioration des méthodes proposées dans la littérature pour les systèmes monoculaire, afin de les adapter avec le contexte multi-caméra. Ensuite, un système totalement décentralisé basé sur l'inférence réparti a été proposé pour la ré-identification distribuée d'objets mobiles inter-caméra. Dans cette conclusion générale, nous faisons tout d'abord un bilan de ces travaux puis proposons un certain nombre de perspectives.

6.1 Synthèse des travaux effectués

Nous avons tout d'abord proposé une nouvelle chaîne de traitement pour la détection d'objets en mouvement. Cette chaîne repose sur plusieurs techniques couramment utilisées en vision par ordinateur. Elle se base sur quatre étapes principales, contenant chacune différentes nouveautés.

La première étape consiste à définir un prétraitement adapté aux types de données utilisées, suivi par une modélisation statistique de l'arrière plan basée sur une mixture gaussienne. En suite, une carte de soustraction binarisée est générée après une opération de soustraction arrière-plan/avant-plan basée sur une soustraction absolue. Les opérations de remplissage des vides dans chaque composante connexe (blob) et l'élimination des bruits en arrière-plan sont assurées par le post-traitement utilisé. Le post-traitement se base sur une convolution de la carte de soustraction érodée par un noyau gaussien. Enfin, pour la labellisation des composantes connexes, une génération d'une carte de soustraction binarisée et étiquetée est assurée par l'algorithme union-find. Les résultats obtenus prouvent la robustesse de notre chaîne de traitement.

La deuxième partie du document a traité du problème de la représentation invariante et discriminante des cibles inter et intra-caméra. Nous avons proposé ensuite un nouveau modèle de représentation de l'apparence des cibles. Le

modèle d'apparence proposé est invariant inter-caméra et discriminant inter-objet. Cet modèle se base sur une mixture ad-hoc des indices de textures de Haralick et des moments statistiques. Les résultats d'évaluation du modèle proposé dans des différentes conditions de captures inter/intra-caméra, prouvent la capacité représentative du modèle d'apparence proposé.

La troisième partie du document a été consacrée à l'amélioration de l'algorithme de suivi par le filtre particulaire afin de l'adapter avec le contexte multi-caméra. Ainsi, une nouvelle version de la méthode filtre particulaire a été proposée nommée «Filtre particulaire contextuel adaptatif a état mixte». Le filtre proposé se décompose en trois étapes essentielles : (1) Initialisation automatique et l'échantillonnage des particules à travers la carte de soustraction binarisée et étiquetée. (2) Estimation de l'importance des particules en utilisant un modèle de vraisemblance visuel basé sur le modèle de l'apparence proposé dans la deuxième partie du document. (3) Estimation de la distribution de probabilité à postériori moyenne, sélection des particules et l'adaptation de l'espace de recherche de chaque particule avec la dimension réelle de la cibles d'intérêt. Enfin, pour assurer le suivi multi-objet en présence de plusieurs cibles, nous avons proposé l'utilisation de plusieurs instances du filtre adaptatif proposé où une instance par cible. Plusieurs expérimentations réalisées prouvent la robustesse du filtre particulaire proposé et son capacité de convergence en évitant les minimums locaux (positions faux positives).

La dernière partie du document a traité du problème de la ré-identification d'objets mobiles inter-caméra dans un réseau multi-caméra à champs disjoints sans fil. Dans cette partie, nous avons proposé une approche totalement distribuée pour la ré-identification d'objets inter-caméra. Cette approche se repose sur la collaboration entre les caméras intelligentes afin de parvenir à un consensus sur les identités des cibles. Les expérimentations réalisées prouvent la robustesse de l'algorithme proposé pour la ré-identification d'objets mobiles inter-caméra.

6.2 Perspectives

Un certain nombre de perspectives peuvent être issues de cette thèse. On peut distinguer des perspectives plus méthodologiques et des perspectives plus applicatives.

1. Détection d'objets en mouvement (mouvement de la caméra)

La chaîne de traitement proposée dans cette thèse pour la détection d'objets en mouvement reste robuste que dans le cas des caméras statiques. Pour adapter la chaîne de traitement proposée au mouvement de la caméra, il faut développer un nouvel estimateur des arrière-plans afin d'assurer une bonne

classification avant/arrière-plan. Ce travail est l'un des axes principaux de nos travaux futurs.

2. Amélioration de l'algorithme de suivi

Jusqu'à présent le suivi multi/mon-objet basé sur les observations visuelles reste un problème au cœur de la recherche, où les algorithmes proposés dans la littérature restent robustes que dans des conditions spécifiques d'applications. La deuxième perspective de cette thèse est le développement des algorithmes de suivi intelligents, adaptables au contexte global et local de la scène. Pour atteindre cet objectif, il faut améliorer les filtres récursifs en trois points essentiels : (1) Développer un modèle d'apparence visuelle d'objet adaptable au contexte de la scène. (2) L'utilisation de l'observation dans la l'étape de prédiction aussi afin de minimiser la divergence. (3) La proposition des mécanismes pour la détection d'occultations et la résolution du problème de l'association de données.

