

الجمهورية الجزائرية الديمقراطية الشعبية  
République Algérienne Démocratique et Populaire  
وزارة التعليم العالي والبحث العلمي  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique



**Université Kasdi Merbah Ouargla**

**Faculté des Nouvelles Technologies de L'Information  
et de la Communication**

**DEPARTEMENT D'ELECTRONIQUE ET TELECOMMUNICATIONS**



**Mémoire en vue de l'obtention du diplôme de  
Master Académique  
Domaine : Sciences et Technologies  
Filière : Génie Electrique  
Spécialité : Automatique**

Par : **BOUSSAFEUR Tiziri**

Encadrée par : **BENZAOUI Wafa**

**Thème**

---

**Fusion de Données Thermiques et Couleurs pour  
Le Suivi de Personnes**

---

**Devant le jury composé de :**

<b>Président :</b>	CHEBBARA Fouad	MCB	UKMO
<b>Examineur :</b>	SAIGAA Mohamed	MAA	UKMO
<b>Examineur :</b>	BENKRINAH Sabra	MAA	UKMO

**Année Universitaire : 2014/2015**

## REMERCIEMENTS

*Au terme de ce travail, je remercie le bon dieu ainsi que toutes les personnes qui ont participé de près ou de loin à l'élaboration de ce travail.*

*Je tiens à exprimer ma grande reconnaissance et ma profonde gratitude à l'égard de :*

*Ma promotrice Madame BENZAOUI Wafa qui m'a témoigné tant de compréhension et qui m'a encouragé et soutenue tout au long de la réalisation de ce modeste travail.*

*Mes remerciements vont également à tout le corps enseignant de département d'Electronique et télécommunication*

*Enfin il m'est très agréable d'exprimer mes meilleurs sentiments envers les membres de jury, qui ont bien voulu m'honorer de leurs présence et assister à la soutenance de ma mémoire afin d'évaluer mes efforts.*

*Tiziri*

## DEDICACES

*Je dédie ce travail à mes chers parents Mohamed et Yamina, qui ont toujours été présents et qui ont su éveiller pour l'accomplissement de mes souhaits, et dont je suis fière, autant qu'ils peuvent l'être de moi.*

*Je dédie ce modeste travail à l'âme de mon grand père «Hadj Slimane Boutaous » et ma grande mère*

*Yaya Hadjila*

*A mes chères sœurs Tinhinane et l'adorable Mélina ainsi que ma tante Linda sans oublier Malika*

*A mes oncles et tantes maternels et paternels qui ont été toujours à mes côtés.*

*A toute la famille Boussafeur, Boutaous et Mazri*

*A ne pas oublier « Mimene, Thafath, Lahna, Abd el Rezak, Zakaria et Faysal »*

*A toutes mes amis et collègues de la promotion  
2014/2015*

*A toute personne qui m'a aidé ou contribué à l'élaboration de ce travail.*

*Tiziri*

## TABLE DES MATIERES

### INTRODUCTION GÉNÉRALE

1. Introduction .....	1
2. Problématique .....	2
3. Objectifs .....	3
4. Plan de travail.....	5

### Chapitre I : Généralité sur la vidéosurveillance

1. Introduction .....	6
2. Système de vidéo surveillance .....	6
3. Générations des systèmes de vidéosurveillance.....	7
4. Structure d'un système de vidéosurveillance intelligente.....	8
4.1. La détection d'objet.....	9
4.2. Reconnaissance d'objets, le suivi et l'évaluation de la performance .....	9
4.3. Analyse comportementale .....	10
4.4. Stockage .....	10
5. Raisons pour l'utilisation de vidéosurveillance intelligente .....	10
6. Applications de la vidéosurveillance .....	11
6.1. Gouvernement et sécurité publique .....	11
6.2. Éducation.....	12
6.3. Commerce de détail .....	12
6.4. Transports .....	13
6.5. Milieu bancaire .....	13
7. Contraintes de la vidéosurveillance .....	13
8. Conclusion .....	15

### Chapitre II : Etat de l'Art

1. Introduction .....	16
2. Fusion d'image.....	16
2.1. Objectif de la fusion d'image .....	16
2.2. Applications de la fusion d'image.....	17
2.3. Niveaux de fusion.....	18
2.3.1. Fusion au niveau de pixel .....	18
2.3.2. Fusion au niveau des caractéristiques.....	19
2.3.3. Niveau decisionel .....	20

2.4. Méthodes de fusion.....	21
I. Algorithme arithmétique.....	21
2.4.1. Combinaison pondérée .....	21
2.4.2 Analyse en composantes principales (ACP).....	22
II. Algorithme biologique .....	22
2.4.3. Approches pyramidales .....	22
2.4.3.1. Fusion utilisant une pyramide Laplacienne .....	23
2.4.3.2. Fusion en utilisant la pyramide de contraste .....	23
2.4.3.3. Fusion utilisant la pyramide de gradient.....	23
2.4.4. Transformée en ondelettes discrète (DWT).....	23
3. Détection de personnes .....	24
3.1. Méthodes de détection du mouvement .....	24
3.1.1. Détection par la méthode de points d'intérêt.....	24
3.1.2 Détection par la méthode du flux optique .....	25
3.1.3. Détection par la méthode soustraction de fond .....	26
4. Suivi du Mouvement.....	27
4.1. Méthodes de suivi.....	27
4.1.1. Représentation et localisation d'objet .....	28
4.1.2. Filtrage et association de données (Data Association) .....	28
5. Conclusion .....	29

### **Chapitre III : Fusion, Suivi et Détection**

1. Introduction .....	30
2. Synoptique général.....	30
3. Acquisition.....	30
3.1. Enregistrement.....	30
4. Fusion.....	32
4.1. Fusion par transformée en ondelettes discrète (DWT).....	32
4.1.1. Principe de DWT : .....	33
4.1.2. Détail et principe de base de DWT.....	35
5. Détection de mouvement .....	37
5.1. Soustraction de l'image de fond .....	37
5.2. Soustraction et moyenne d'images.....	38
5.3. Algorithme.....	39
6. Suivi de mouvement.....	41

6.1. Le filtre de Kalman.....	41
6.2. Fonctionnement .....	41
6.3. Algorithme.....	43
6.4. Principe et points forts.....	43
6.5. Paramètres du filtre.....	44
7. Conclusion .....	44

### **Chapitre IV : Résultats et Discussions**

1. Introduction.....	46
2. L'implémentation de la solution proposée.....	46
3. L'acquisition .....	47
4. Base de données utilisée .....	48
4.1. Base de données 1 : .....	48
4.2. Base de données 02 : .....	49
5. Résultats de Fusion .....	49
5.1. Évaluation qualitative.....	49
5.2. Evaluation quantitative.....	50
5.2.1. Entropie (EN) .....	50
5.2.2. Information mutuelle (MI : Mutual Information).....	50
5.2.3. Erreur quadratique moyenne (MSE : Mean Square Error).....	51
5.2.4. Rapport de signal crête sur bruit ( PSNR :Peak Signal to Noise Ratio ).....	51
5.2.5. Contenu structurelle (SC) .....	52
5.2.6. Corrélation croisée normalisée(NCC : Normalized Cross Correlation).....	52
5.2.7. Erreur absolue normalisée (NAE : NormalizedAbsoluteError) : .....	52
5.2.8.Gradient moyen (AVG : Average Gradient:) .....	53
5.3. Résultats .....	53
5.3.1. Interprétation des résultats.....	58
5.4. L'importance de la fusion.....	59
6. détection par soustraction de fond .....	61
6.1. Arrière-plan .....	61
6.2. Résultat de l'arrière-plan .....	61
6.3. Résultats de soustraction de fond : .....	62
6.4. Détection.....	63
7. suivi.....	64
8. Conclusion.....	69
<b>Conclusion Générale.....</b>	<b>70</b>

## LISTE DES FIGURES

<b>Figure 0.1.</b> Plan général de notre travail.....	4
<b>Figure 1.1.</b> Poste de surveillance d'un système de télévision en circuit fermé.....	6
<b>Figure 1.2.</b> Architecture de système de vidéosurveillance intelligente.....	9
<b>Figure 1.3.</b> L'architecture du système de surveillance multi-caméras.....	10
<b>Figure 1.4.</b> Exemple de changement d'illumination. Les images ont été capturées à 1 seconde d'intervalle.....	15
<b>Figure 1.5.</b> Exemple de scène souffrant d'un manque de résolution.....	15
<b>Figure 2.1.</b> Principe de fusion de données.....	17
<b>Figure 2.2.</b> Fusion au niveau de pixel.....	19
<b>Figure 2.3.</b> Représentation de fusion des données du capteur au niveau des caractéristiques	20
<b>Figure 2.4.</b> Représentation de fusion des données du capteur au niveau décisionnel.....	20
<b>Figure 2.5.</b> Fusion d'image par la combinaison pondérée.....	21
<b>Figure 2.6.</b> Processus d'une fusion d'image par ondelette.....	24
<b>Figure 2.7.</b> Détection par la méthode de point d'intérêt.....	25
<b>Figure 2.8.</b> Détection par la méthode du flux optique.....	26
<b>Figure 2.9.</b> Détection par soustraction de fond.....	27
<b>Figure 3.1.</b> Synoptique général de la présente étude.....	30
<b>Figure 3.2.</b> Décalage entre les deux caméras.....	31
<b>Figure 3.3.</b> Bloc diagramme d'une fusion à base de DWT.....	33
<b>Figure 3.4.</b> Schéma bloc de fusion d'image à base de DWT.....	36
<b>Figure 3.5.</b> Soustraction de l'image de fond.....	37
<b>Figure 3.6.</b> Schéma de principe d'un algorithme de soustraction de fond.....	38
<b>Figure 3.7.</b> Cycle de filtre de kalman.....	41
<b>Figure 4.1.</b> Organigramme de la présente étude.....	47

<b>Figure 4.2.</b> Résultats de fusion par DWT, Gradient, laplacienne et PCA pour le scénario1.....	54
<b>Figure 4.3.</b> Résultat de fusion avec Gradient, DWT, Laplacienne et PCA pour le scénario2.....	55
<b>Figure 4.4.</b> Résultats de fusion par DWT, Gradient, Laplacienne et PCA pour le scénario 3.....	56
<b>Figure 4.5.</b> Résultats de fusion par DWT, PCA Laplacienne et Gradient pour le scénario 4.....	57
<b>Figure 4.6.</b> Difficulté de détection d'objet dans une scène filmée par une caméra couleur.....	60
<b>Figure 4.7.</b> Difficulté de détection d'objet dans une scène filmée par une caméra infrarouge.....	60
<b>Figure 4.8.</b> Possibilité de détection d'objet dans une scène filmée par le système Fusionnée.....	61
<b>Figure. 4.9.</b> Arrière-plan.....	62
<b>Figure 4.10.</b> Possibilité de détection d'objet dans une scène filmée par le système Fusionnée.....	63
<b>Figure 4.10</b> résultat de suivi dans un parking avec un seul objet mobile.....	65
<b>Figure 4.12.</b> Résultat de suivi dans un parking, avec deux objets mobiles.....	66
<b>Figure 4.13.</b> Résultat de suivi, avec deux objets mobiles proches.....	68



## LISTE DES ABBREVIATIONS

---

AVG	Gradient moyen (Average Gradient)
CCD	Capteur photographique Charge-Coupled Device
CCTV	Circuit de télévision fermé (closed-circuit television)
CT	Computed-Tomography
DWT	Transformée en ondelette discrète (Discrete Wavelet Transform)
FT	Transformée de Fourier
IR	Caméra Infrarouge
IVS	System de video surveillance intelligent (Intelligent video surveillance system)
KF	Kalman filter
MI	Information mutuelle (Mutual Information)
MRI	Résonance d'image magnétique (Magnetic Resonance-Image)
MSE	Erreur quadratique moyenne (Mean Square Error)
NAE	Erreur absolue normalisée (Normalized Absolute Error)
NCC	Corrélation croisée normalisée (Normalized Cross Correlation)
PCA	Analyse en composantes principales (principal component analysis)
PSNR	Rapport de signal crête sur bruit (Peak Signal to Noise Ratio)
RGB	Camera couleur (Red, Green, Blue)
SC	Contenu structurelle (Structural Content)

# ***INTRODUCTION GENERALE***

## *INTRODUCTION GÉNÉRALE*

### **1. Introduction**

De nos jours, les technologies d'analyse numérique d'images en temps réel sont de plus en plus utilisées, notamment dans le domaine de la surveillance et la sécurité de lieux sensibles (Banques, routes, lieux publics aéroports, usines...) mais aussi pour le suivi qualité des produits en industrie, ou tout simplement la surveillance domestique.

Avec les développements technologiques, la surveillance intelligente est apparue. Cela consiste en un système d'analyse vidéo en temps réel, de nuit comme de jour. Ces systèmes sont capables de surveiller des lieux par fois avec des installations comportant plusieurs caméras. Ces techniques apportent une assistance très précieuses, parfois essentielles, lors de la surveillance humaine, améliorant la vigilance et faisant accroître ainsi la sécurité, sans compter la réduction des coûts de personnel.

De ce fait, cela représente un marché très important, tant économique que stratégiques. Même si le commerce des caméras est très répondeu et représente un chiffre d'affaire mondial énorme, celui des logiciels de traitement intelligent reste encore en développement.

En effet, il faudra assurer une fiabilité dans la gestion des flux d'images, de leur stockage, mais aussi une robustesse à toute épreuve notamment lors des changement de conditions d'éclairage, ou travailler de nuit comme de jour, ou en cas de réverbération ou d'excès d'enseillement. Il est donc nécessaire d'élargir les conditions d'utilisation. Plusieurs équipes dans le monde travaillent en concert ou en concurrence afin de développer des techniques numériques assurant ces fonctionnalités et améliorant leur fiabilité.

Différentes approches de modélisation permettant le suivi des objets sont développés. Il y a celles basées sur les histogrammes de couleur, celles exploitant les mouvements, d'autres les contours des objets, ou d'autres des caractéristiques de leur émission thermique, notamment dans le domaine des ondes infrarouges.

Cependant, le suivi des séquences d'images pose un vrai problème lié aux conditions dans lesquelles se trouve l'objet, telles que la météo, le changement de luminosité, la nuit ou le jour, ...etc.

Par exemple, pour les camera couleur, il est très difficile de distinguer un mouvement d'une couleur évoluant sur fond de même couleur.

Une combinaison de plusieurs informations est alors nécessaire pour palier à ces bruits parasites, qui nuisent à la qualité de suivi et de la reconnaissance des objets.

Par exemple, une exploitation de la redondance et de la complémentarité des informations que peuvent donner une caméra couleur et une caméra infrarouge permettra de palier au problème de suivi continue dans l'ombre et sous la lumière, de nuit ou de jours. La caméra thermique permet la détection des personnes en fonction des températures, et donc indépendamment des couleurs.

Mais l'utilisation de plusieurs caméras nécessite une gestion d'informations provenant de différentes sources, afin de proposer une interprétation de bonne qualité des événements et des objets.

## **2. Problématique**

De nos jours, la vidéosurveillance est omniprésente et on la retrouve dans de nombreux secteurs d'activité (banque, transports, industrie, grande distribution, etc.) ou lieux de vie (villes, immeubles de bureau, équipements collectifs, etc.). La plupart des responsables souhaitent accroître la sécurité en protégeant les biens et les personnes par la vidéosurveillance. Aujourd'hui, le développement technologique permet d'obtenir de très bons résultats lorsqu'on sait exprimer un besoin et lui faire correspondre un matériel efficace et évolutif afin de pérenniser l'installation.

L'efficacité d'un système de surveillance augmente continuellement, mais elle est limitée par l'information acquise par les caméras.

Lors du suivi d'un objet dans une scène, la détection automatique et le suivi de ce dernier présentent des difficultés telle que la détermination du mouvement de l'objet ainsi que les erreurs émises par le capteur à travers les bruits s'ajoutant aux mesures de la camera.

Le marché de la vidéosurveillance offre présentement deux types de solutions distinctes : les caméras thermiques et les caméras couleur. Les caméras couleur sont efficaces le jour et elles sont surtout utilisées pour sécuriser des lieux et pour faire le suivi de personnes. Elles s'avèrent cependant inefficaces dans un environnement de faible luminosité ou en pleine noirceur (pénombre). De leur côté, les caméras thermiques peuvent détecter un intrus dans le noir le plus total. Toutefois, l'infrarouge ne permet pas l'identification des personnes. Ces

deux systèmes sont donc inadéquats pour répondre aux multiples besoins de l'industrie de la surveillance vidéo.

Afin de remédier à ces contraintes et améliorer la vision et les performances de l'image, l'utilisation combinée de caméras infrarouges thermiques et de caméras couleurs permet de recueillir davantage d'informations car elle fusionne les images thermiques et couleur en une seule et même image. La vidéosurveillance intelligente permet de suivre des objets de jour comme de nuit, dans un environnement extérieur aux conditions atmosphériques changeantes. La fusion des images augmente l'efficacité de la détection et la robustesse de la caméra, quelles que soient les variations climatiques, en plus d'abaisser le taux de fausses alarmes. De cette façon, il est possible de poursuivre la surveillance d'une personne dans l'obscurité ou lorsqu'elle est exposée au vent, ce qui peut la rendre alors invisible en infrarouge

Les questions qui se posent se sont

Comment pouvoir exploiter le maximum d'informations provenant de deux capteurs d'image, couleur et infrarouge ?

Comment pourvoir détecter le mouvement d'un objet dans une scène filmée par les deux caméras ?

Comment pouvoir suivre le mouvement de l'objet détecté ?

### **3. Objectifs**

Dans ce travail, nous visons à réaliser un traitement informatique et numérique de la combinaison d'images issues de caméra thermiques (infrarouge) et de caméra couleur, en réalisant leur fusion en temps réel dans le but de détecter puis suivre le mouvement d'un objet dans une scène.

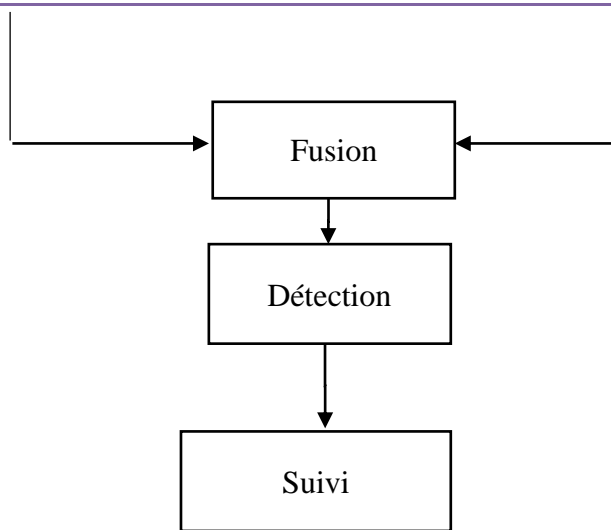
Comme le montre le schéma synoptique présenté dans la figure 1, la scène est supposée filmée par deux caméras, couleur et infrarouge. D'abord une étape de fusion des deux vidéos sera réalisée pour avoir le maximum d'information. Suivie d'une étape de détection d'objet puis son suivi.



**Image couleur**



**Image IR**



**Figure 0.1.** Plan général de notre travail

#### **4. Plan de travail**

Après une définition de la problématique à traiter, la première partie portera une généralité de la vidéosurveillance.

La deuxième partie portera une revue bibliographique décrivant l'état de l'art portant sur les techniques de fusion d'image. On définira ce qu'est cette technique, sa structuration ainsi que son objet. Suivront alors les descriptions des méthodes de fusion d'images usuelles, trouvées dans la littérature une brève description des méthodes de détection ainsi que celle de suivi est aussi présentée dans le deuxième chapitre.

Le chapitre 3 sera consacré à l'utilisation de ces techniques afin de développer un algorithme permettant de réaliser la fusion d'images. La technique DWT sera développé et leur évaluation présentée. L'algorithme de détection par soustraction de fond et celui de suivi, filtre de Kalman , seront aussi présentés dans ce chapitre.

Le chapitre 4 sera consacré à l'analyse des résultats à travers lesquels on proposera des conclusions et des perspectives à ce travail.

## ***I. Généralités sur la vidéosurveillance***



## **1. Introduction**

La vidéosurveillance est de plus en plus utilisée dans la vie quotidienne. Elle s'implique surtout pour augmenter la sûreté des êtres humains.

Dans le présent chapitre, nous présentons une brève description des systèmes de vidéosurveillance et leur évolution. L'architecture des systèmes de vidéosurveillance intelligente est décrite, suivie de ses applications, ensuite ses contraintes applicatives.

## **2. Système de vidéo surveillance**

L'évolution actuelle des appareils électroniques est due à l'apparition des nouvelles techniques nécessaires pour la vie humaine. La vidéosurveillance est l'une parmi elles. Elle consiste en un système de caméras et de transmission d'images, disposé dans un espace public ou privé pour le surveiller. Les images obtenues avec ce système, peuvent être traitées automatiquement. Elle a une grande importance dans le maintien de la sécurité public et privée. [1]



**Figure 1.1.** Poste de surveillance d'un système de télévision en circuit fermé.[5]

### 3. Générations des systèmes de vidéosurveillance

Au cours des deux dernières décennies, les systèmes de surveillance sont devenues un domaine de recherche. Récemment, beaucoup de recherches ont été concentrées sur les systèmes de surveillance basés sur la vidéo, en particulier dans le domaine de la sécurité et des systèmes de transport publics. [2]

Les données sont recueillies par des sources distribuées puis elles sont généralement transmises à quelques centres de contrôle à distance. Le système de contrôle offre un mécanisme pour signaler les événements potentiellement anormaux par la découverte de la structure anormale de l'activité et signalant les moins probables. Les systèmes de surveillance sont classés en trois générations. Chaque génération est fondée sur la précédente et marque un certain changement en termes d'algorithmes nécessaires, afin de permettre plus d'applications, face à de nouveaux défis, et abordant différents domaines de recherche. [2]

Le Tableau 1.1 résume la différence entre les systèmes de surveillance des trois générations et les défis correspondants ainsi que la recherche actuelle. [3]

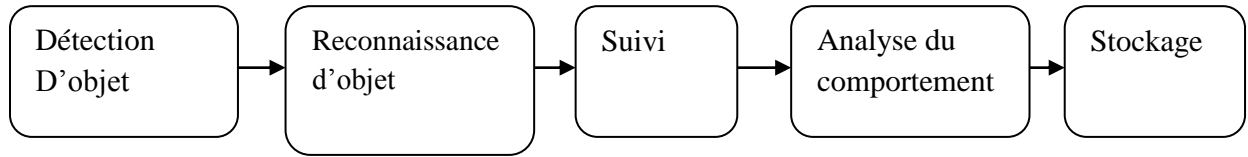
**Tableau 1.1.** Les trois générations des systèmes de vidéosurveillance

	<b>Techniques</b>	<b>Avantages</b>	<b>Inconvénients</b>	<b>Recherche actuelle</b>
<b>1ere génération</b>	Systèmes CCTV analogiques	Ils donnent de bonnes performances dans certaines situations -technologie évoluée	Utilise des techniques analogiques pour la distribution et le stockage d'image	comparative Numérique analogique L'enregistrement de vidéo numérique - Compression de vidéo CCTV

<p style="writing-mode: vertical-rl; transform: rotate(180deg);"><b>2eme génération</b></p>	<p>Automatise les techniques de surveillance visuelle en combinant la technologie de vision par ordinateur avec des systèmes de vidéosurveillance</p>	<p>Accroît l'efficacité de la surveillance des systèmes de vidéosurveillance</p>	<p>Nécessité d'algorithmes robustes de détection et de suivi pour l'analyse comportementale</p>	<p>-L'apprentissage automatique de la variabilité de scène et les modèles de comportements - Comblent le fossé entre l'analyse statistique d'une scène et des interprétations en langage naturel</p>
<p style="writing-mode: vertical-rl; transform: rotate(180deg);"><b>3eme génération</b></p>	<p>Système intelligent de surveillance pour des grands espaces.</p>	<p>Des informations plus précises en raison de la combinaison de différents types de capteurs</p>	<p>Diffusion de l'information (intégration de la communication) - Méthodologie de conception - Plates-formes de transfert, les plates-formes multi-capteurs</p>	<p>Surveillance distribuée par rapport intelligence centralisée - La fusion de données - Cadre de raisonnement probabiliste - Les techniques de surveillance multi-caméras</p>

#### 4. Structure d'un système de vidéosurveillance intelligente

Un système de vidéosurveillance intelligente nécessite pour son fonctionnement plusieurs blocs. Une configuration typique de modules de traitement est illustrée dans la figure. 1.2 Ces modules constituent les blocs de construction de premier niveau nécessaires pour tout système de surveillance distribuée. [3]



**Figure 1.2.** Architecture de système de vidéosurveillance intelligente [3]

Par conséquent, chacune des sections ci-dessous décrit les techniques de traitement d'image les plus populaires utilisés dans chacun de ces modules. [3]

#### **4.1. La détection d'objet**

Il existe deux principales approches conventionnelles à la détection d'un objet: la «différence temporelle» et «soustraction de fond ».

La première approche consiste en la soustraction de deux trames consécutives suivie par seuillage. La seconde technique est basée sur la soustraction d'un arrière-plan ou modèle de référence à l'image courante suivie d'un processus d'étiquetage.

Après l'application de l'une de ces approches, les opérations morphologiques sont appliquées pour réduire le bruit issu de la différence d'image. [3]

La technique de différence temporelle a de bonnes performances dans des environnements dynamiques car elle est très adaptable, mais elle a une mauvaise performance sur l'extraction de tous les pixels pertinents de l'objet. [3]

D'autre part, la soustraction de fond a de meilleures performances pour extraire des informations de l'objet, mais elle est sensible aux changements dynamiques de l'environnement. [3]

#### **4.2. Reconnaissance d'objets, le suivi et l'évaluation de la performance**

Les techniques de suivi peuvent être divisés en deux approches principales: modèles 2-D avec ou sans modèles de formes explicites et des modèles 3-D. [3]

### 4.3. Analyse comportementale

La prochaine étape d'un système de surveillance reconnaît et comprend les activités et les comportements des objets suivis. Cette étape correspond généralement à un problème de classification des données de fonction variant dans le temps qui sont fournis par les étapes précédentes. [3]

### 4.4. Stockage

Une des dernières étapes dans un système de surveillance est le stockage et la récupération de données. Relativement, peu de travaux ont été faits sur la façon de stocker et de récupérer toutes les informations de surveillance obtenues d'une manière efficace, en particulier quand il est possible d'avoir différents formats de données et des types d'information à récupérer. [3]

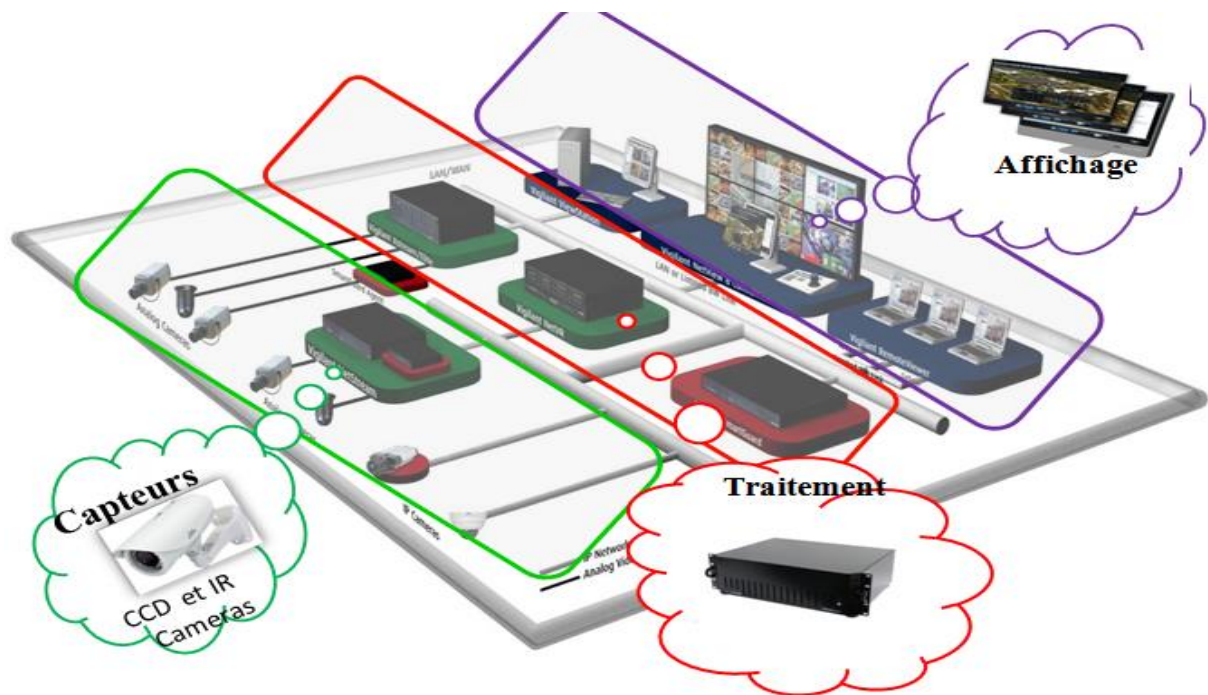


Figure 1.3. L'architecture du système de surveillance multi-caméras

## 5. Raisons pour l'utilisation de vidéosurveillance intelligente

Plusieurs études démontrent les limites de la surveillance humaine. Après seulement 20 minutes à regarder et analyser des écrans de surveillance vidéo, l'attention de la plupart des

individus chute sous un niveau acceptable. Un surveillant ne peut suivre attentivement 9 à 12 caméras plus de 15 minutes. [4]

Dans ce contexte, la vidéosurveillance intelligente présente de nombreux avantages:

- Elle est en fonction 24 heures par jours, 7 jours par semaine.
- Elle peut enclencher une alarme qui sera traitée par un opérateur humain ou commander le déplacement ou zoom d'une caméra pour une surveillance plus précise de l'événement, permettant ainsi une intervention en temps réel, plutôt qu'après l'événement.
- Elle réduit la bande passante et l'espace d'archivage nécessaire en ne transmettant ou n'enregistrant que les données sur les événements pertinents.
- Elle libère le personnel de sécurité d'une surveillance continue.
- Elle permet la recherche rapide d'événements pertinents dans les séquences vidéo archivées.
- Elle permet d'identifier les objets dans une scène et de suivre leurs activités. [4]

### **6. Applications de la vidéosurveillance**

À son origine, la vidéosurveillance a d'abord été utilisée par les services publics (police, transports, administrations). Elle fut ensuite adoptée par des entreprises désirant protéger des actifs stratégiques, telles que les raffineries, les centrales nucléaires, les barrages fluviaux, les usines agroalimentaires et les complexes pharmaceutiques. Les casinos apparaissent aussi comme des précurseurs dans le déploiement de grands systèmes de vidéosurveillance. De nos jours, les caméras de surveillance sont présentes dans divers lieux publics et privés : immeubles, commerces, stationnements, gares, aéroports, routes, transports publics, banques, etc. [4]

#### **6.1. Gouvernement et sécurité publique**

A différents niveaux, les gouvernements doivent assurer la sécurité de la population, ainsi que des infrastructures publiques. Au niveau national, la vidéosurveillance est utilisée, par exemple, pour surveiller les éléments suivants :

- Infrastructures sensibles
- Frontières
- Édifices et sites gouvernementaux
- Laboratoires

- Bases militaires
- Prisons

## **6.2. Éducation**

La vidéosurveillance est de plus en plus présente dans les établissements scolaires. Son utilisation vise à assurer la sécurité des enseignants et étudiants, ainsi que la protection des actifs contre le vandalisme et le vol. Dans ce milieu, la vidéosurveillance sert notamment à :

- surveiller l'accès au périmètre de l'établissement, parfois étendu, dans le cas d'un campus universitaire.
- surveiller le matériel, les équipements et les données.
- détecter et suivre les actes de vandalisme, vols, méfaits, comportements inappropriés.
- reconnaître les plaques d'immatriculation.
- aider les enquêtes criminelles.
- contrôler les accès. [4]

## **6.3. Commerce de détail**

Le commerce de détail représente un marché en expansion pour la vidéosurveillance. Celle-ci y est utilisée autant pour des aspects de sécurité intérieure (magasin, entrepôt) qu'extérieure (stationnement). Même les plus petits commerces se dotent de caméras pour, au moins, conserver des preuves vidéo en cas de vols ou d'incidents. Dans les chaînes de magasins, des systèmes de vidéosurveillance beaucoup plus sophistiqués sont déployés pour permettre la surveillance centralisée des différentes succursales. Dans l'ensemble du secteur, la vidéosurveillance visera notamment :

- la surveillance des caisses et des transactions (vols, fraudes par les employés).
- la protection des matériels et des infrastructures.
- la surveillance de l'inventaire, des marchandises (livraisons).
- la protection du personnel et des clients.

- le contrôle d'accès aux aires verrouillées.
- la vérification des situations d'urgence (incendie, alarmes, etc.).
- la surveillance des stationnements, des véhicules, des entrées et sorties. [4]

#### **6.4. Transports**

La sécurité et le bon fonctionnement des aéroports, des gares, des ports et des transports publics est critique pour l'économie d'un pays. Un incident de sécurité peut gravement perturber leurs opérations et se traduire en pertes significatives.

Or, étant donné les flots importants de passagers qui y transitent et l'étendue de leurs infrastructures, les systèmes de transports font face à des défis de sécurité hors du commun. Les actes terroristes commis dans différents moyens de transports dans le monde ont exacerbé ces défis. [4]

#### **6.4. Milieu bancaire**

La vidéosurveillance est très utilisée pour assurer la sécurité des banques. Dans un premier temps, la présence de caméras opère un certain pouvoir de dissuasion sur la perpétration de vols à mains armées et d'agressions. Si un tel délit survient, les séquences vidéo archivées servent à l'enquête et l'identification des malfaiteurs.

Les guichets automatiques sont particulièrement à risque pour les actes criminels. Les caméras de surveillance aident à détecter les fraudes, comme par exemple, l'installation d'un dispositif pour lire l'information magnétique des cartes de guichet.