3. Description visuelle invariante inter-caméra

L'algorithme proposé dans cette thèse pour la ré-identification distribuée d'objets inter-caméra intègre une opération d'adaptation de la couleur inter-caméra. Cette opération minimise la variation inter-caméra dans le modèle d'apparence proposé. Cette calibration de couleur influence sur la scabilité de notre système, ce qui nécessite le développement des nouveaux descripteurs visuels invariants inter-caméra. La troisième perspective de cette thèse est le développement d'un modèle d'apparence invariant inter-caméra en utilisant une combinaison de plusieurs modalités de captures, afin d'assurer une grande scabilité de notre système.

4. Résolution d'ambiguïtés dans la décision globale (Consensus)

L'un des grands problèmes dans le suivi et la ré-identification inter-caméra est la ressemblance visuelle entre des objets différents, par exemple : deux objets différents portant les vêtements avec la même couleur, dans ce cas la décision globale à-propos des identités d'objets devienne très compliquée. La présence de telle ambiguïté dans la décision globale nécessite l'utilisation des mécanismes spécifiques pour la résoudre, par exemple : l'utilisation des contraintes spatiales et temporelles, attribution des degrés de confiances à chaque caméra intelligente, etc. La quatrième perspective de cette thèse est l'intégration d'un mécanisme de résolution d'ambiguïté dans l'algorithme proposé, afin d'améliorer l'opération de la re-identification inter-caméra.

Bibliographie

- [Agarwal 2006] Ankur Agarwal. *Machine learning for image based motion capture*. PhD thesis, Institut National Polytechnique de Grenoble-INPG, 2006.
- [Andreas 2008] Griesser Andreas. *OSCAR - an Opportunistic SCAnner @ONLINE*, 2008.
- [Arulampalam 2004] M. Arulampalam, S. Maskell, N. Gordon et T. Clapp. *A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking*. Signal Processing, IEEE Transactions on, vol. 50, no. 2, pages 174–188, 2004.
- [Bak 2010] Sławomir Bak, Etienne Corvee, Francois Brémont et Monique Thonnat. *Person re-identification using haar-based and dcd-based signature*. In Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on, pages 1–8. IEEE, 2010.
- [Baker 2011] Simon Baker, Daniel Scharstein, JP Lewis, Stefan Roth, Michael J Black et Richard Szeliski. *A database and evaluation methodology for optical flow*. International Journal of Computer Vision, vol. 92, no. 1, pages 1–31, 2011.
- [Bay 2006] Herbert Bay, Tinne Tuytelaars et Luc Van Gool. *Surf : Speeded up robust features*. Computer Vision–ECCV 2006, pages 404–417, 2006.
- [Bekkari 2011] A Bekkari, S Idbraim, D Mammass et ME Yassa. *Exploiting spectral and space information in classification of high resolution urban satellites images using Haralick features and SVM*. In Multimedia Computing and Systems (ICMCS), 2011 International Conference on, pages 1–4. IEEE, 2011.
- [Bellotto 2007] Nicola Bellotto et Huosheng Hu. *People tracking with a mobile robot : A comparison of Kalman and particle filters*. In Proc. of the 13th IASTED Int. Conf. on Robotics and Applications, pages 388–393. Citeseer, 2007.
- [Benezeth 2008] Yannick Benezeth, Pierre-Marc Jodoin, Bruno Emile, Hélène Laurent et Christophe Rosenberger. *Review and evaluation of commonly-implemented background subtraction algorithms*. In Pattern Recognition, 2008. ICPR 2008. 19th International Conference on, pages 1–4. IEEE, 2008.

- [Benfold 2011] Ben Benfold et Ian Reid. *Stable multi-target tracking in real-time surveillance video*. In Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, pages 3457–3464. IEEE, 2011.
- [Berg 2005] Alexander Christiansen Berg, Tamara L Berg et Jitendra Malik. *Shape matching and object recognition using low distortion correspondences*. In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, volume 1, pages 26–33. IEEE, 2005.
- [Bob 2002] Bailey Bob. *Moments in image processing@ONLINE*, Novembre 2002.
- [Bousetouane 2010] Fouad Bousetouane. *Méta-apprentissage pour la reconnaissance des structures cérébrales : application au irm-3d*. Master's thesis, University of Badji Mokhtar-Annaba, 2010.
- [Bousetouane 2011] Fouad Bousetouane, Lynda Dib et Hichem Snoussi. *Robust detection and tracking pedestrian object for real time surveillance applications*. In Proc. SPIE, volume 8285, page 828508, 2011.
- [Bousetouane 2012] Fouad Bousetouane et Dib Lynda. *Robust Particle Filtering with Multiple-cues for Non-rigid Object Tracking*. In Artificial Intelligence Doctorials (AID), 2012, pages 81–90. USTHB-Algiers, 2012.
- [Bousetouane 2013a] Fouad Bousetouane, Motamed Cina et Bib Lynda. *Contextual Adaptive Particle Filtering for Robust Real-time Non-Rigid Object Tracking*. In 8th International Workshop on Systems, Signal Processing and their Applications (WoSSPA), pages 97–102. IEEE, 2013.
- [Bousetouane 2013b] Fouad Bousetouane, Lynda Dib et Hichem Snoussi. *Improved mean shift integrating texture and color features for robust real time object tracking*. The Visual Computer, vol. 29, pages 155–170, 2013.
- [Bousetouane 2013c] Fouad Bousetouane, Cina Motamed et Lynda Dib. *Consensus-based Inter-camera Re-identification Across Non-overlapping Views*. In 8th International Conference on Computer Vision Theory and Applications (VISAPP), pages 202–211. INSTICC, 2013.
- [Bouwman 2008] Thierry Bouwman, Fida El Baf et Bertrand Vachon. *Background modeling using mixture of gaussians for foreground detection-a survey*. Recent Patents on Computer Science, vol. 1, no. 3, pages 219–237, 2008.