Pour le milieu bancaire, la vidéosurveillance intelligente peut augmenter l'efficacité de la surveillance. Elle permet d'assurer un suivi sur l'ensemble du réseau des succursales, afin d'y détecter les individus ou comportements suspects. Elle permettra, entre autres, de retrouver toutes les séquences vidéo provenant de toutes les succursales où apparaît un certain individu, grâce à des techniques de reconnaissance faciales. [4]

### **7. Contraintes de la vidéosurveillance**

Les systèmes de vidéosurveillance sont en pratique, particulièrement contraints. On peut citer :

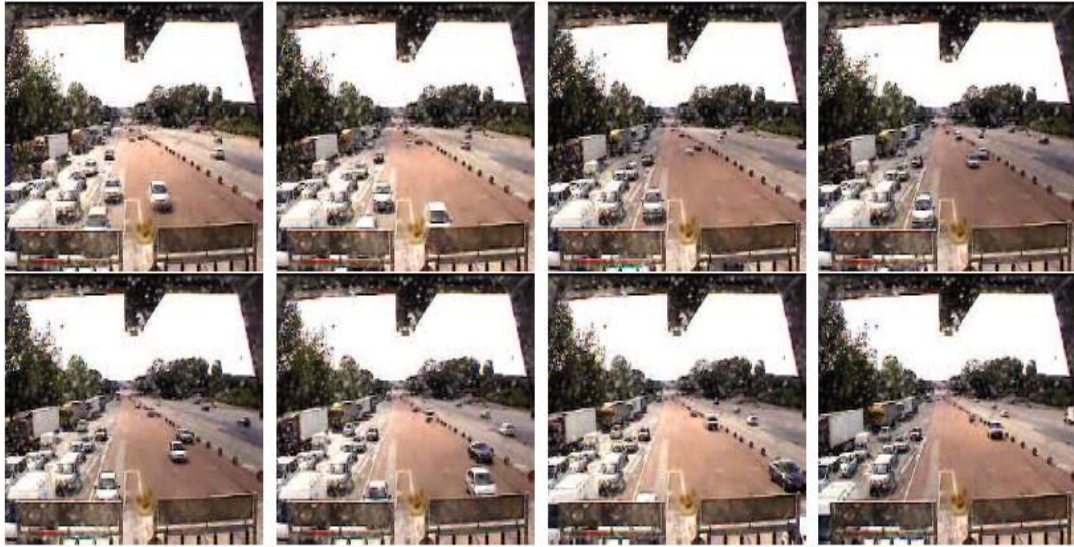


- Des contraintes techniques, comme par exemples la calibration de caméra ou de réseaux de caméras, les éventuelles distorsions provoquées par les capteurs, des taux de rafraîchissement faibles dans certaines situations, etc.
- Des contraintes environnementales, comme le manque de résolution dans certaines situations et les changements d'illumination... etc. Ce sont des problèmes récurrents en vision par ordinateur et il n'existe pas encore de solutions « miracle ». Ils se traitent en général au cas par cas selon la scène observée, ce qui rend la généralisation d'un système de vidéosurveillance d'autant plus difficile.
- Des contraintes sémantiques, les systèmes mis en œuvre sont généralement ciblés pour un type ou une plage d'évènements bien spécifiques. Par exemple, la détection de personnes trop proche du bord d'un quai de gare ou la détection de présence de personnes sur la chaussée, etc.

En pratique, le système sera défini soit par des règles bien précises correspondant à la situation, on parlera d'approches basées modèles, soit par des approches statistiques, on parlera dans ce cas d'approches basées apprentissage. Plus spécifiquement, lors de l'analyse de scènes denses, telles que des foules, on rencontre de nouveaux problèmes tels que :

- Des contraintes de traitement de volume d'information importantes. Une foule représente une quantité phénoménale de données à traiter, ce qui peut poser des problèmes de temps de calcul pour la vidéosurveillance temps-réel.
- Des contraintes de point de vue qui sont d'autant plus sensibles que sur des vidéos denses, certains points de vue peuvent souffrir d'occultations très importantes et très longues. [5]

Deux exemples de contraintes environnementales sont présentés dans les figures 1.4 et 1.5.



**Figure 1.4.** Exemple de changement d'illumination. Les images ont été capturées à 1 seconde d'intervalle. [5]



**Figure 1.5.** Exemple de scène souffrant d'un manque de résolution. [5]

## 8. Conclusion

Dans ce chapitre nous avons présenté la vidéosurveillance en général, puis nous avons insisté sur la vidéosurveillance intelligente. Cette dernière, semble pouvoir résoudre beaucoup de problème de sécurité, mais a encore des limites et des difficultés. Dans le prochain chapitre, nous allons parler de l'une des solutions possibles.

## *II. Etat de l'Art*

## **1. Introduction**

La détection d'objet et son suivi sont des vrais challenges dans le domaine de la vidéosurveillance. La fusion de données est une solution qui peut soulever plusieurs problèmes de la vidéosurveillance. Elle sera étudiée dans la première partie du chapitre. Ensuite, une étude bibliographique sur les méthodes de détection et celles de suivi sera exposée, afin de pouvoir choisir les techniques qui seront utilisées par la suite

## **2. Fusion d'image**

La fusion provenant de plusieurs capteurs a habituellement pour but de recueillir plus d'information qu'il n'est possible d'en recueillir avec un seul capteur et pour rendre le système robuste aux variations de la scène, comme la poussière ou la fumée, et les conditions de l'environnement, c.-à-d. de jour ou de nuit.

La fusion donne une certaine forme de redondance qui permet de corrélérer les données acquises et de réduire l'influence du bruit. De plus, avec l'utilisation de capteurs de types différents (comme dans le cas de caméras visibles et de caméras infrarouges) la fusion apporte des informations complémentaires. [1]

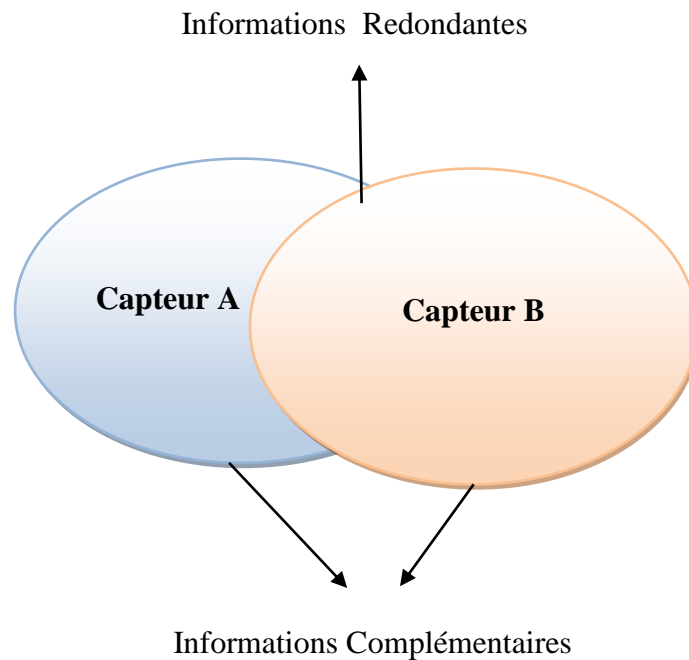
Dans notre étude, la fusion est un appariement entre un objet détecté en infrarouge et un objet détecté en visible. Cette fusion est utilisée pour suivre l'objet dans le temps et faire face aux occultations, en plus de permettre une modélisation plus complète.

### **2.1. Objectif de la fusion d'image**

La fusion est devenue un aspect important de traitement de l'information dans plusieurs domaines très différents elle peut présenter plusieurs intérêts :

- ✓ Améliorer la fiabilité (par informations redondantes).
- ✓ Améliorer la capacité (par l'information complémentaire).
- ✓ Améliorer la résolution spatiale.
- ✓ Combiner l'information diachronique ou la réduction du nombre d'image a traité.
- ✓ Améliorer la correction géométrique nécessaire d'enregistrement dans l'image.

-Dans l'application de la stéréophotogrammétrie des ensembles de stéréo sont créés pour surmonter le manque d'information dus aux obstacles tels que la nébulosité. [6]



**Figure 2.1.** Principe de fusion de données

-Dans des applications liées à la surveillance géologique ou des phénomènes liés aux conditions atmosphériques la fusion peut:

- ✓ Détecter des changements de données temporelles
- ✓ Substituer le manque d'information et surmonter celles qui surgissent en raison d'obstacles ambiants pendant l'acquisition d'image. [6]

## 2.2. Applications de la fusion d'image

La fusion a un très vaste domaine d'application et on la trouve dans différents domaines on peut citer : [6]

### ➤ **Domaine militaire**

Pour des applications comme la détection, l'identification et la poursuite d'une cible, la détection des mines, l'évaluation d'une situation, l'authentification d'une personne. Les aviations civile et militaire, ont employé des techniques de fusion pour faciliter la navigation. Une application classique consiste à fusionner des images issues de capteurs différents pour améliorer la navigation dans des conditions météorologiques défavorables. [6]

### ➤ **La télédétection**

Est un domaine pour lequel l'un de ces objectifs est de fusionner des images aériennes ou satellitaires utilisant des capteurs multiples caractérisés par des résolutions spatiales

et/ou fréquentielles différentes. Ce domaine d'application concentre aujourd'hui une bonne partie des travaux de recherche en fusion

L'enjeu est la détection de cibles naturelles ou artificielles : les rivières, les rues, les bâtiments, etc. [6]

➤ **Le domaine de l'imagerie médicale**

La fusion joue un rôle important dans le diagnostic, le traitement ou la modélisation du corps humain. En particulier, on retrouve des techniques d'images relativement récentes comme la résonance magnétique (MRI Magnetic -Resonance-Image) ou la tomographie à rayons X (CT Computed-Tomography). Ces techniques permettent la détection de tissus de différentes duretés, et un processus de fusion permet l'intégration de ces informations, l'inscription spatiale de la surface 3-D et peut aussi faire la chirurgie assistée par ordinateur. [6]

➤ **La robotique**

-Exiger la commande de mouvement, basée sur les commentaires de l'environnement visuel, tactile, et d'autres types de capteurs

-Contrôler la visualisation intelligente

-Reconnaître automatiquement la cible et l'objet suivi

➤ **D'autres applications** possibles comme la surveillance d'une ligne de production, le contrôle de la qualité, l'observation des défauts, ou l'amélioration de la vision par ordinateur. [6]

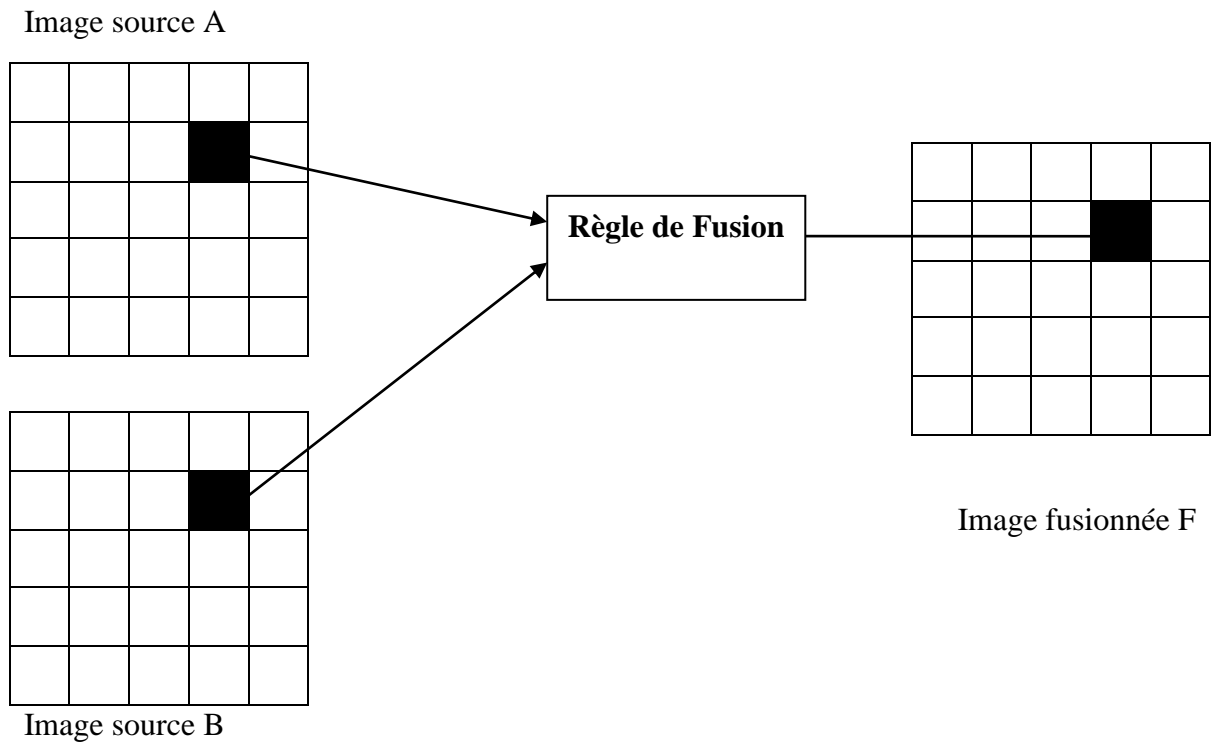
### **2.3. Niveaux de fusion**

La fusion d'images peut être effectuée à l'un des trois niveaux suivants: niveau signal ou de pixel, niveau des caractéristiques, et le niveau de décision. Une élaboration sur les deux premiers niveaux est présentée dans ce chapitre et la fusion au niveau de décision est toujours considéré comme théorique. [1]

#### **2.3.1. Fusion au niveau de pixel**

Dans la fusion à base de pixels, les images source sont traitées pixel par pixel. Une règle de fusion est appliquée aux deux pixels au même endroit dans les différentes images source.

La valeur de pixel résultante est ensuite utilisée pour représenter le pixel à cet emplacement dans l'image composite. Le procédé est illustré sur la Figure. 2.2. [1]



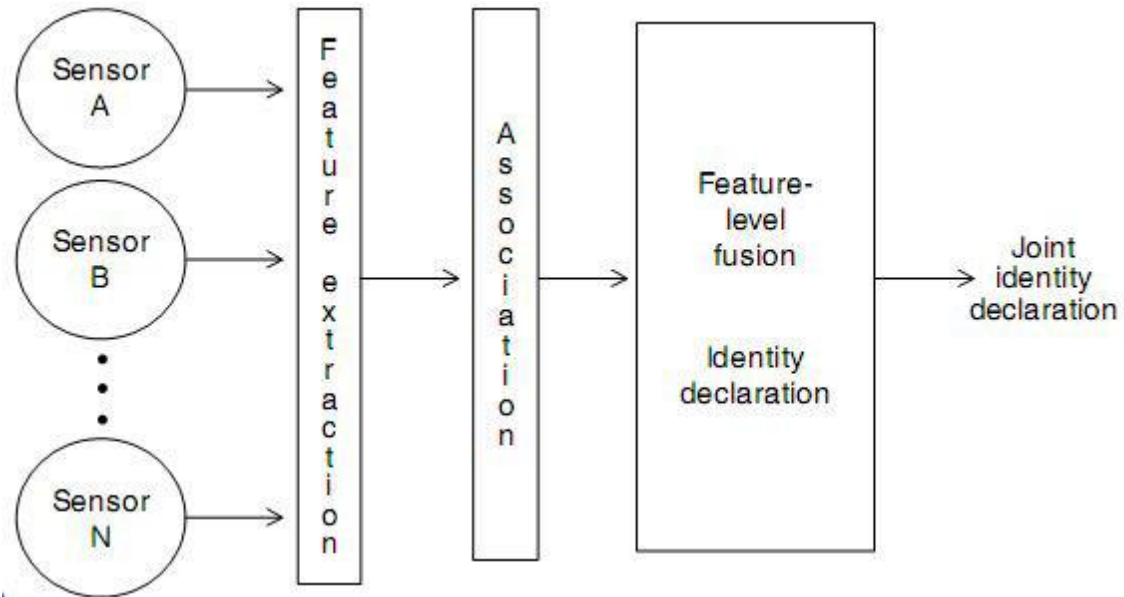
**Figure 2.2.** Fusion au niveau de pixel [1]

### 2.3.2. Fusion au niveau des caractéristiques

Elle implique la représentation des données de capteurs en termes de vecteurs de caractéristiques. Le résultat est obtenu par l'extraction de caractéristiques spécifiques à partir des images en utilisant des techniques de segmentation, qui sont ensuite fusionnées en utilisant des approches statistiques ou réseaux de neurones artificiels (ANN). [7]

L'objectif est d'obtenir un nombre limité de caractéristiques comme preuve confirme que les humains utilisent une fonction cognitive basée sur les caractéristiques de reconnaître les objets. Dans le cas de multi-capteurs, les caractéristiques sont extraites de multiples observations de capteurs et combinées en un seul vecteur enchaîné qui est une entrée à des techniques de reconnaissance des formes telles que les réseaux de neurones

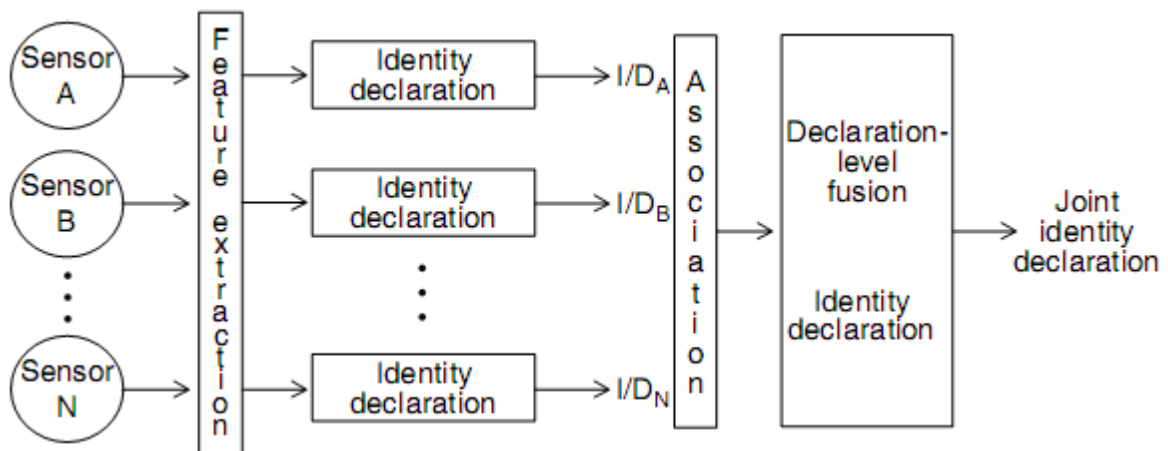
Une illustration de fusion au niveau de pixel est représentée sur la figure 2.3



**Figure 2.3.** Représentation de fusion des données du capteur au niveau des caractéristiques [7]

### 2.3.3. Niveau décisionnel

Représente une méthode qui utilise des données à valeur ajoutée où les images d'entrée sont traitées individuellement pour l'extraction de l'information, dans lequel chaque capteur permet une détermination préliminaire de l'emplacement de l'entité, ses attributs, et son identité. L'information obtenue est ensuite combinée avec l'application des règles de décision pour renforcer l'interprétation commune et fournir une meilleure compréhension des objets observés. [7]



**Figure 2.4.** Représentation de fusion des données du capteur au niveau décisionnel [7]



## 2.4. Méthodes de fusion

Les algorithmes de fusion les plus utilisés sont regroupés en deux catégories: algorithme à base arithmétique et à base biologique. [1]

### I. Algorithme arithmétique

#### 2.4.1. Combinaison pondérée

C'est la forme la plus simple de fusion. Le pixel F fusionné à l'emplacement x, y,  $f(x, y)$ , est calculé en fonction d'une combinaison pondérée des pixels dans les images source, A et B, à l'emplacement x, y, représentée en (2.1). [1]

$$F(x, y) = w_1 \cdot A(x, y) + w_2 \cdot B(x, y) \quad (2.1)$$

Où  $w_1$  et  $w_2$  sont des poids choisis en conséquence. La méthode de fusion moyenne définit comme suit:

$$F(x, y) = 0.5 * (A(x, y) + B(x, y)) \quad (2.2)$$

La combinaison pondérée est une méthode simple et a calcul efficace. En outre, il supprime le bruit trouvé dans les images sources en raison de la fonction moyenne. Cependant, l'image fusionnée souffre de la perte de contraste et de suppression des principales caractéristiques.[1]

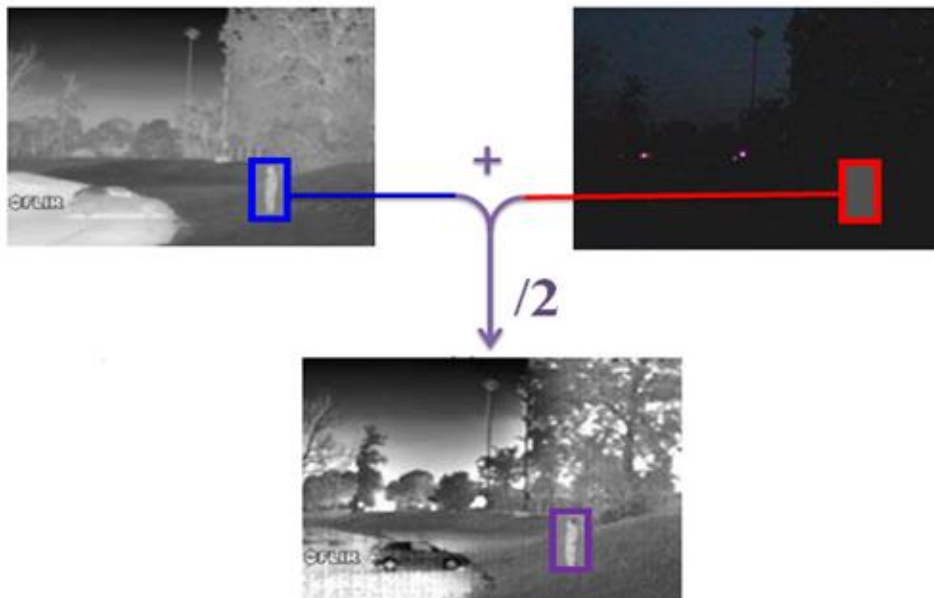


Figure 2.5. Fusion d'image par la combinaison pondérée

### 2.4.2 Analyse en composantes principales (ACP)

L'analyse en composante principale (ACP) a beaucoup d'applications dans le traitement d'image. Elle s'agit d'une approche à la fois géométrique (les variables étant représentées dans un nouvel espace, selon des directions d'inertie maximale) et statistique (la recherche portant sur des axes indépendants expliquant au mieux la variabilité et la variance des données). Lorsqu'on veut compresser un ensemble de  $N$  variables aléatoires, les  $n$  premiers axes de l'analyse en composantes principales sont un meilleur choix, du point de vue de l'inertie ou de la variance. Les algorithmes basés arithmétique connus utilisent la combinaison pondérée.

La matrice de covariance des images sources est obtenu par:[1]

$$C = \begin{bmatrix} V_A & C_{AB} \\ C_{AB} & V_B \end{bmatrix} \quad (2.3)$$

$$V_A = \frac{1}{m*n} \sum_{x,y} (A(x,y) - \mu_A)^2 \quad (2.4)$$

$$V_B = \frac{1}{m*n} \sum_{x,y} (B(x,y) - \mu_B)^2 \quad (2.5)$$

$$C_{AB} = \frac{1}{m*n} \sum_{x,y} (A(x,y) - \mu_A)(B(x,y) - \mu_B) \quad (2.6)$$

Avec  $\mu_A$  et  $\mu_B$  sont les moyenne des deux images A et B respectivement .

Après dérivation de la matrice de covariance des images sources A et B, les optimaux de poids sont choisis en tant que les éléments du vecteur propre de la plus grande valeur propre de la matrice de covariance C.

Toutefois, la PCA a tendance à choisir l'une des images source à la place de leur fusion.

En outre, elle amplifie le bruit dans les images sources. [1]

## II. Algorithme biologique

### 2.4.3. Approches pyramidales

#### Une image pyramide

Se compose d'un ensemble de passe-bas ou passe-bande des copies d'une image, chaque copie représentant des informations de configuration d'une échelle différent Typiquement dans une pyramide d'images tous les niveaux est un facteur deux plus petits que son prédécesseur, et les niveaux plus élevés se concentreront sur la partie inférieure des

fréquences spatiales. Une pyramide d'images ne contient pas toutes les informations nécessaires pour reconstruire l'image originale.

#### **2.4.3.1. Fusion utilisant une pyramide Laplacienne**

La fusion de la pyramide de Laplace est un processus itératif de calcul de pyramides gaussiennes et Laplacienne de chaque image source. Pour calculer la pyramide Laplacienne on soustrait les images de chaque niveau de la pyramide gaussienne avec l'image du niveau inférieur agrandi. [1].

#### **2.4.3.2. Fusion en utilisant la pyramide de contraste**

Le principe de la fusion pyramidale de contraste est toujours le même sauf que cette fois-ci, on remplace l'opérateur par un calcul de contraste.[1]

#### **2.4.3.3. Fusion utilisant la pyramide de gradient**

Une pyramide de gradient est obtenue en appliquant un ensemble de quatre filtres de gradient directionnel (horizontales, verticales et 2 diagonales) à la pyramide gaussienne. A chaque niveau, ces quatre pyramides de gradients directionnels sont combinées ensemble pour obtenir une pyramide de gradient combiné qui est similaire à une pyramide laplacienne. La fusion de la pyramide de gradient est donc la même que la fusion à l'aide pyramide laplacienne excepté le remplacement de la pyramide de Laplace avec la pyramide de gradient combiné. [1]

#### **2.4.4. Transformée en ondelettes discrète (DWT)**

DWT (Discrète Wavelet Transforme) est similaire à FT en termes de représentation et d'analyse des données. DWT représente un signal  $f(X)$  comme une superposition d'ondelettes comme représenté en (2.7).

$$f(x) = \sum_{p,q} C_{p,q} \psi_{p,q}(x) \quad (2.7)$$

Où les fonctions ondelettes  $\psi_{p,q}(x)$  sont dilatées

Ce qui constitue la différence principale avec la transformée en ondelettes continue, où son implémentation en séries discrètes dans le temps parfois appelée transformée en ondelettes continue discrète dans le temps (DT-CWT, discrète-time-continuos-wavelet-transform).L'ondelette peut être construite à partir d'une fonction d'échelle décrivant les propriétés d'échelle du signal. [1]

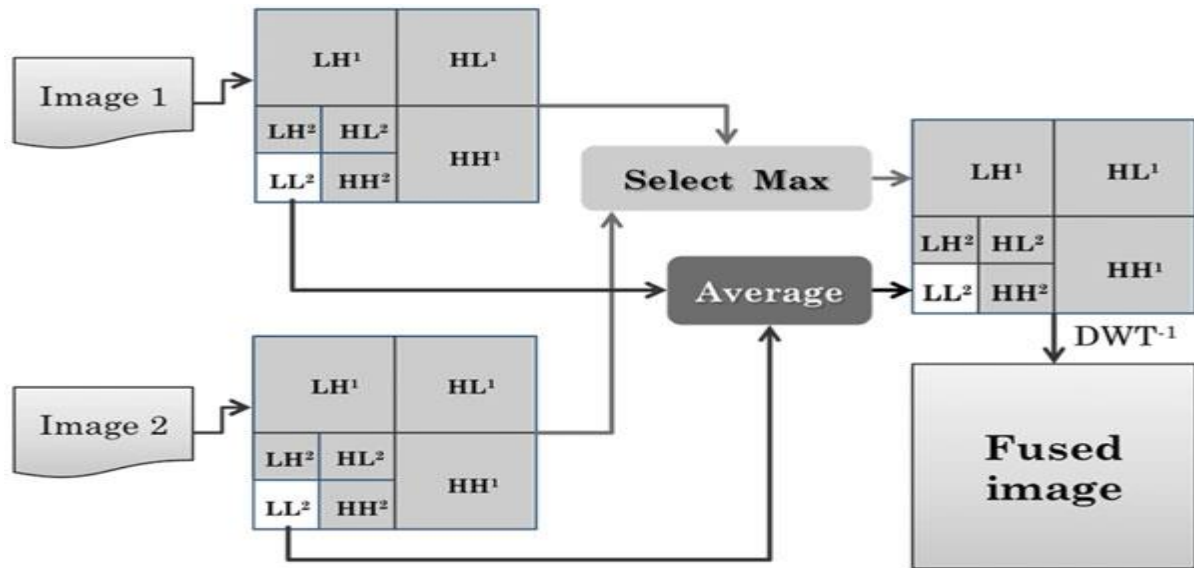


Figure 2.6. Processus d'une fusion d'image par ondelette [1]

### 3. Détection de personnes

La détection d'objets est une étape clé dans plusieurs applications de vidéosurveillance. Certaines de ces applications comprennent la reconstruction 3-D, la compression, l'imagerie médicale, récupération d'image et la surveillance. Elle est particulièrement importante pour la visualisation des capteurs de surveillance. Comme elle segmente l'image d'entrée en premier et arrière-plan.[8]

#### 3.1. Méthodes de détection du mouvement

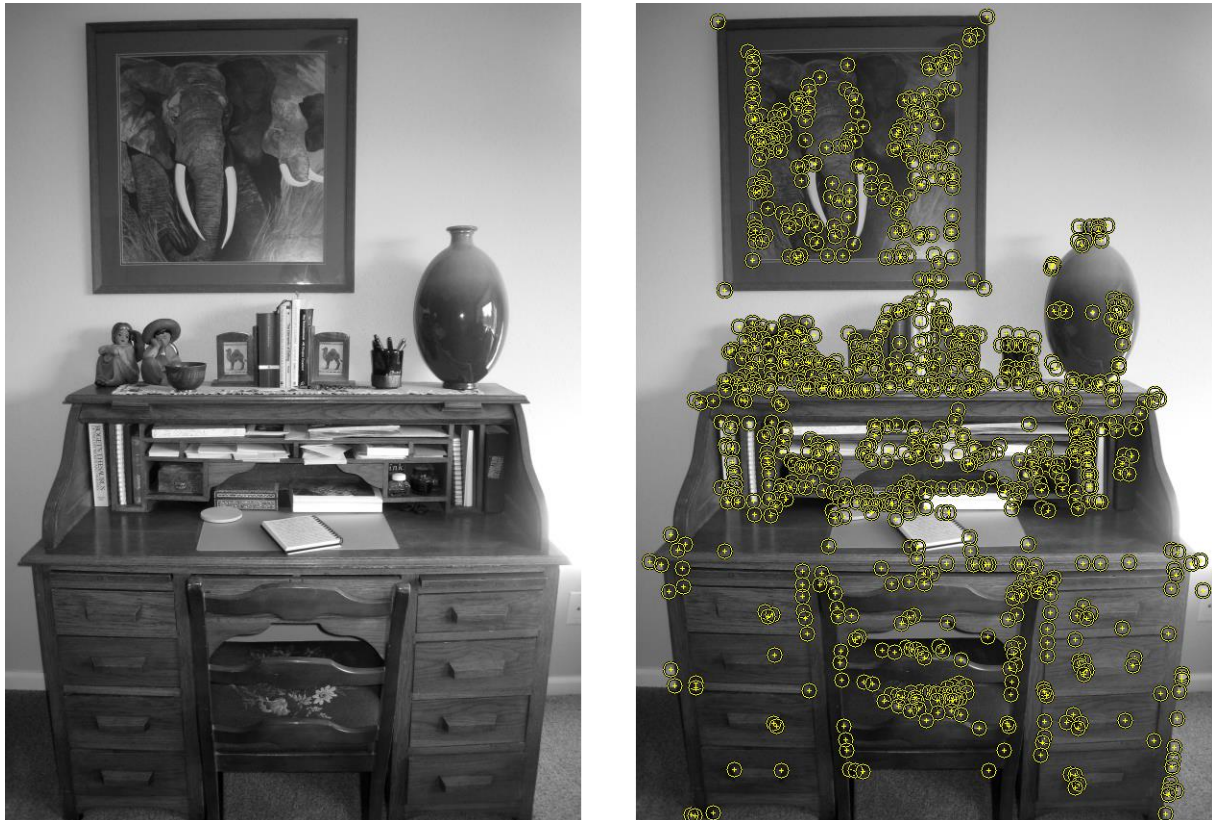
La détection de l'objet a reçu beaucoup d'attention en raison de sa nécessité dans une large gamme d'applications de vision. Il existe plusieurs techniques pour la détection d'objets suivants principalement trois approches: flux optique, le point d'intérêt, et soustraction de fond

##### 3.1.1. Détection par la méthode de points d'intérêt

L'idée générale de cette méthode est de détecter des points d'intérêt dans l'image courante et de les comparer avec ceux du modèle de fond. Le modèle de fond doit donc être une image de luminance.

Les points d'intérêt sont détectés avec un seuil faible sur le score de Harris afin d'avoir des points sur toutes les zones texturées de l'image. Cependant l'extraction de points d'intérêt,

surtout avec un seuil faible, n'est pas un processus stable (comme illustré en figure 2.7). Il est possible qu'un point d'intérêt détecté dans l'image de fond ou l'image courante ne soit pas détecté dans l'autre image. Pour cette raison une fois que les points d'intérêt sont détectés dans chacune des images, nous construisons la liste de l'union des points d'intérêt détectés dans les deux images. Ainsi, même si un point d'intérêt n'est détecté que dans l'une des images il sera tout de même pris en compte dans les deux et est donc garanti d'avoir un candidat potentiel à l'appariement. [8]



**Figure 2.7.** Détection par la méthode de point d'intérêt

Les points n'ayant pu être appariés mettent donc en évidence un changement de texture et donc la présence d'un objet d'intérêt. Nous utilisons pour cela une méthode d'estimation non paramétrique pour estimer la fonction densité de probabilité de présence. [8]

### 3.1.2 Détection par la méthode du flux optique

La mesure du flux optique est une étape de traitement d'image dite de bas niveau. On lui trouve dans de nombreuses applications comme l'analyse de mouvements en physique expérimentale, la compression de séquences d'images vidéo par compensation de

mouvement, ou son utilisation pour des phases de traitement des images de plus haut niveau, comme la reconstruction de scènes tridimensionnelles. [9]

Le flux optique figure (2.8) désigne le champ de vitesses apparent observé entre deux images successives d'une même scène. Son estimation est un intermédiaire essentiel dans de très nombreuses applications et nous proposons d'appliquer des techniques venant de la communauté de la vision par ordinateur à la vélocimétrie par images de particules.



**Figure 2.8.** Détection par la méthode du flux optique

De nombreux auteurs ont vite compris l'avantage qu'il y avait à effectuer une mesure Multi-échelles du mouvement. Pratiquement toutes les méthodes ont une déclinaison multi-échelles. Le gain est de réduire la complexité de calcul pour les méthodes d'appariement, et d'augmenter la gamme des déplacements mesurables pour les méthodes basées sur des lettrages en éliminant un choix a priori de paramètre d'échelle qui limitait cette gamme de mesures. [10]

### **3.1.3. Détection par la méthode soustraction de fond**

La soustraction de fond est calculée en prenant la différence d'intensité de l'image actuelle avec l'image de fond. L'image de fond c'est une image fixe de l'image sous objets et l'image actuelle est l'image résultat Figure (2.9) de l'acquisition de la vidéo par la caméra. [11]

Cette méthode populaire est utilisée par plusieurs applications. Cette façon détecte la région de mouvement en soustrayant pixel par pixel l'image courante à l'image de fond.  
[12]



**Figure 2.9.** Détection par soustraction de fond

#### **4. Suivi du Mouvement**

L'objectif du suivi est de localiser comment les objets de premier plan se déplacent au fil du temps en reliant les instances du même objet dans des frames successives. Cela est essentiel pour plusieurs applications telles que la sécurité et la surveillance, la compression de vidéo, traitement médical et de chambres élégantes, le trafic la gestion et les jeux informatiques interactifs.

##### **4.1. Méthodes de suivi**

Le Suivi d'objet est une tâche importante et difficile en vidéosurveillance. un grand nombre d'approches ont été proposées pour résoudre ce problème. Ceux-ci peuvent être classés suivant deux façons.