- [Bruhn 2005] A. Bruhn, J. Weickert, C. Feddern, T. Kohlberger et C. Schnorr. *Variational optical flow computation in real time*. Image Processing, IEEE Transactions on, vol. 14, no. 5, pages 608–615, 2005.
- [Cai 1999] Qin Cai et Jake K Aggarwal. *Tracking human motion in structured environments using a distributed-camera system*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 21, no. 11, pages 1241–1247, 1999.
- [CAVIAR 2004] CAVIAR. *CAVIAR Test Case Scenarios @ONLINE*, 2004.
- [Chan 2004] Yung-Kuan Chan et Chih-Ya Chen. *Image retrieval system based on color-complexity and color-spatial features*. Journal of Systems and Software, vol. 71, no. 1, pages 65–70, 2004.
- [Chen 2008] Kuan-Wen Chen, Chih-Chuan Lai, Yi-Ping Hung et Chu-Song Chen. *An adaptive learning method for target tracking across multiple cameras*. In Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, pages 1–8. IEEE, 2008.
- [Chen 2011] Kuan-Wen Chen, Chih-Chuan Lai, Pei-Jyun Lee, Chu-Song Chen et Yi-Ping Hung. *Adaptive Learning for Target Tracking and True Linking Discovering across Multiple Non-Overlapping Cameras*. Multimedia, IEEE Transactions on, vol. 13, no. 4, pages 625–638, 2011.
- [Chitrakala 2009] S Chitrakala, P Shamini et D Manjula. *Multi-class enhanced image mining of heterogeneous textual images using multiple image features*. In Advance Computing Conference, 2009. IACC 2009. IEEE International, pages 496–501. IEEE, 2009.
- [Collins 2001] Robert T Collins, Alan J Lipton, Hironobu Fujiyoshi et Takeo Kanade. *Algorithms for cooperative multisensor surveillance*. Proceedings of the IEEE, vol. 89, no. 10, pages 1456–1477, 2001.
- [Colombo 2008] Alberto Colombo, James Orwell, Sergio Velastin *et al.* *Colour constancy techniques for re-recognition of pedestrians from multiple surveillance cameras*. In Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications-M2SFA2 2008, 2008.
- [Comaniciu 2000] Dorin Comaniciu, Visvanathan Ramesh et Peter Meer. *Real-time tracking of non-rigid objects using mean shift*. In Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on, volume 2, pages 142–149. IEEE, 2000.
- [Cucchiara 2003] Rita Cucchiara, Costantino Grana, Massimo Piccardi et Andrea Prati. *Detecting moving objects, ghosts, and shadows in video streams*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 25, no. 10, pages 1337–1342, 2003.

- [Deng 2007] Hongli Deng, Wei Zhang, Eric Mortensen, Thomas Dietterich et Linda Shapiro. *Principal curvature-based region detector for object recognition*. In Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on, pages 1–8. IEEE, 2007.
- [Dick 2005] Anthony Dick et Michael Brooks. *A stochastic approach to tracking objects across multiple cameras*. AI 2004 : Advances in Artificial Intelligence, pages 53–84, 2005.
- [Doretto 2011] Gianfranco Doretto, Thomas Sebastian, Peter Tu et Jens Rittscher. *Appearance-based person reidentification in camera networks : problem overview and current approaches*. Journal of Ambient Intelligence and Humanized Computing, vol. 2, no. 2, pages 127–151, 2011.
- [Erdem 2010] Erkut Erdem, Séverine Dubuisson et Isabelle Bloch. *Particle Filter-Based Visual Tracking by Fusing Multiple Cues with Context-Sensitive Reliabilities Suivi d'objet dans des séquences d'images par filtrage particulaire et fusion d'informations, prenant en compte leur fiabilité*. 2010.
- [Farenzena 2010] Michela Farenzena, Loris Bazzani, Alessandro Perina, Vittorio Murino et Marco Cristani. *Person re-identification by symmetry-driven accumulation of local features*. In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, pages 2360–2367. IEEE, 2010.
- [Fazli 2009] Saeid Fazli, Hamed Moradi Pour et Hamed Bouzari. *Particle filter based object tracking with sift and color feature*. In Machine Vision, 2009. ICMV'09. Second International Conference on, pages 89–93. IEEE, 2009.
- [François 1999] Alexandre RJ François et Gérard G Medioni. *Adaptive color background modeling for real-time segmentation of video streams*. In Proceedings of the International Conference on Imaging Science, Systems, and Technology, pages 227–232, 1999.
- [García-Martín 2010] Álvaro García-Martín et José M Martínez. *Robust real time moving people detection in surveillance scenarios*. In Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on, pages 241–247. IEEE, 2010.
- [Gilbert 2006] Andrew Gilbert et Richard Bowden. *Tracking objects across cameras by incrementally learning inter-camera colour calibration and patterns of activity*. Computer Vision–ECCV 2006, pages 125–136, 2006.
- [Hamdoun 2008] Omar Hamdoun, Fabien Moutarde, Bogdan Stanciulescu et Bruno Steux. *Person re-identification in multi-camera system by si-*