Une façon de représentation et localisation d'objet qui appartient à l'un des quatre groupes: suivi basé sur contour, basé sur région, à base de fonction, et le suivi basé sur un modèle. L'autre façon base sur le filtrage. [1]

#### **4.1.1. Représentation et localisation d'objet**

Toutes les approches reposent sur les différences d'apparences et leur suivi d'une frame à l'autre. Elles sont basées sur la segmentation du mouvement. Le processus comprend généralement les étapes suivantes pour associer les objets dans différents cadres : Le cadre est d'abord segmenté pour identifier son déplacement (détection d'objet). Chaque objet est représenté suivant un certain modèle ou par certaines caractéristiques (extraction de caractéristiques). Les objets sont suivis par leurs fonctionnalités correspondantes au cours du temps (objet correspondant). [1]

#### **4.1.2. Filtrage et association de données (Data Association)**

Le filtrage estime la position des objets en mouvement dans le nouveau cadre, étant donné certaines connaissances a priori sur ces caractéristiques et positions dans la trame précédente.

Dans le suivi du mouvement, le filtre de Kalman est une méthode de prédiction intéressante. Un filtre de Kalman est essentiellement un ensemble d'équations récursives. Elles sont utilisées pour aider à modéliser et évaluer très précisément le mouvement d'un système dynamique linéaire. Le filtre de Kalman est appelé 'discret' parce que l'algorithme est discret dans le temps.

En utilisant une multitude de différents vecteurs et des matrices, le filtre de Kalman est capable de construire une représentation interne du système et grâce à cette représentation, on peut estimer l'état futur du système.

Dans le filtrage de Kalman, les équations décrivent l'état du système à des temps successifs. Grâce à ces équations, on peut prédire l'état futur en utilisant l'état actuel sans dépendre des états antérieurs dans les calculs.

Dans le processus du suivi, le filtre Kalman 'regarde' un objet quand il se déplace, c'est à dire il prend les informations sur l'état de l'objet à ce moment-là. Puis il utilise ces informations pour prédire où se trouve l'objet dans la trame prochaine. La position de l'objet est corrigée ensuite en considérant la prédiction et aussi l'observation. [13]



## **5. Conclusion**

Dans ce chapitre, on s'est intéressé par la fusion d'images et ses niveaux. Nous avons présenté quelques techniques utilisées dans la littérature pour fusionner deux images, néanmoins chaque technique a ses avantages et ses inconvénients. Quelques méthodes de suivi ont été données en général. Parmi elles, la méthode de soustraction de fond, qui donne des résultats satisfaisants d'après la littérature, que nous allons adopter pour résoudre notre problème. Et finalement, l'algorithme de suivi qui a prouvé ses performances dans plusieurs applications, le filtre de Kalman, il sera utilisé pour garantir le suivi, dans le prochain chapitre.

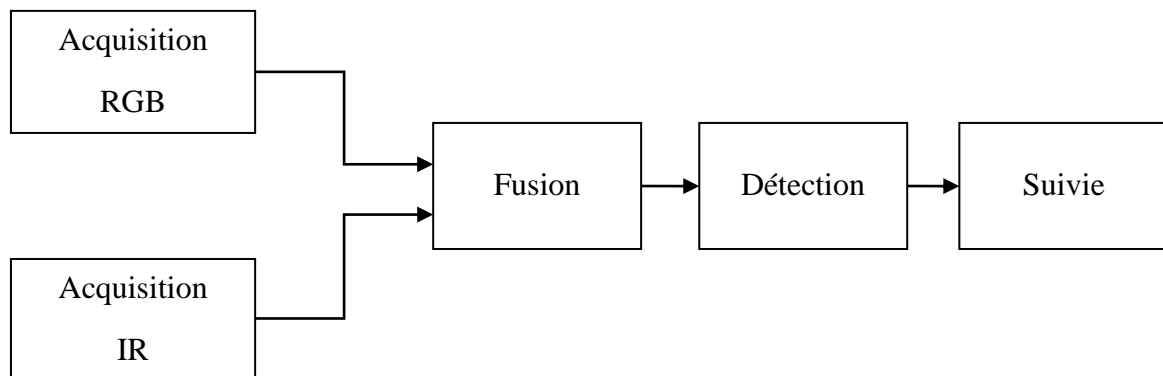
### *III. Fusion, Détection et Suivi*

## 1. Introduction

Ce travail a pour but de réaliser un système de détection et de suivi, qui combine deux types de vidéos : couleur et infrarouge. Dans le présent chapitre, nous exposons la solution proposée pour notre problématique. D'abord on donne un schéma synoptique général du système, puis nous détaillons les différentes parties du système : l'acquisition, la fusion, la détection et le suivi.

## 2. Synoptique général

L'objectif de notre travail est de détecter et suivre un objet en mouvement dans une scène filmée par deux types de caméras, couleur et infrarouge. Afin de pouvoir résoudre le problème, on opte pour la solution résumée par le synoptique ci-dessous.



**Figure 3.1.** Synoptique général de la présente étude

## 3. Acquisition

Il existe une panoplie de modèles de caméras répondant à différents besoins de surveillance. Elles sont analogiques ou numériques et peuvent être motorisées ou non. Pour notre étude, on utilise une caméra Infrarouge(IR) et une couleur (RGB), vu qu'elles sont complémentaires.

### 3.1. Enregistrement

La caméra infrarouge utilise des détecteurs thermiques ou photoniques qui sont capables de faire la différence dans le rayonnement infrarouge de différents objets. Les capteurs électro-optiques, par exemple CCD ou CMOS caméras, capturent la réflexion de la lumière d'objets

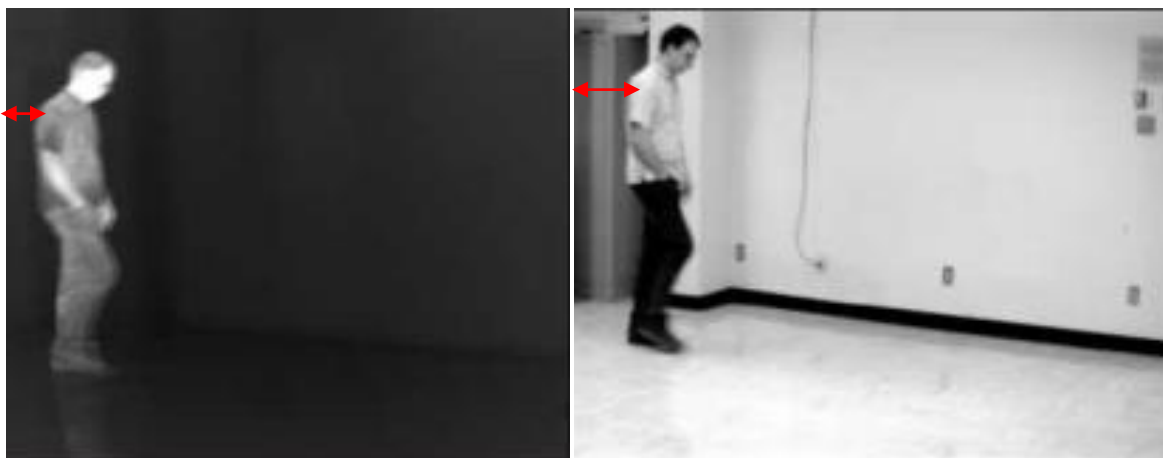
[14]. Par conséquent, la fusion des deux caméras peut fournir les informations complémentaires sur la scène [14].

Avant tout autre traitement, les images couleur et IR des séquences vidéo doivent être enregistrées de façon que les pixels correspondants aux deux images soient associés aux mêmes points physiques de la scène.

L'enregistrement des images à partir de capteur CCD et infrarouge peut être mis en œuvre par l'étalonnage des deux caméras. Liu, Z. Propose une procédure d'étalonnage [15], où une grille de fils chauffés électriquement a été utilisée. Cette grille de fil est apparue comme une grille de fil rouge dans une image en couleur et une grille métallique lumineuse dans une image thermique. L'étalonnage permet d'assurer les lignes de la grille dans les deux cadres de l'appareil photo correspondants. Les lignes de grille détectées à partir de l'image IR correspondront aux lignes extraites de l'image visuelle.

Les images provenant de plusieurs capteurs ont généralement différentes représentations géométriques, elles doivent être transformées en une représentation commune pour la fusion. Cette représentation conserve la meilleure résolution de l'un des capteurs. L'alignement des images multi-capteurs est aussi l'une des étapes de prétraitement les plus importantes de la fusion d'images. L'enregistrement des multi-capteurs est également affecté par les différences dans les images de ces derniers.

En règle générale, la procédure d'enregistrement d'image se compose de quatre étapes de base: détection de fonction, fonction correspondants, conception de la fonction de cartographie, et l'image de transformation et de ré-échantillonnage. [15]



**Figure 3.2.** Décalage entre les deux caméras [15]

## **4. Fusion**

La fusion d'images, comme expliqué précédemment, est le processus qui combine des informations à partir de plusieurs images de la même scène. Ces images peuvent être capturées à partir de différents capteurs, acquises à des moments différents, ou ayant des caractéristiques spatiales et spectrales différentes. [1]

L'objet de la fusion d'images est de conserver les caractéristiques les plus désirables de chaque image. Avec la disponibilité de données de capteurs multiples dans de nombreux domaines, la fusion d'images a reçu une attention croissante dans les recherches pour un large éventail d'applications. [6]

La fusion d'images a pour but d'extraire les informations nécessaires à partir de deux images d'entrée. La résultante sera les images présentées complètes. Ceci est réalisé en faisant la moyenne des deux images. Dans le présent travail, en se concentrant sur DWT qui contient différents filtres haar, biorthogonal&Daubechies pour mesurer la qualité de l'image des performances. [6]

### **4.1. Fusion par transformée en ondelettes discrète (DWT)**

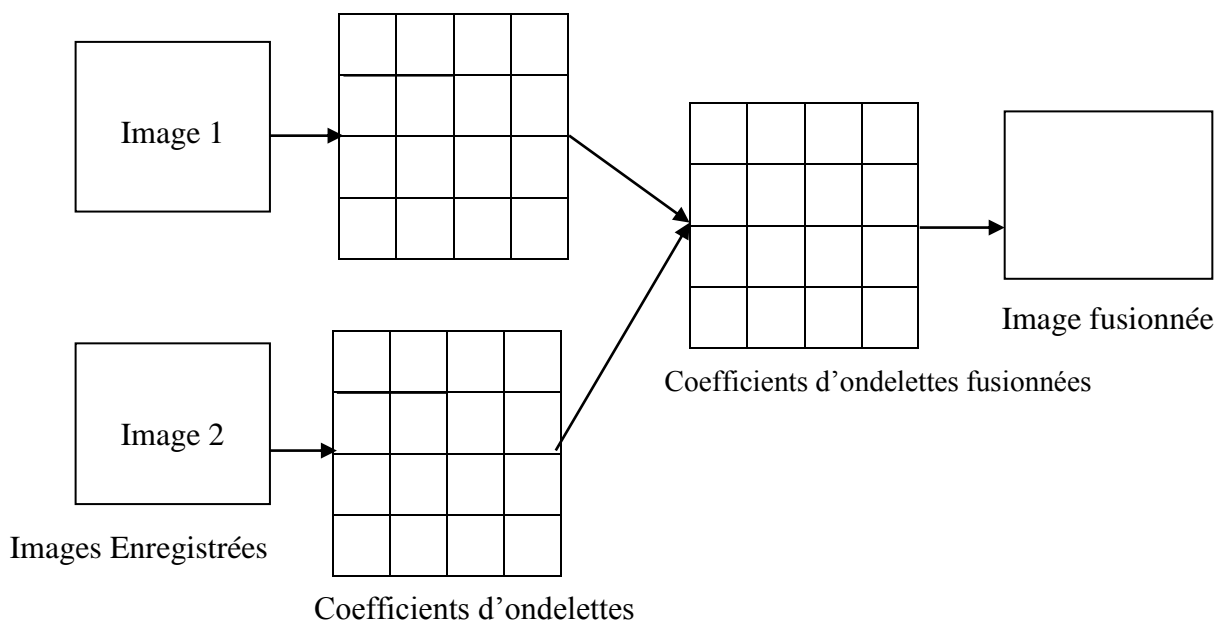
La transformée en ondelette DWT bidimensionnelle est devenu l'un des outils standard pour la fusion d'images dans le domaine de traitement d'image et du signal. Le processus de DWT est effectué par filtrage successifs passe-bas et passe-haut de l'image ou des images numériques. Ce processus est appelé l'algorithme de Mallat ou la décomposition Mallat-arbre. La signification principale de l'algorithme de Mallat est celle qui relie le temps continu multi-résolution de filtres en temps discret. [1]

L'idée de la fusion d'images par ondelettes est de fusionner les décompositions en ondelettes des deux images originales en appliquant des méthodes de fusion à des approximations et des coefficients. En observant la performance de toutes les techniques de fusion d'images, la DWT donne des résultats efficaces. [1]

L'exigence pour la fusion d'image est que les images doivent être correctement alignées sur une base pixel par pixel. Dans ce projet, les images pour être combinés sont supposées être déjà parfaitement enregistrées.[1]

La figure (3.3) montre le diagramme en bloc de niveau de fusion d'image en utilisant la transformée en ondelettes. Les images 1 et 2 deux images d'entrée qui sont capturées à partir de la caméra visible et infrarouge respectivement. La transformée en ondelettes décompose l'image en bandes bas-bas, bas-haut, haut-bas et haute-haute fréquence.

Les coefficients d'ondelettes sont générés par l'application de la transformée en ondelettes sur les images d'entrée. Les coefficients d'ondelettes des images d'entrée sont fusionnés en prenant la moyenne des images d'entrée. L'image résultante fondue est obtenue en appliquant la transformée en ondelettes inverse. [16]



**Figure 3.3.** Bloc diagramme d'une fusion à base de DWT [1]

#### 4.1.1. Principe de DWT :

La DWT d'un signal  $x$  est calculée en faisant passer à travers une série de filtres. D'abord, les échantillons sont passés à travers d'un filtre passe-bas à réponse impulsionnelle  $g$  entraîne une convolution des deux

$$y[n] = (x * g)[n] = \sum_{k=-\infty}^{\infty} x[k] g[n - k] \quad (3.1)$$

Le signal est également décomposé en utilisant simultanément un filtre passe-haut  $h$ . Les sorties donnant les coefficients de détail (du filtre passe-haut) et les coefficients d'approximation (du filtre passe-bas).

Il est important que les deux filtres sont liés les uns aux autres et ils sont connus en tant que filtre miroir en quadrature.

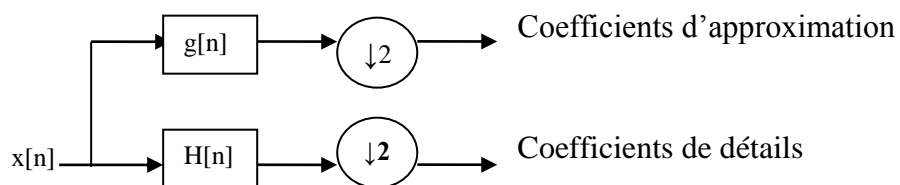
Cependant, depuis la moitié des fréquences du signal ont été retirés, la moitié des échantillons peuvent être jetées selon la règle de Nyquist. Les sorties des filtres sont ensuite sous-échantillonnées par 2

$$y_{low}[n] = \sum_{k=-\infty}^{\infty} x[k]g[2n - k] \quad (3.2)$$

$$y_{high}[n] = \sum_{k=-\infty}^{\infty} x[k]h[2n - k]$$

Cette décomposition a réduit de moitié la résolution de temps depuis que la moitié de chaque sortie de filtre caractérise le signal.[16]

Cependant, chaque sortie a la moitié de la bande de fréquence de l'entrée de sorte que la résolution de fréquence a été doublée.



Avec  $\downarrow$  opérateur de sous-échantillonnage.

$$(y \downarrow k)[n] = y[kn] \quad (3.3)$$

la somme ci-dessus peut être écrite de façon plus concise :

$$y_{low} = (x * g) \downarrow 2 \quad (3.4)$$

$$y_{high} = (x * h) \downarrow 2$$

Cependant le calcul d'une complète convolution  $x * g$  avec sous-échantillonnage ultérieur serait perdre du temps de calcul. Le système de levage est une optimisation où ces deux calculs sont fusionnés. [16]

#### **4.1.2. Détail et principe de base de DWT**

DWT- IDWT joue un rôle vital dans la technique de fusion d'images. Dans ce mémoire, les images enregistrées ont été envisagées pour le procédé de fusion, puisqu'il est nécessaire que les images doivent être correctement alignées sur une base pixel par pixel, pour obtenir avec succès la fusion de l'image. DWT a été modélisé à l'émetteur et IDWT a été modélisé au niveau du récepteur. [17]

Les transformées en ondelettes des images ont été calculé. Les images enregistrées ont été transmises comme signaux d'entrée à travers deux unidimensionnels filtres numériques différents H0 et H1 respectivement. Les filtres numériques H0 et H1 qui sont effectués successivement passe-haut et passe-bas filtrant les opérations pour les deux images d'entrée. La sortie de chacun des 2 filtres est suivie par un sous-échantillonnage. [17]

Cette étape est désignée comme la compression de ligne et le résultat est appelé en tant que composante de basse fréquence L et la composante à haute fréquence H. Les sorties de bas échantillonné ont encore été passé à l'un des deux filtres numériques afin d'obtenir une compression de la colonne. [17]

Le HH haut-haut, HL haut-bas, LH-Bas-Haut et LL Bas-Bas sont les composantes de fréquence de sortie obtenus après deux compressions au niveau des deux images d'entrée. La figure 3.4 montre le schéma bloc de fusion d'image à base de DWT qui se compose de deux images d'entrée, blocs de DWT, blocs de fusion et de blocs de IDWT.