- gnature based on interest point descriptors collected on short video sequences.* In Distributed Smart Cameras, 2008. ICDS-C 2008. Second ACM/IEEE International Conference on, pages 1–6. IEEE, 2008.
- [Hanbury 2003] Allan Hanbury et Jean Serra. *Colour image analysis in 3d-polar coordinates.* Pattern Recognition, pages 124–131, 2003.
- [Haralick 1973] Robert M Haralick, Karthikeyan Shanmugam et Its' Hak Dinstein. *Textural features for image classification.* Systems, Man and Cybernetics, IEEE Transactions on, no. 6, pages 610–621, 1973.
- [Haritaoglu 1998] Ismail Haritaoglu, David Harwood et Larry Davis. *W 4 S : A real-time system for detecting and tracking people in 2 1/2D.* Computer Vision—ECCV'98, pages 877–892, 1998.
- [Hero III 2002] Alfred O Hero III, Bing Ma, Olivier JJ Michel et John Gorman. *Applications of entropic spanning graphs.* Signal Processing Magazine, IEEE, vol. 19, no. 5, pages 85–95, 2002.
- [Hogg 1983] David Hogg. *Model-based vision : a program to see a walking person.* Image and vision computing, vol. 1, no. 1, pages 5–20, 1983.
- [Horprasert 1999] Thanarat Horprasert, David Harwood et Larry S Davis. *A statistical approach for real-time robust background subtraction and shadow detection.* In IEEE ICCV, volume 99, pages 256–261, 1999.
- [Hu 2004] Weiming Hu, Tieniu Tan, Liang Wang et Steve Maybank. *A survey on visual surveillance of object motion and behaviors.* Systems, Man, and Cybernetics, Part C : Applications and Reviews, IEEE Transactions on, vol. 34, no. 3, pages 334–352, 2004.
- [Hu 2010] Jiangping Hu et Gang Feng. *Distributed tracking control of leader-follower multi-agent systems under noisy measurement.* Automatica, vol. 46, no. 8, pages 1382–1387, 2010.
- [Huang 1997] Timothy Huang et Stuart Russell. *Object identification in a bayesian context.* In International Joint Conference on Artificial Intelligence, volume 15, pages 1276–1283. Citeseer, 1997.
- [Iftene 2004] Tahar Iftene et Abdelmounaine Safia. *Comparaison entre la matrice de cooccurrence et la transformation en ondelettes pour la classification texturale des images HRV (XS) de Spot.* Télédétection, vol. 4, no. 1, pages 39–49, 2004.
- [Ilyas 2009] Atif Ilyas, Mihaela Scuturici et Serge Miguet. *Real time foreground-background segmentation using a modified codebook model.* In Advanced Video and Signal Based Surveillance, 2009. AVSS'09. Sixth IEEE International Conference on, pages 454–459. IEEE, 2009.

- [Isard 1998] Michael Isard et Andrew Blake. *Condensation—conditional density propagation for visual tracking*. International journal of computer vision, vol. 29, no. 1, pages 5–28, 1998.
- [Jaffré 2004] Gaël Jaffré, Philippe Joly et al. *Costume : A new feature for automatic video content indexing*. In Proceedings of RIAO, pages 314–325, 2004.
- [Jain 1995] Ramesh Jain et Koji Wakimoto. *Multiple perspective interactive video*. In Multimedia Computing and Systems, 1995., Proceedings of the International Conference on, pages 202–211. IEEE, 1995.
- [Javed 2005] Omar Javed, Khurram Shafique et Mubarak Shah. *Appearance modeling for tracking in multiple non-overlapping cameras*. In Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on, volume 2, pages 26–33. IEEE, 2005.
- [Javed 2008a] Omar Javed, Khurram Shafique, Zeeshan Rasheed et Mubarak Shah. *Modeling inter-camera space-time and appearance relationships for tracking across non-overlapping views*. Computer Vision and Image Understanding, vol. 109, no. 2, pages 146–162, 2008.
- [Javed 2008b] Omar Javed et Mubarak Shah. *Tracking in multiple cameras with disjoint views*. Automated Multi-Camera Surveillance : Algorithms and Practice, pages 1–26, 2008.
- [Ju 2010] Ming-Yi Ju, Chen-Sen Ouyang et Hao-Shiu Chang. *Mean shift tracking using fuzzy color histogram*. In Machine Learning and Cybernetics (ICMLC), 2010 International Conference on, volume 6, pages 2904–2908. IEEE, 2010.
- [Julien 2007] Osmalsky Julien. *Caractérisation d’un air de musique et reconnaissance avec un son dans une base de données*. Master’s thesis, Université de Liège, 2007.
- [Kettner 1999] V. Kettner et R. Zabih. *Bayesian multi-camera surveillance*. In Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on., volume 2, pages –259 Vol. 2, 1999.
- [Khan 2005] Zia Khan, Tucker Balch et Frank Dellaert. *MCMC-based particle filtering for tracking a variable number of interacting targets*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 27, no. 11, pages 1805–1819, 2005.
- [Khan 2011] Zulfiqar Hasan Khan, Irene Yu-Hua Gu et Andrew G Backhouse. *A Robust Particle Filter-Based Method for Tracking Single Visual Object Through Complex Scenes Using Dynamical Object Shape and Appearance Similarity*. Journal of Signal Processing Systems, vol. 65, no. 1, pages 63–79, 2011.