Les HH, HL, LH et LL composantes de fréquence d'une image d'entrée est fusionné avec les composantes HH, HL, LH et LL de seconde image respectivement. HH composantes des deux images ont été ajoutés, puis la sortie résultante a été divisée par un facteur 2. De même, la moyenne des composantes HL, LH et LL ont été prises. Ce processus est connu comme la fusion d'images. Ce résultat a été en moyenne suivi par le processus de reconstruction c'est-à-dire transformée en ondelettes inverse. [17]

IDWT est le processus inverse de DWT. En cours IDWT, les composantes HH, HL, LH et LL ont d'abord été échantillonné exploité puis filtré. Les sous-bandes ont été ajoutées ou additionnées pour obtenir l'image reconstruite résultante. La technique de fusion d'images de



DWT produit l'image naturelle fusionnée même lorsque les images qui pourront être combinés ont été prises à partir de différentes caméras. [17]

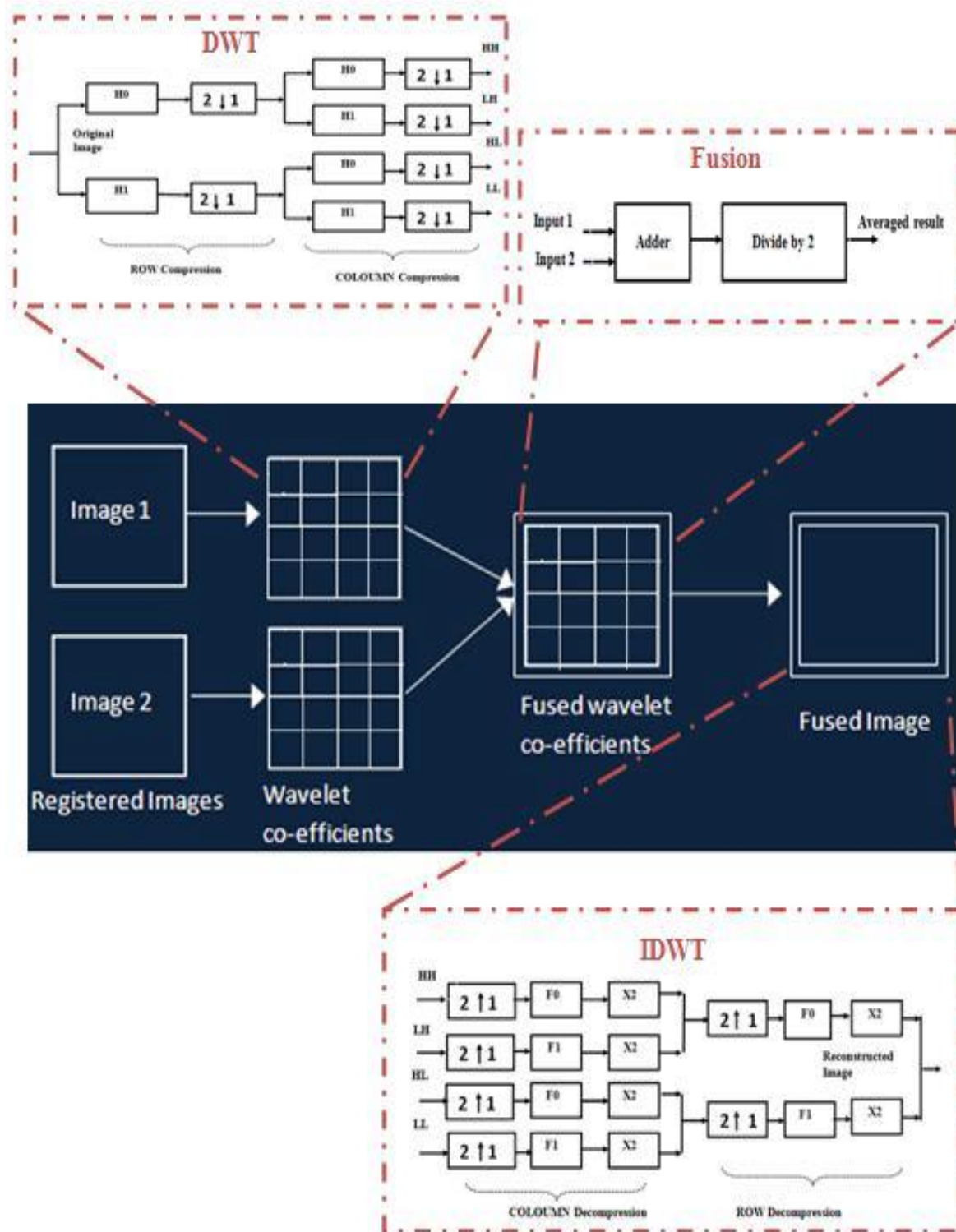


Figure 3.4. Schéma bloc de fusion d'image à base de DWT [17]

## 5. Détection de mouvement

La détection d'objet en mouvement en temps réel est une tâche difficile dans les systèmes de surveillance. Il s'agit souvent une première étape pour un traitement ultérieur comme la classification de l'objet mobile détecté. Afin d'effectuer des opérations plus sophistiquées telles que la classification, nous devons d'abord développer un procédé efficace et précis de détection d'objets en mouvement.

Un algorithme classique de détection d'objet en mouvement a les caractéristiques suivantes: (a) estimation de la partie fixe (fond) (b) obtention des images de différentes trames (c) la valeur de seuil appliqué (d) suivi (e) la vitesse d'estimation.

### 5.1. Soustraction de l'image de fond

La soustraction de fond est particulièrement une technique couramment utilisée pour la segmentation de mouvement dans les scènes statiques.

Cette méthode est très populaire et elle est utilisée par plusieurs applications. Cette façon détecte la région de mouvement en soustrayant pixel par pixel l'image courante à l'image de fond. [18]

$$X_t(x, y) = \text{Seuil}(|I_t(x, y) - b(x, y)|)$$

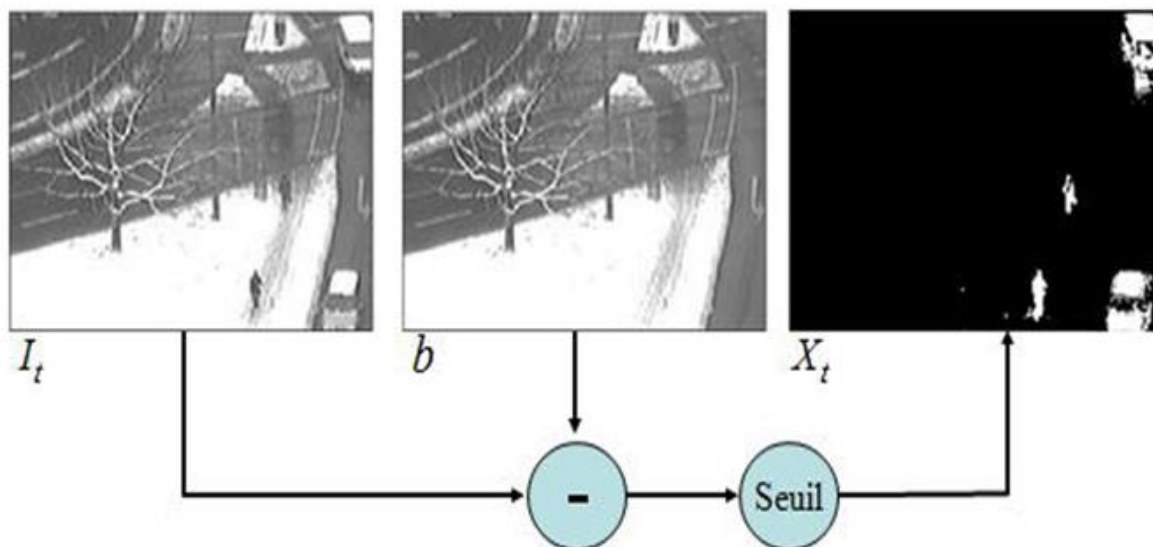
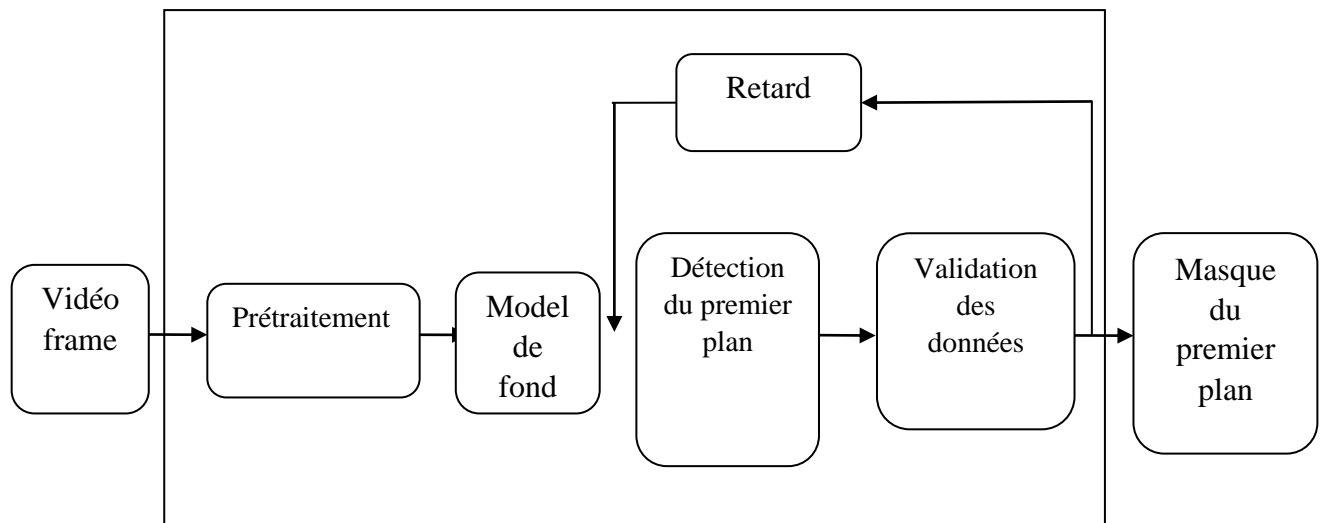


Figure 3.5. Soustraction de l'image de fond



**Figure 3.6.** Schéma de principe d'un algorithme de soustraction de fond. [19]

## 5.2. Soustraction et moyenne d'images

La soustraction d'images est utilisée dans certaines applications pour mettre en évidence certains détails caractéristiques. Ainsi, en radiographie, on peut être intéressé à mettre en évidence certains organes par utilisation de matière colorante: on effectue une prise d'image aux rayons X d'un sujet  $I_1(x, y)$  puis après injection de matière colorante, on prend des images TV du même sujet :  $I_2(x, y)$  On effectue alors la différence.

$$g(x, y) = I_1(x, y) - I_2(x, y) \quad (3.5)$$

Qui permet de donner une visualisation "dynamique" des détails intéressants. Dans un autre ordre d'idée, on peut effectuer des sommations d'images. Considérons une image

$I_1(x, y)$  entachée d'un bruit  $b(x)$ . Le résultat est une image bruitée

$$g(x, y) = I(x, y) + b(x, y). \quad (3.6)$$

Imaginons alors que l'on prenne M images du même sujet. Pour chaque image de numéro k, on aura

$$g_k(x, y) = I(x, y) + b_k(x, y) \quad (3.7)$$

La valeur moyenne (notée ici  $\langle \dots \rangle$ ) des résultats est :

$$\langle g(x, y) \rangle = \frac{1}{M} \sum_{k=1}^M g_k(x, y) = I(x, y) + \frac{1}{M} \sum_{k=1}^M b_k(x, y) \quad (3.8)$$

On peut recommencer l'opération un certain nombre de fois et considérer la distribution statistique des valeurs  $\langle g(x, y) \rangle$  obtenues. Elle est caractérisée par son espérance mathématique  $\langle\langle g \rangle\rangle$  et son écart-type ( $\delta \langle g \rangle$ ). En considérant légitimement que le bruit a une valeur moyenne nulle, on obtient les résultats suivants :

$$\langle\langle g \rangle\rangle = I(x, y) \text{ et } \delta \langle g \rangle = \frac{\sigma \langle b \rangle}{\sqrt{M}} \quad (3.9)$$

Ces résultats indiquent que la moyenne des images bruitées  $g_k(x, y)$  tend vers l'image "propre"  $I(x, y)$  quand  $M$  est assez grand. [18]

### 5.3. Algorithme

Soit  $I$  image à l'instant  $t$ .  $B$  est l'image de fond à l'instant  $t$ . L'objet du mouvement se compose des pixels qui satisfont l'équation suivante :

$$\max |I(t)(x, y) * C - B(t)(x, y) * C|, C = R, G, B \geq S(t)(x, y) \quad (3.10)$$

Où  $S(x, y)$  est un seuil de la position  $(x, y)$  à l'instant  $t$ . Cette méthode nous permet d'obtenir la forme complète d'objet et compter exactement les objets du mouvement parce que l'image de fond n'a pas l'objet mouvant.

Cependant, en réel, le changement de l'espace a lieu souvent (Spécialement, c'est le changement de la lumière : le matin, le soir, il y a du soleil, il pleut etc.). Cela influence le résultat si l'image de fond n'est pas mise à jour. Ou, dans un autre cas, l'objet stoppe dans une durée de temps long.

On a besoin de mettre à jours cet objet à l'image de fond. Evidemment, on a besoin d'un seuil pour déterminer qu'un objet va s'ajouter à l'image de fond s'il a stoppé dans  $N$  trames. Alors, on doit mettre à jours l'image de fond. Selon Yiğithan Dedeoğlu.[11]. Il y a une façon

simple pour mettre à jours l'image de fond et la matrice de seuil: soit O l'ensemble de l'objet mouvant.

$$B_{t+1}(x, y) = \begin{cases} \alpha B_t(x, y) + (1 - \alpha) I_t(x, y), & (x, y) \in B \\ \beta I_t(x, y) + (1 - \beta) I_t(x, y), & (x, y) \in O \end{cases} \quad (3.11)$$

$$B_{t+1}(x, y) = \begin{cases} \alpha S_t(x, y) + (1 - \alpha)(\gamma x |I_t(x, y) - B_t(x, y)|), & (x, y) \in B \\ S_t(x, y), & (x, y) \in O \end{cases} \quad (3.12)$$

Où  $\alpha, \beta \in (0, 1)$  sont les constantes et sont décidés par la pratique.  $\gamma$  est le taux d'influence de la différence entre l'image et le fond.

Un problème, comment on peut choisir  $\alpha$ , et  $\gamma$  pour tous les cas? C'est très difficile. Ainsi, on utilise une méthode qui est plus simple que la méthode de Yiğithan Dedeoğlu. On met à jour seulement l'image de fond de la manière suivante:[19]

$$B_{t+1}(x, y) = \begin{cases} B_t(x, y), & (x, y) \in O \\ I_t(x, y), & (x, y) \in B \end{cases} \quad (3.13)$$

## 6. Suivi de mouvement

Lors de l'exécution de détection d'objet et de suivi en vidéo, une étape clé est de mettre en œuvre un filtre de suivi. Les données d'une caméra de surveillance est généralement bruyantes et de mauvaise qualité, ce qui implique qu'il peut y avoir des incertitudes associées aux mesures des objets.

Un filtre devrait prendre ces incertitudes de mesure en considération et les combiner avec des prédictions de la façon dont les objets se déplacent, pour arriver à une piste optimale.

Le type de filtre mis en œuvre dépend de l'application. Plus précisément, cela dépend de comment nous modélisons le bruit qui est prévu, la relation entre les mesures et les états qui sont suivis ainsi que la prédiction de la manière dont les états changent au fil du temps (le modèle de prédiction).

Si le bruit est modélisé comme additif gaussien de moyenne nulle, et la mesure et la prévision sont linéaires, nous pouvons mettre en place un filtre de Kalman. [20]

### 6.1. Le filtre de Kalman

Le filtre de Kalman est une méthode de prédiction souvent utilisée pour le suivi spatio-temporel des objets. Il permet également de calculer le niveau d'incertitude de la prédiction. On doit toutefois supposer que la distribution statistique des paramètres d'état est uni-modale. Les occultations, le mouvement en arrière-plan et la présence d'une dynamique complexe de mouvements sont des exemples de distribution statistique multimodale qui nécessite l'utilisation du filtre généralisé de Kalman. Pour le présent projet, certaines simplifications permettent l'utilisation du filtre de Kalman comme si la distribution des paramètres d'état était uni-modale. [20]

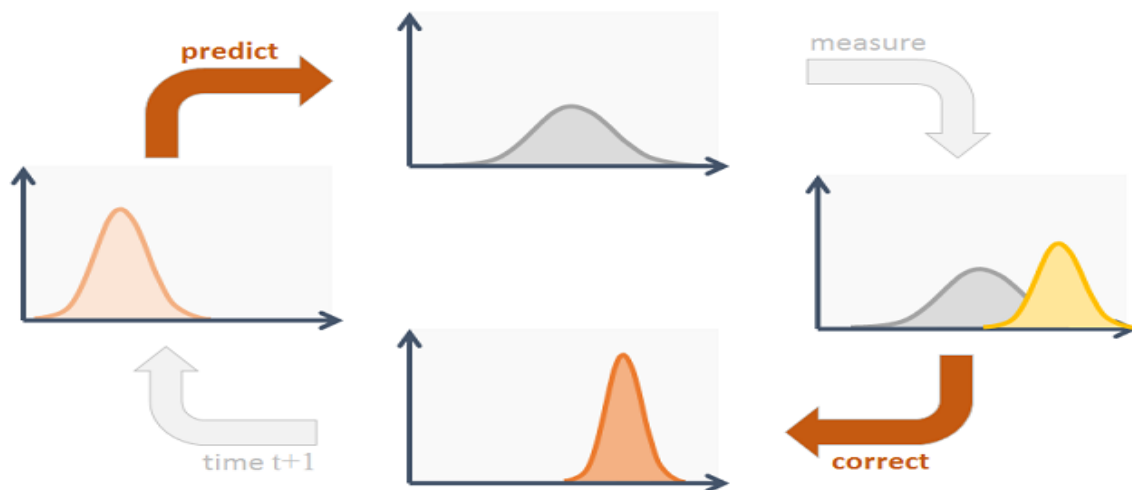


Figure 3.7. Cycle de filtre de kalman

### 6.2. Fonctionnement

Le principe d'un filtre de Kalman est comme suit. Tout d'abord, les équations d'état et les équations liant l'état précédent à l'instant suivant qui étaient linéaires dans le cas du filtre de Kalman classique sont maintenant non linéaires. Il est donc impossible de l'écrire sous forme matricielle. On remplace donc les équations [20]

$$\begin{cases} Y = H \cdot X + B \\ \hat{X}_k^+ = A \cdot \hat{X}_k \end{cases} \text{ par } \begin{cases} Y = h(X, B) \\ \hat{X}_k^+ = f(\hat{X}_k) \end{cases} \quad (3.14)$$

On est donc obligé d'appliquer ces équations non linéaires pour le calcul de la prédiction et la mise à jour du vecteur d'état. Les équations de Kalman deviennent donc :

La phase de prédiction

$$\begin{aligned}\hat{X}_k^+ &= f(\hat{X}_k) \\ P_k^+ &= A \cdot P_k \cdot A^T + Q\end{aligned}$$

La phase de mise à jour

$$\begin{aligned}K_{k+1} &= P_k^+ \cdot H_{k+1}^T \cdot (R_{k+1} + H_{k+1} \cdot P_k^+ \cdot H_{k+1}^T)^{-1} \\ P_{k+1} &= (I - K_{k+1} \cdot H_{k+1}) \cdot P_k^+ \\ \hat{X}_{k+1}^+ &= \hat{X}_k^+ + K_{k+1} \cdot (y_{k+1} - h(\hat{X}_k^+, 0))\end{aligned}$$

On remarque que pour mettre à jour le vecteur d'état, on utilise l'équation d'état non linéaire avec un bruit nul (second paramètre de la fonction  $h$  à 0).

Il reste un problème en ce qui concerne le calcul de la covariance de l'erreur et du gain de Kalman. En effet, on utilise toujours les matrices d'observation et de transition  $H$  et  $A$ . Pour pouvoir utiliser ces formules, il faut donc que l'on linéarise localement les fonctions  $H$  et  $F$ . On obtient donc les matrices d'observation et de transition en prenant les matrices des dérivés partielles des équations non linéaires

$$H_k = \left. \frac{\partial h}{\partial X} \right|_{\hat{X}_k^+} \tag{3.15}$$

$$A_k = \left. \frac{\partial f}{\partial X} \right|_{\hat{X}_{k-1}}$$

Avec ces Jacobiennes, il est donc possible d'appliquer le filtre de Kalman tel que défini ci-dessus. Il suffit de recalculer les matrices aux dérivées partielles à chaque nouvel échantillon à traiter et d'utiliser ces matrices dans les équations. [20]

Par contre, on se rend bien compte que l'on linéarise localement les équations afin d'appliquer le filtre de Kalman. Cette linéarisation est locale, ce qui entraîne donc une convergence locale

du filtre de Kalman étendu. Ce filtre ne garantis donc pas une convergence globale (à l'inverse du filtre de Kalman classique). La stabilité d'un KF est donc plus difficile à garantir et dépend souvent de sa bonne initialisation. [20]

### 6.3. Algorithme

Le filtre de Kalman estime le processus sous la forme d'une boucle de rétroaction récursive (tableau 3.1). Connaissant l'état au moment temporel  $k$ , le filtre estime l'état  $k + 1$  du processus, ce qui donne l'estimation a priori  $\hat{x}_{k+1}$ . La rétroaction est ensuite assurée par la mesure (bruitée.). [20]

**TAB.3.1.** Algorithme du filtre de kalman

Prédiction	Correction
$\hat{x}_k^- = A\hat{x}_{k-1} + Bu_{k-1}$ $P_k^- = AP_{k-1}A^T + Q$	$K_k = P_k^- H^T (HP_k^- H^T + R)^{-1}$ $\hat{x}_k = \hat{x}_k^- + K_k(z_k - H\hat{x}_k^-)$ $P_k = (1 - K_k H)P_k^-$

### 6.4. Principe et points forts

Le filtre de Kalman est une méthode visant à estimer des paramètres d'un système évoluant dans le temps à partir de mesures bruitées.

On retrouve ce filtre dans bon nombre de domaines relatifs au traitement du signal, radar, traitement d'images etc.

La force de ce filtre est sa capacité de prédiction des paramètres et de rectification des erreurs, non seulement des capteurs mais aussi du model lui-même.

En effet, pour appliquer un filtre de Kalman, il faut avant tout modéliser le système pour lequel on veut estimer les paramètres, de manière linéaire. (Des variantes du filtre de Kalman existent pour la prise en compte du modèle non linéaire).

Dans une méthode d'estimation classique (par exemple, la méthode des moindres carrés), une simple erreur dans la modélisation du système entraine inévitablement une erreur au niveau de l'estimation. La force du filtre de Kalman est d'intégrer un terme d'imprécision sur le modèle



lui-même, ce qui lui permet de donner des estimations correctes malgré les erreurs de modélisation (pour peu que les erreurs restent raisonnables).

Un autre point fort du filtre de Kalman (mais que l'on retrouve aussi dans la méthode des moindres carrés par exemple) est sa capacité à déterminer l'erreur moyenne de son estimation. En effet, l'outil Kalman fournit un vecteur contenant les paramètres estimés, mais aussi une matrice de covariance de l'erreur. Cette matrice nous renseigne donc sur la précision de l'estimation, ce qui peut être utile dans de nombreuses applications. Un autre atout du filtre de Kalman est que la convergence de cette erreur est garantie. [21]

Le fonctionnement du filtre de Kalman peut se diviser en deux étapes

Une première étape de prédiction de l'estimation selon le modèle du système. Pour ce faire, le filtre de Kalman reprend l'estimation précédente des paramètres et de l'erreur et prédit les nouveaux paramètres et la nouvelle erreur en fonction de la modélisation du système.

La seconde étape va faire la mise à jour de cette prédiction grâce aux nouvelles mesures. Ces mesures (par définition bruitées) vont permettre d'obtenir une estimation des paramètres et de l'erreur à partir de la prédiction faite. Si jamais le modèle comporte des erreurs, cette étape de mise à jour permettra de les rectifier. [21]

#### 6.5. Paramètres du filtre

Les matrices de covariance  $R$  et  $P$  correspondent respectivement au bruit du processus et au bruit de la mesure. Le bruit de la mesure est calculé à partir d'échantillons connus alors pour le bruit du processus, la source de bruit est souvent l'incertitude sur le modèle choisi. Si  $R_k$  et  $Q_k$  sont constantes, l'estimation de la covariance de l'erreur  $P_k$  et du facteur  $k_k$  auront une stabilisation rapide. Dans ce cas, tous les paramètres du filtre peuvent être calculés. Les paramètres  $R$  et  $P$  doivent être ajustés pour obtenir un filtre performant. [21]

#### 7. Conclusion

Nous avons étudié dans ce chapitre le principe de fusion d'image par la méthode de DWT qui est connue dans la littérature par ses performances dans la fusion des données. Nous avons aussi présenté une méthode de suivi, qui est la soustraction de fond, qui est la méthode la plus

triviale et robuste. Finalement, nous avons donné le formalise mathématique du filtre de Kalman afin de l'utiliser pour suivre le mouvement de l'objet détecté.

## *IV. Résultats et Discussions*

## **1. Introduction**

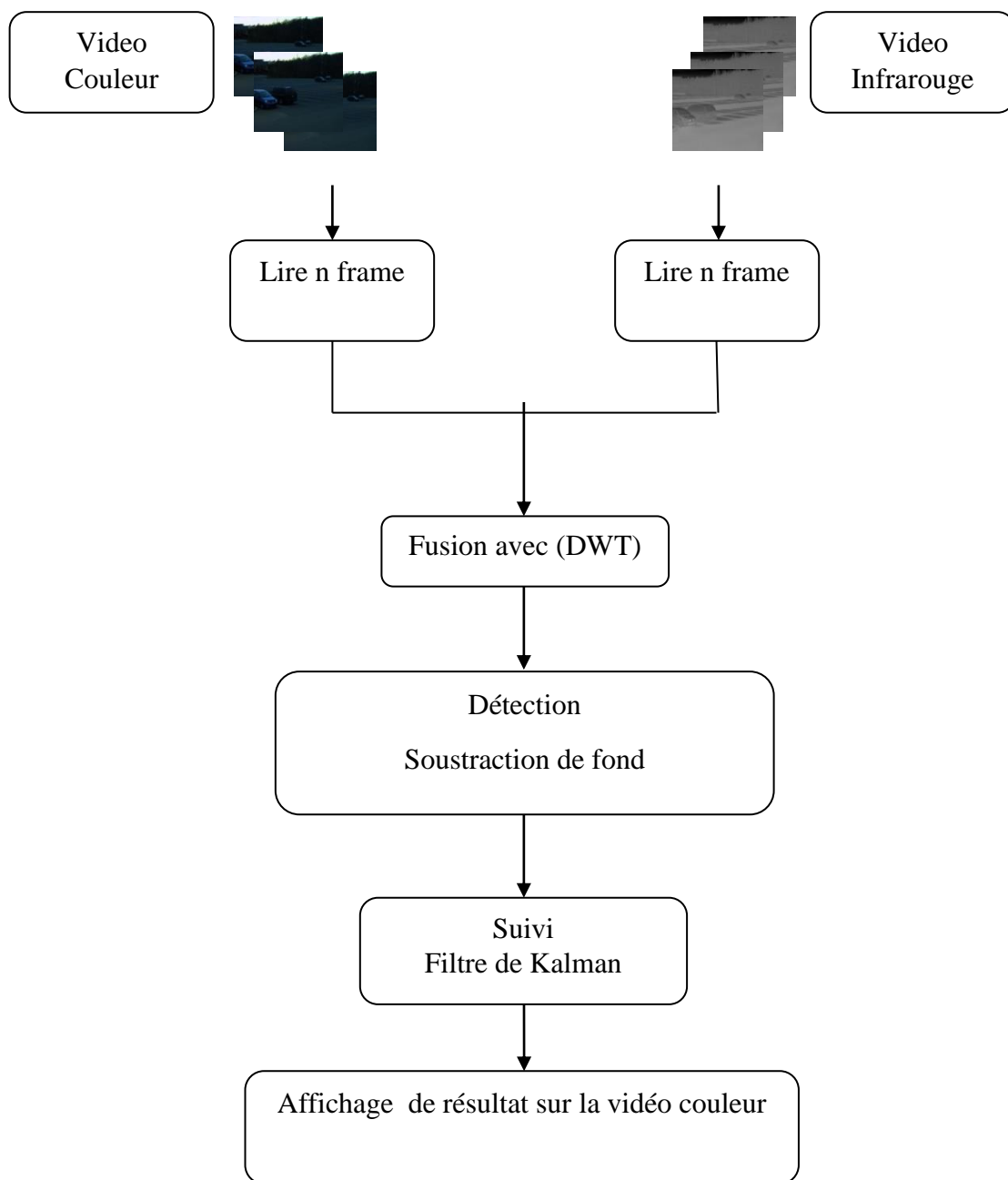
Dans ce chapitre, les résultats des méthodes de fusion, détection et suivi seront présentés. Nous décrivons toutes les mesures d'évaluation et les algorithmes de fusion détection et suivi. La solution proposée est implémentée sur Matlab.

## **2. L'implémentation de la solution proposée**

L'algorithme suivi dans ce travail est présenté dans la figure 4.1 ci-dessous. Il décrit les étapes générales qui sont utilisés pour obtenir une solution pour notre problématique, qui est la détection et le suivi des objets dans le cadre de vidéosurveillance intelligente.

La première étape de notre travail consiste à acquérir deux vidéos, une avec une caméra couleur et l'autre avec une caméra infrarouge ensuite pour chaque frame des deux vidéos, on effectue le traitement suivant :

La fusion des deux images à l'aide de DWT. En utilisant l'image fusionnée précédente, la détection de la personne en mouvement dans la scène est effectuée. La dernière étape consiste à suivre cette personne, en utilisant le filtre de Kalman.



**Figure 4.1.** Organigramme de la présente étude

### 3. L'acquisition

La première étape est de filmer la scène en question de deux caméras, une est couleur qui capture les images avec des couleurs et avec une résolution d'image de haute qualité, l'autre est infrarouge qui nous donne des informations avec des mesures thermique. Cette

dernière convient dans tous les temps avec une bonne résolution et prend également plus d'informations de l'objet ou de la cible par rapport à la caméra CCD.

Dans le cas pratique, et comme mentionné dans le chapitre précédent, les vidéos doivent être enregistrées, *i.e.*, subissent des opérations de mise à l'échelle et mise en correspondance (pour ne prendre en considération que l'information commune).

Dans notre cas, vu que les caméras ne sont pas à notre disponibilité, nous avons utilisé des bases de données, qui sont généralement utilisées par la communauté scientifique de fusion et suivi.

#### **4. Base de données utilisée**

Dans notre étude, nous avons utilisé deux bases de données avec deux scénarios pour chacun afin de mieux montrer l'efficacité de notre travail.

##### **4.1. Base de données 1 :**

Cet ensemble de données est utilisé pour l'évaluation des algorithmes de fusion, et porte sur le thème suivant :

La détection d'objet en fonction Fusion-couleur et de l'imagerie thermique,[22] "Background-Soustraction utilisant une fusion d'imagerie thermique et visible base-Contour," Computer Vision and Image Understanding, vol 106, n ° 2-3, 2007, pp. 162-182. James W. Davis, jwdavis [at] cse.ohio-state.edu

Les deux capteurs thermique et couleurs ont les détails suivant

Capteur thermique: Raytheon Palmir 250D, objectif 25 mm

Capteur de couleur: Sony TRV87 Handycam

Nous avons choisis pour notre travail deux vidéos dont le scénario est filmé dans l'intersection de la voie Occupé sur le campus de l'Université d'État de l'Ohio elle a un nombre d'image de 17089 qui ont le format : 8 bits en niveaux de gris bitmap, Couleur: couleur 24 bits bitmap et résolution d'image de 320 x 240 pixels, et un taux d'échantillonnage d'environ 30Hz

**4.2. Base de données 02 :** Les ensembles de données images /vidéos sont tirées de l'INO qui est une société de conception et de développement de technologie optique et des solutions photoniques pour les PME et les grandes entreprises, Ontario CANADA.

## **5. Résultats de Fusion**

Dans les chapitres précédents, nous avons présenté les techniques de fusion. Dans notre travail, nous avons testé quelques-unes.

L'évaluation de la performance est une partie essentielle dans tout système de fusion. Elle est divisée en deux catégories principales : l'évaluation qualitative et l'évaluation quantitative.

Les anciens jugent l'exécution de l'algorithme de fusion en inspectant visuellement l'image fusionnée et en la comparant à des images sources. Bien que cette méthode simple ne nécessite aucun traitement, elle repose toujours sur une inspection de l'opérateur humain qui ne parvient pas à évaluer l'exécution de l'algorithme de fusion ou de comparer plusieurs schémas de fusion.

D'autre part, L'évaluation quantitative est en mesure de comparer les différents algorithmes sur la base de mesures quantitatives et sans aucune intervention humaine, ce qui est n'est pas approprié ni pour un fonctionnement en temps réel, ni pour un système de surveillance d'urgence dans lequel les décisions doivent être entièrement autonome.

### **5.1. Évaluation qualitative**

Comme indiqué précédemment, la qualité d'un système de fusion d'images peut être évaluée de façon qualitative en utilisant l'inspection visuelle humaine. En d'autres termes, les expériences sont menées sur plusieurs personnes pour déterminer la qualité de la fusion par l'application des tests tels que la détection d'une cible ou la détermination de l'orientation d'image.

Les tests peuvent être tenus de déterminer et d'identifier une cible sur le fond naturel ainsi que savoir si l'image est droite ou inversée. Le temps de réponse et les taux d'erreur sont comparés en regardant l'image fusionnée et en regardant l'image sources séparément. Les images fusionnées ayant des temps de réponse plus rapide et taux d'erreur plus faible indiquent un meilleur algorithme de fusion d'image.

## **5.2. Evaluation quantitative**

Les mesures qualitatives peuvent être divisés en trois catégories: les caractéristiques de l'image fusionnées, relation entre image fusionnée et images sources, et la relation entre l'image fusionnée et une image parfaitement fusionné.

Les mesures de performance sont jugées importantes pour déterminer les aspects bénéfiques de la transformée en ondelettes et aussi de comparer les résultats obtenus pour des images différentes. Ce chapitre vise à souligner que DWT donne de meilleurs résultats par rapport aux autres techniques testées (Laplacienne, gradient et PCA). Les mesures quantitatives qu'on a utilisées à savoir :

### **5.2.1. Entropie (EN)**

Les performances d'un algorithme de fusion peuvent être évaluées par l'évaluation de la quantité d'informations contenues dans l'image fusionnée ou le degré de clarté de l'image fusionnée. A cet effet, l'entropie de l'image pourrait être utilisée pour mesurer la quantité d'information à l'aide de l'équation suivante : [15]

$$En = \sum_{j=0}^G (P(i) \text{Log}(P(i)))$$

Avec:

G est le nombre de niveaux de gris d'image et P (i) est la probabilité normalisée d'apparition de chaque niveau de gris

### **5.2.2. Information mutuelle (MI : Mutual Information)**

La relation entre l'image source et l'image fusionnée est capable de déterminer combien d'information est transférée à l'image fusionnée et la quantité d'informations perdue.