- [Kim 2005] Kyungnam Kim, Thanarat H Chalidabhongse, David Harwood et Larry Davis. *Real-time foreground-background segmentation using codebook model*. Real-time imaging, vol. 11, no. 3, pages 172–185, 2005.
- [Koller 1994] Dieter Koller, J Weber, T Huang, J Malik, G Ogasawara, B Rao et S Russell. *Towards robust automatic traffic scene analysis in real-time*. In Pattern Recognition, 1994. Vol. 1-Conference A : Computer Vision & Image Processing., Proceedings of the 12th IAPR International Conference on, volume 1, pages 126–131. IEEE, 1994.
- [Laptev 2006] Ivan Laptev et Tony Lindeberg. *Tracking of Multi-state Hand Models Using Particle Filtering and a Hierarchy of Multi-scale Image Features*. Scale-Space and Morphology in Computer Vision, pages 63–74, 2006.
- [Leibe 2007] Bastian Leibe, Konrad Schindler et Luc Van Gool. *Coupled detection and trajectory estimation for multi-object tracking*. In Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on, pages 1–8. IEEE, 2007.
- [Leichter 2010] Ido Leichter, Michael Lindenbaum et Ehud Rivlin. *Mean shift tracking with multiple reference color histograms*. Computer Vision and Image Understanding, vol. 114, no. 3, pages 400–408, 2010.
- [Li 2011] Yiming Li et Bir Bhanu. *Utility-based camera assignment in a video network : A game theoretic framework*. Sensors Journal, IEEE, vol. 11, no. 3, pages 676–687, 2011.
- [Liu 2005] Tianjian Liu et Shanan Zhu. *Eyes detection and tracking based on entropy in particle filter*. In Control and Automation, 2005. ICCA'05. International Conference on, volume 2, pages 1002–1007. IEEE, 2005.
- [Lo Presti 2009] Liliana Lo Presti, Stan Sclaroff et Marco La Cascia. *Object matching in distributed video surveillance systems by LDA-based appearance descriptors*. Image Analysis and Processing-ICIAP 2009, pages 547–557, 2009.
- [Loza 2007] A Loza, Alin Achim, David Bull et Nishan Canagarajah. *Statistical image fusion with generalised Gaussian and Alpha-Stable distributions*. In Digital Signal Processing, 2007 15th International Conference on, pages 268–271. IEEE, 2007.
- [Madden 2007] Christopher Madden, Eric Dahai Cheng et Massimo Piccardi. *Tracking people across disjoint camera views by an illumination-tolerant appearance representation*. Machine Vision and Applications, vol. 18, no. 3, pages 233–247, 2007.
- [Makris 2004] Dimitrios Makris, Tim Ellis et James Black. *Bridging the gaps between cameras*. In Computer Vision and Pattern Recognition, 2004.

- CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, volume 2, pages II–205. IEEE, 2004.
- [Mao 2011] Kezhi Z Mao et Wenyin Tang. *Recursive Mahalanobis separability measure for gene subset selection*. Computational Biology and Bioinformatics, IEEE/ACM Transactions on, vol. 8, no. 1, pages 266–272, 2011.
- [Meden 2011] Boris Meden, Patrick Sayd et Frédéric Lerasle. *Mixed-state particle filtering for simultaneous tracking and re-identification in non-overlapping camera networks*. Image Analysis, pages 124–133, 2011.
- [Migdal 2005] Joshua Migdal et W Eric L Grimson. *Background subtraction using markov thresholds*. In Application of Computer Vision, 2005. WACV/MOTIONS'05 Volume 1. Seventh IEEE Workshops on, volume 2, pages 58–65. IEEE, 2005.
- [Mohamed 2010] Shahrizat Shaik Mohamed, Nooritawati Md Tahir et Ramli Adnan. *Background modelling and background subtraction performance for object detection*. In Signal Processing and Its Applications (CSPA), 2010 6th International Colloquium on, pages 1–6. IEEE, 2010.
- [Motamed 2007] Cina Motamed et Olivier Wallart. *A temporal fusion strategy for cross-camera data association*. Pattern recognition letters, vol. 28, no. 2, pages 233–245, 2007.
- [Musicki 2004] Darko Musicki et Robin Evans. *Joint integrated probabilistic data association : JIPDA*. Aerospace and Electronic Systems, IEEE Transactions on, vol. 40, no. 3, pages 1093–1099, 2004.
- [Muybridge 1979] Eadweard Muybridge. *Muybridge's complete human and animal locomotion : all 781 plates from the 1887 animal locomotion*, volume 3. Dover Publications, 1979.
- [Nakajima 2003] Chikahito Nakajima, Massimiliano Pontil, Bernd Heisele et Tomaso Poggio. *Full-body person recognition system*. Pattern recognition, vol. 36, no. 9, pages 1997–2006, 2003.
- [Nummiaro 2002] Katja Nummiaro, Esther Koller-Meier et Luc Van Gool. *Object tracking with an adaptive color-based particle filter*. Pattern Recognition, pages 353–360, 2002.
- [Olfati-Saber 2007] Reza Olfati-Saber, J Alex Fax et Richard M Murray. *Consensus and cooperation in networked multi-agent systems*. Proceedings of the IEEE, vol. 95, no. 1, pages 215–233, 2007.
- [Oliva 2007] Aude Oliva, Antonio Torralba et al. *The role of context in object recognition*. Trends in cognitive sciences, vol. 11, no. 12, pages 520–527, 2007.

- [Pantrigo 2005] Juan Pantrigo, Antonio Montemayor et Raúl Cabido. *Scatter search particle filter for 2d real-time hands and face tracking*. Image Analysis and Processing—ICIAP 2005, pages 953–960, 2005.
- [Pasula 1999] Hanna Pasula, Stuart Russell, Michael Ostland et Yaacov Rivtsov. *Tracking many objects with many sensors*. In International Joint Conference on Artificial Intelligence, volume 16, pages 1160–1171. LAWRENCE ERLBAUM ASSOCIATES LTD, 1999.
- [Pérez 2002] Patrick Pérez, Carine Hue, Jaco Vermaak et Michel Gangnet. *Color-based probabilistic tracking*. Computer Vision—ECCV 2002, pages 661–675, 2002.
- [Pets 2000] Pets. *Motion Tracking : PETS 2000 sequence @ONLINE*, 2000.
- [Pets 2009] Pets. *Motion Tracking : PETS 2009 sequence @ONLINE*, 2009.
- [Piccardi 2004] Massimo Piccardi. *Background subtraction techniques : a review*. In Systems, Man and Cybernetics, 2004 IEEE International Conference on, volume 4, pages 3099–3104. IEEE, 2004.
- [Ponsa 2005] Daniel Ponsa, Antonio López, Joan Serrat, Felipe Lumbreras et Thorsten Graf. *Multiple vehicle 3d tracking using an unscented kalman*. In Intelligent Transportation Systems, 2005. Proceedings. 2005 IEEE, pages 1108–1113. IEEE, 2005.
- [Porikli 2003a] F. Porikli et A. Divakaran. *Multi-camera calibration, object tracking and query generation*. In Multimedia and Expo, 2003. ICME '03. Proceedings. 2003 International Conference on, volume 1, pages I–653–6 vol.1, 2003.
- [Porikli 2003b] Fatih Porikli. *Inter-camera color calibration by correlation model function*. In Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on, volume 2, pages II–133. IEEE, 2003.
- [Power 2002] P Wayne Power et Johann A Schoonees. *Understanding background mixture models for foreground segmentation*. In Proceedings Image and Vision Computing New Zealand, volume 2002, 2002.
- [Prosser 2008] B. Prosser, S. Gong et T. Xiang. *Multi-camera Matching using Bi-Directional Cumulative Brightness Transfer Functions*. In Proceedings of the British Machine Vision Conference, pages 64.1–64.10. BMVA Press, 2008. doi :10.5244/C.22.64.
- [Qi 2010] Guo-Jun Qi, Xian-Sheng Hua, Yong Rui, Jinhui Tang et Hong-Jiang Zhang. *Image classification with Kernelized spatial-context*. Multimedia, IEEE Transactions on, vol. 12, no. 4, pages 278–287, 2010.
- [Qu 2007] Wei Qu, Dan Schonfeld et Magdi Mohamed. *Real-time distributed multi-object tracking using multiple interactive trackers and a*

- magnetic-inertia potential model*. Multimedia, IEEE Transactions on, vol. 9, no. 3, pages 511–519, 2007.
- [Quaritsch 2007] Markus Quaritsch, Markus Kreuzthaler, Bernhard Rinner, Horst Bischof et Bernhard Strobl. *Autonomous multicamera tracking on embedded smart cameras*. EURASIP Journal on Embedded Systems, vol. 2007, no. 1, pages 35–35, 2007.
- [Radke 2005] Richard J Radke, Srinivas Andra, Omar Al-Kofahi et Badrinath Roysam. *Image change detection algorithms : a systematic survey*. Image Processing, IEEE Transactions on, vol. 14, no. 3, pages 294–307, 2005.
- [Rahimi 2004] A. Rahimi, B. Dunagan et T. Darrell. *Simultaneous calibration and tracking with a network of non-overlapping sensors*. In Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on, volume 1, pages I–187–I–194 Vol.1, 2004.
- [Rahman 2007] Md M Rahman, Prabir Bhattacharya et Bipin C Desai. *A framework for medical image retrieval using machine learning and statistical similarity matching techniques with relevance feedback*. Information Technology in Biomedicine, IEEE Transactions on, vol. 11, no. 1, pages 58–69, 2007.
- [Rittscher 2000] Jens Rittscher, Jien Kato, Sébastien Joga et Andrew Blake. *A probabilistic background model for tracking*. Computer Vision—ECCV 2000, pages 336–350, 2000.
- [Runnalls 2007] Andrew R Runnalls. *Kullback-Leibler approach to Gaussian mixture reduction*. Aerospace and Electronic Systems, IEEE Transactions on, vol. 43, no. 3, pages 989–999, 2007.
- [Schlogl 2004] T Schlogl, C Beleznai, M Winter et Horst Bischof. *Performance evaluation metrics for motion detection and tracking*. In Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on, volume 4, pages 519–522. IEEE, 2004.
- [Schwartz 2009] William Robson Schwartz et Larry S Davis. *Learning discriminative appearance-based models using partial least squares*. In Computer Graphics and Image Processing (SIBGRAPI), 2009 XXII Brazilian Symposium on, pages 322–329. IEEE, 2009.
- [Schwerdt 2000] Karl Schwerdt et James L Crowley. *Robust face tracking using color*. In Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on, pages 90–95. IEEE, 2000.

- [Seigneur 2004] Jean-Marc Seigneur, Daniel Solis et Fergal Shevlin. *Ambient intelligence through image retrieval*. Image and Video Retrieval, pages 1960–1960, 2004.
- [Serre 2007] Thomas Serre, Lior Wolf, Stanley Bileschi, Maximilian Riesenhuber et Tomaso Poggio. *Robust object recognition with cortex-like mechanisms*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 29, no. 3, pages 411–426, 2007.
- [Sherrah 2009] Jamie Sherrah, Branko Ristic et Nicholas J Redding. *Evaluation of a particle filter to track people for visual surveillance*. In Digital Image Computing : Techniques and Applications, 2009. DICTA'09., pages 96–102. IEEE, 2009.
- [Smith 2006] Duncan Smith et Sameer Singh. *Approaches to multisensor data fusion in target tracking : A survey*. Knowledge and Data Engineering, IEEE Transactions on, vol. 18, no. 12, pages 1696–1710, 2006.
- [Smith 2007] Kevin Smith. *Bayesian methods for visual multi-object tracking with applications to human activity recognition*. Lausanne, EPFL, 2007.
- [Soman 2010] Jyothish Soman, Kothapalli Kishore et P J Narayanan. *A fast GPU algorithm for graph connectivity*. In Parallel & Distributed Processing, Workshops and Phd Forum (IPDPSW), 2010 IEEE International Symposium on, pages 1–8. IEEE, 2010.
- [Soto 2009] Cristian Soto, Bi Song et Amit K Roy-Chowdhury. *Distributed multi-target tracking in a self-configuring camera network*. In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pages 1486–1493. IEEE, 2009.
- [Spanos 2005] Demetri P Spanos, Reza Olfati-Saber et Richard M Murray. *Approximate distributed Kalman filtering in sensor networks with quantifiable performance*. In Proceedings of the 4th international symposium on Information processing in sensor networks, page 18. IEEE Press, 2005.
- [SPEVI 2007] SPEVI. *Surveillance Performance Evaluation Initiative@ONLINE*, 2007.
- [Stauffer 2000] Chris Stauffer et W. Eric L. Grimson. *Learning patterns of activity using real-time tracking*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 22, no. 8, pages 747–757, 2000.
- [Stauffer 2005] Chris Stauffer. *Learning to Track Objects Through Unobserved Regions*. In Application of Computer Vision, 2005. WACV/MOTIONS '05 Volume 1. Seventh IEEE Workshops on, volume 2, pages 96–102, 2005.

- [Stenger 2001] Bjoern Stenger, Paulo R. S. Mendonça et Roberto Cipolla. *Model-based hand tracking using an unscented kalman filter*. In Proc. British Machine Vision Conference, volume 1, pages 63–72, 2001.
- [Takacs 2010] Gabriel Takacs, Vijay Chandrasekhar, Sam Tsai, David Chen, Radek Grzeszczuk et Bernd Girod. *Unified real-time tracking and recognition with rotation-invariant fast features*. In Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on, pages 934–941. IEEE, 2010.
- [Truong Cong 2009] Dung Truong Cong, Catherine Achard, Louahdi Khouddour et Lounis Douadi. *Video sequences association for people re-identification across multiple non-overlapping cameras*. Image Analysis and Processing–ICIAP 2009, pages 179–189, 2009.
- [Verbeke 2007] Nicolas Verbeke. *Suivi d’objets en mouvement dans une séquence vidéo*. PhD thesis, Université Paris Descartes, 2007.
- [Wang 2003] Liang Wang, Weiming Hu et Tieniu Tan. *Recent developments in human motion analysis*. Pattern recognition, vol. 36, no. 3, pages 585–601, 2003.
- [Weng 2006] Shiuh-Ku Weng, Chung-Ming Kuo et Shu-Kang Tu. *Video object tracking using adaptive Kalman filter*. Journal of Visual Communication and Image Representation, vol. 17, no. 6, pages 1190–1208, 2006.
- [Wren 1997] Christopher Richard Wren, Ali Azarbayejani, Trevor Darrell et Alex Paul Pentland. *Pfinder : Real-time tracking of the human body*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 19, no. 7, pages 780–785, 1997.
- [Wu 2007] Bo Wu, Xuefeng Song, Vivek Singh et Ram Nevatia. *Evaluation of USC human tracking system for surveillance videos*. Multimodal Technologies for Perception of Humans, pages 183–189, 2007.
- [Wu 2009] Ying Wu et Jialue Fan. *Contextual flow*. In Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, pages 33–40. IEEE, 2009.
- [Wu 2010] Hao Wu, Aswin C Sankaranarayanan et Rama Chellappa. *Online empirical evaluation of tracking algorithms*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 32, no. 8, pages 1443–1458, 2010.
- [Yilmaz 2006] Alper Yilmaz, Omar Javed et Mubarak Shah. *Object tracking : A survey*. Acm Computing Surveys (CSUR), vol. 38, no. 4, page 13, 2006.

- [Ying 2010] Hongwei Ying, Xuena Qiu, Jiatao Song et Xiaobo Ren. *Particle filtering object tracking based on texture and color*. In Intelligence Information Processing and Trusted Computing (IPTC), 2010 International Symposium on, pages 626–630. IEEE, 2010.
- [Zang 2006] Qi Zang et Reinhard Klette. *Parameter analysis for mixture of gaussians model*. 2006.
- [ZHANG 2010] Tao ZHANG et Shumin FEI. *TARGET TRACKING WITH PARTICLE FILTER BASED ON MULTIPLE CUES FUSION*. Journal of Systems Science and Mathematical Sciences, vol. 6, page 006, 2010.
- [Zheng 2009] Wei-Shi Zheng, Shaogang Gong et Tao Xiang. *Associating groups of people*. In British machine vision conference, volume 1, pages 1–11, 2009.
- [Živković 2004] Zoran Živković et Ferdinand van der Heijden. *Improving the selection of feature points for tracking*. Pattern Analysis & Applications, vol. 7, no. 2, pages 144–150, 2004.

Liste des publications

1. Fouad Bousetouane, Cina Motamed and Lynda Dib, " Contextual Adaptive Particle Filtering for Robust Real-time Non-Rigid Object Tracking", In proc of IEEE-WOSSPA'2013, Algiers, Algeria, May 2013. Pages 172-132
2. Fouad Bousetouane, Cina Motamed and Lynda Dib, "Consensus-based Inter-camera Re-identification - Across Non-overlapping Views", In proc of the 8th International Conference on Computer Vision Theory and Applications (VISAPP 2013) Barcelona, Spain. Pages 341-346
3. Fouad Bousetouane, Lynda Dib, "Robust Particle Filtering with Multiple-cues for Non-rigid Object Tracking", Proceedings of Artificial Intelligence Doctorials (AID2012), USTHB-Algiers, Algeria, 20-21 May 2012. Pages 81-90.
4. Fouad Bousetouane, Lynda Dib, Hichem Snoussi, "Improved Mean Shift Integrating Texture and Color Features For Robust Real Time Object Tracking. ", The Visual Computer Journal, Springer, 23.9, 2013, Pages 155-170. DOI : 10.1007/s00371-012-0677-0.

5. Fouad Bousetouane, Lynda Dib, Hichem Snoussi, "Robust Detection and Tracking Pedestrian Object for Real Time Surveillance Applications", Proc. SPIE 8285, 828508 (2011); DOI : 10.1117/12.913034.

Author Biography



Fouad Bousetouane, Received his BSc (bachelor degree) in computer science and mathematics, in 2008 and Master by Research degree in Artificial Intelligence and Pattern Recognition from Badji Mokhtar University, Algeria, in 2010. He obtained his PhD in Artificial Intelligence and Computer Vision from UBMA-University, Algeria, in 2014. He is valedictorian (Major de Promotion), member of International Association of Computer Science and Information Technology (IACSIT) and Computer Vision Foundation (CVF). He is listed in "Whos Who in the World : 2012". He collaborated with researchers from CNRS-Lille(France) and LISIC-Laboratory for developing computer vision algorithms for multi-object tracking and re-identification across multi-sensor networks. Actually, he is an associate teacher of computer science at EPSE (National preparatory school of business), Algeria. He is Co-Founder of Robotics and Intelligent Computing Startup. He authored many technical articles in machine learning, computer vision, satellite image processing, and served as a reviewer at IET-Image Processing journal and many high level conferences including(IEEE-IROS2012). His research interests include artificial intelligence, pattern recognition, Probabilistic Graphical models and bayesian computation, machine learning, computer vision and image processing.