A cet effet, l'information mutuelle (MI) est utilisée pour mesurer le degré de similarité entre l'image fusionnée et chacune des images sources.

Donc pour mesurer la performance de fusion d'images, l'information mutuelle entre l'image fusionnée et l'image source est calculée, puis on considère la somme des deux mutuelles.

Plus le MI est grand plus la fusion est meilleure. Les équations suivantes représentent le processus.



$$MI = Mi1(fuse, ir) + Mi2(fuse, ir)$$

Avec :

$$Mi1(fuse, ir) = - \sum P(fuse, ir) \text{Log} \left\{ \frac{P(fuse, ir)}{P(fuse).P(ir)} \right\}$$

$$Mi2(fuse, vis) = - \sum P(fuse, vis) \text{Log} \left\{ \frac{P(fuse, vis)}{P(fuse).P(vis)} \right\}$$

Avec

P (fuse, ir) et P (fuse, vis) sont les histogrammes conjointes de l'image source Iir, Ivis et l'image fusionnée ifuse . [15]

### 5.2.3. Erreur quadratique moyenne (MSE : Mean Square Error)

MSE est une mesure fréquemment utilisée pour trouver la différence entre les valeurs de l'entrée et la sortie reconstruit.

L'équation mathématique de MSE est donnée par :

$$MSE = \frac{1}{m.n} \sum_{i=1}^m \sum_{j=1}^n (A_{ij} - B_{ij})^2$$

Où m.n correspond au pas de pixels dans l'image.

A<sub>ij</sub>- valeur de pixel à l'emplacement (i, j) de l'image d'entrée.

B<sub>ij</sub>- valeur de pixel à l'emplacement (i, j) de l'image de sortie.

### 5.2.4. Rapport de signal crête sur bruit ( PSNR :Peak Signal to Noise Ratio )

PSNR est la mesure la plus utilisée pour juger de la qualité d'image reconstruite. Le signal est dans ce cas les données d'origine et le bruit est l'erreur introduite dans les valeurs de gris de l'image pendant le traitement. PSNR est utilisé comme une approximation de la perception humaine sur la qualité de la reconstruction.

Un PSNR plus élevé serait normalement indiquer que la reconstruction est de meilleure qualité. Sa formule mathématique est donnée par :

$$PSNR = 20 \log_{10} \frac{L^2}{\frac{1}{m \cdot n} \sum_{i=1}^m \sum_{j=1}^n (I_r(i, j) - I_f(i, j))^2}$$

Où  $L$  est le nombre de niveau de gris de l'image

### 5.2.5. Contenu structurelle (SC)

La mesure de contenu structurel est utilisée pour comparer deux images, dans un certain nombre de patches de petites image, le contenu structurel représente les patches en commun.

$$SC = \frac{\sum_{i=1}^m \sum_{j=1}^n (A_{ij})^2}{\sum_{i=1}^m \sum_{j=1}^n (B_{ij})^2}$$

### 5.2.6. Corrélation croisée normalisée (NCC : Normalized Cross Corrélation)

La corrélation est souvent utilisé pour rechercher des signaux similaires qui se répètent dans une série de temps - Ceci est connu comme le filtrage adapté. Parce que la corrélation de deux signaux de forte amplitude aura tendance à donner de grands nombres, on ne peut pas déterminer la similitude des deux signaux simplement en comparant l'amplitude de leur corrélation croisée.

$$NCC = \frac{\sum_{i=1}^m \sum_{j=1}^n (A_{ij} * B_{ij})}{\sum_{i=1}^m \sum_{j=1}^n (A_{ij})^2}$$

### 5.2.7. Erreur absolue normalisée (NAE : Normalized Absolute Error) :

Une valeur plus grande de l'erreur absolue normalisée, veut dire une image de mauvaise qualité le NAE est défini par l'équation suivante :

$$NAE = \frac{\sum_{i=1}^m \sum_{j=1}^n (|A_{ij} - B_{ij}|)}{\sum_{i=1}^m \sum_{j=1}^n (A_{ij})}$$

### **5.2.8. Gradient moyen (AVG : Average Gradient:)**

Il est donné par la formule :

$$\bar{g} = \frac{1}{N} \sum \sqrt{\frac{\Delta I_x^2 + \Delta I_y^2}{2}}$$

Où  $\Delta I_x, \Delta I_y$  sont les différences dans les directions x et y.

Une valeur plus grande de gradient moyen, est interprétée par une image plus nette.

### **5.3. Résultats**

Les résultats expérimentaux obtenus par fusion d'une image IR et une image CCD en utilisant des différentes méthodes sont présentés dans cette section. Trois séries d'images ont été sélectionnés comme entrées pour les algorithmes de fusion.

Les figures 4.2 à 4.5 montrent les résultats des techniques de fusion pour les quatre scénarisons dans la liste ci-dessus. Les techniques que nous avons implémentées sont la fusion avec le gradient, DWT, Pyramide Laplacienne et PCA.



a) Camera CCD image source



b) Camera IR image source



c) Résultat de fusion avec DWT



d) Résultat de fusion avec Gradient



e) Résultat de fusion avec Laplacienne



f) Résultat de fusion avec PCA

**Figure 4.2. Résultats de fusion par DWT, Gradient, laplacienne et PCA pour le scénario 1**



a) Camera CCD image source



b) Camera IR image source



c) Résultat de fusion avec gradient



d) Résultat de fusion avec DWT



e) Résultat de fusion avec Laplacienne



f) Résultat de fusion avec PCA

**Figure 4.3. Résultat de fusion avec Gradient, DWT, Laplacienne et PCA pour le scénario 2**



a) Camera CCD image source



b) Camera IR image source



d) Résultat de fusion avec DWT



e) Résultat de fusion avec Gradient



f) Résultat de fusion avec Laplacienne



g) Résultat de fusion avec PCA

**Figure 4.4. Résultats de fusion par DWT, Gradient, LAplacienne et PCA pour le scénario 3**



a) Camera CCD image source



b) Camera IR image source



c) Résultat de fusion avec DWT



d) Résultat de fusion avec PCA



e) Résultat de fusion avec Laplacienne



f) Résultat de fusion avec Gradient

**Figure 4.5. Résultats de fusion par DWT, PCA Laplacienne et Gradient pour le scénario 4**

Les résultats des tests montrent que le régime de fusion DWT visuellement meilleurs que les trois autres régimes. Il est des endroits plus visibles comme, par exemple, le brouillage (les bords de l'arrière-plan et de petits détails) et la perte de la texture dans l'image fusionnée obtenue par le rapport de pyramide et PCA. En outre, dans l'image fusionnée par la méthode PCA, quelques détails sur les images et le fond ont été complètement perdu, ainsi que dans l'image composite moyenne.

### **5.3.1. Interprétation des résultats**

Ces comparaisons visuelles subjectives coïncident avec l'évaluation métrique proposée, et qui est présentée dans le tableau 4.1.

Il est clair d'après les expériences que MI métrique et PSNR attribuent souvent la plus forte valeur de la mesure de la performance de fusion à l'algorithme DWT.

Les valeurs obtenues à partir des mesures proposées sont bien corrélées à la qualité subjective des images fusionnées, ce qui n'était pas possible par la mesure de la performance de fusion MI standard et PSNR. En outre, les mesures proposées ne sont pas dépendante, de façon significative, de la taille de la fenêtre d'analyse en tant que la différence de performance de fusion ne varie pas beaucoup avec la variation de la taille de la fenêtre.



**Tableau 4.1.** Evaluation quantitative pour les 4 scénarios

	méthode	EN	PSNR	MI	MSE	SC	NAE	Avg	ncc	Temps
1	dwt	<b>16.655</b>	<b>68.936</b>	<b>2.019</b>	<b>0.020</b>	<b>0.834</b>	<b>0.105</b>	<b>0.064</b>	<b>1.007</b>	<b>0.574</b>
	laplacienne	14.760	63.446	0.966	0.002	0.902	0.247	0.052	1.002	0.665
	gradient	15.069	66.116	1.018	0.015	1.012	0.187	0.039	0.965	0.726
	pca	13.057	63.025	1.531	0.324	1.639	0.308	0.030	0.753	1.787
2	dwt	<b>15.241</b>	<b>80.119</b>	<b>1.701</b>	<b>0.002</b>	<b>0.485</b>	<b>0.091</b>	<b>0.054</b>	<b>0.996</b>	<b>0.607</b>
	laplacienne	14.288	73.607	1.379	0.002	0.964	0.109	0.019	1.003	0.723
	gradient	14.233	74.347	1.431	0.001	1.049	0.105	0.015	0.963	0.763
	pca	12.110	63.046	0.866	0.029	0.697	0.533	0.010	1.060	1.622
3	dwt	<b>14.514</b>	<b>80.141</b>	<b>1.842</b>	<b>0.006</b>	<b>0.297</b>	<b>0.096</b>	<b>0.036</b>	<b>0.997</b>	<b>0.657</b>
	laplacienne	13.522	79.102	1.374	0.007	1.025	0.106	0.011	0.981	0.715
	gradient	13.419	75.865	1.147	0.001	1.182	0.157	0.008	0.910	0.757
	pca	8.419	59.467	1.450	0.073	0.408	1.643	0.016	1.186	1.101
4	dwt	<b>15.947</b>	<b>87.668</b>	<b>3.269</b>	<b>0.001</b>	<b>0.686</b>	<b>0.014</b>	<b>0.049</b>	<b>0.992</b>	<b>0.538</b>
	laplacienne	14.969	65.300	1.128	0.019	0.979	0.151	0.039	0.985	0.792
	gradient	14.974	67.338	1.138	0.012	1.026	0.132	0.031	0.971	0.671
	pca	13.803	84.144	2.574	0.002	0.960	0.023	0.033	1.020	0.975

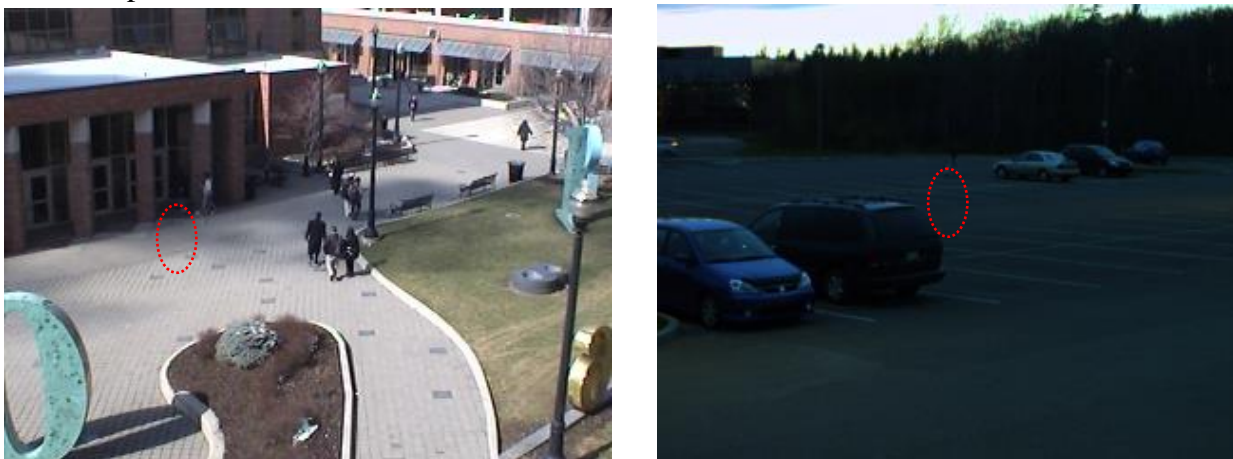
De ces résultats, il en résulte que le régime DWT de fusion semble donner la meilleure performance globale. Dans la littérature, il a été trouvé qu'un système de fusion en utilisant une transformée en ondelettes (comme la méthode de DWT) a des avantages sur la fusion comme l'augmentation directionnel d'informations, pas de blocage des artefacts qui se produisent souvent dans les images fusionnées et amélioration de la perception comparées en utilisant une analyse humaine .

#### 5.4. L'importance de la fusion

Nous allons montrer l'importance de la fusion, et la justification de cette étape. Pour le faire, nous considérons deux scénarios.

Si le système de vidéosurveillance ne dispose que de caméras couleurs, le résultat affiché sera comme le montre la figure 4.6.

On remarque que c'est impossible de détecter les objets dans les positions indiquées avec des cercles vu le manque de luminosité et aussi la couleur de l'objet est proche de celle de l'arrière-plan.



**Figure 4.6.** Difficulté de détection d'objet dans une scène filmée par une caméra couleur

Si on utilise des caméras infrarouges au lieu de caméras couleurs, les images obtenues sont présentées dans la figure 4.7.

Nous remarquons que la caméra infrarouge a pu parfaitement détecter les personnes qui n'étaient pas visibles dans la scène filmée par la caméra couleur.

Par contre, les images IR perdent les détails, tels que la texture.



**Figure 4.7.** Difficulté de détection d'objet dans une scène filmée par une caméra infrarouge

D'où l'idée de fusionner les deux images, de l'image couleur on récupère les détails, et de l'image infrarouge on récupère la détection. La figure 4.8. montre la fusion des deux, le résultat contient au même temps les détails (à part la couleur) et les objets à détecter



**Figure 4.8.** Possibilité de détection d'objet dans une scène filmée par le système fusionnée

### 6. détection par soustraction de fond

Pour pouvoir suivre les mouvements dans la vidéo, la première étape est de détecter les mouvements. Cette étape joue un rôle très important dans le système de vidéo surveillance car le résultat de cette étape va influencer toutes les étapes suivantes.

#### 6.1. Arrière-plan

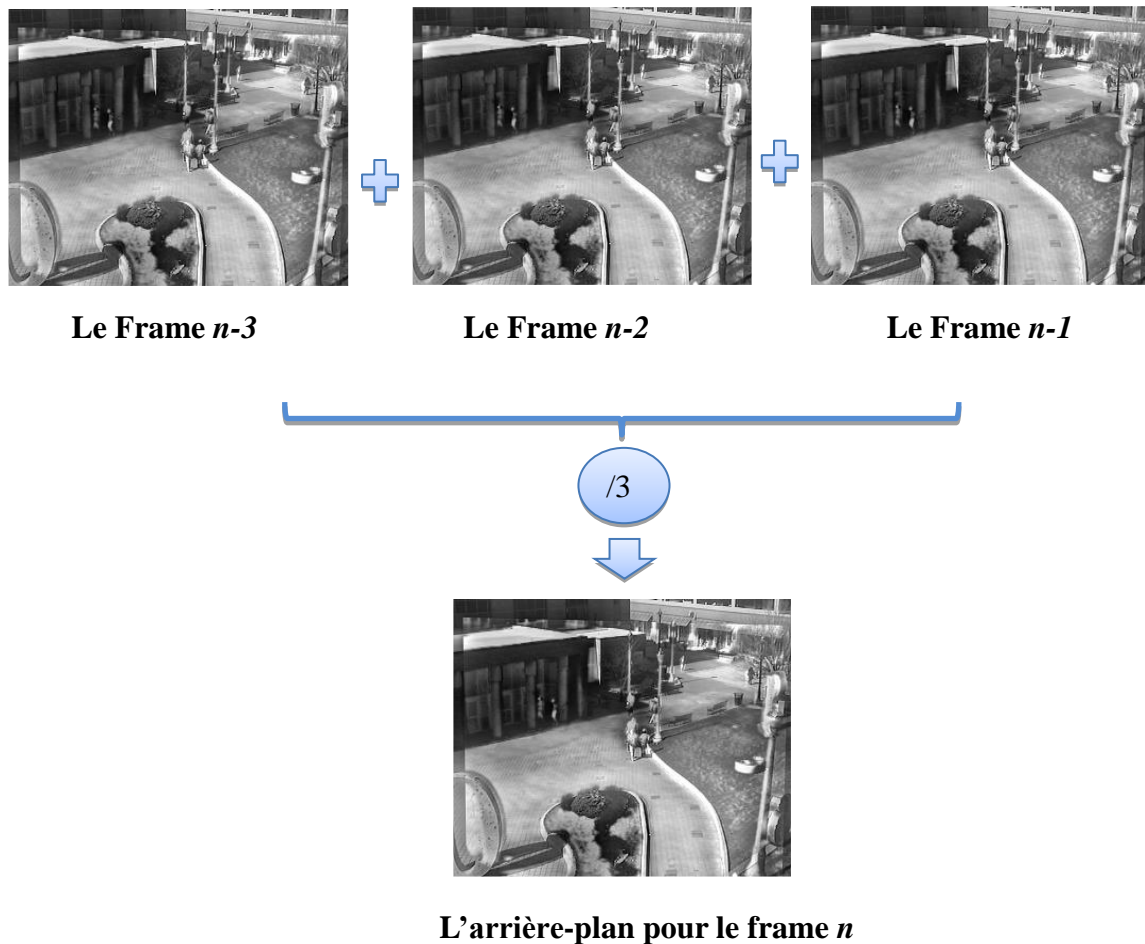
La méthode de soustraction de fond nécessite l'image arrière-plan. Il peut être effectué par plusieurs façons, mais ça dépend de l'application.

Par exemple, dans un scénario de surveillance d'une salle sécurisée, et qui ne doit pas contenir aucun mouvement pendant une durée précise. Dans ce cas, l'arrière-plan est choisi comme l'image de la scène sans objet indésirable.

Dans notre cas, la scène est supposée dynamique, c-à-dire, elle contient un ou plusieurs objets mobiles d'une façon aléatoire.

#### 6.2. Résultat de l'arrière-plan

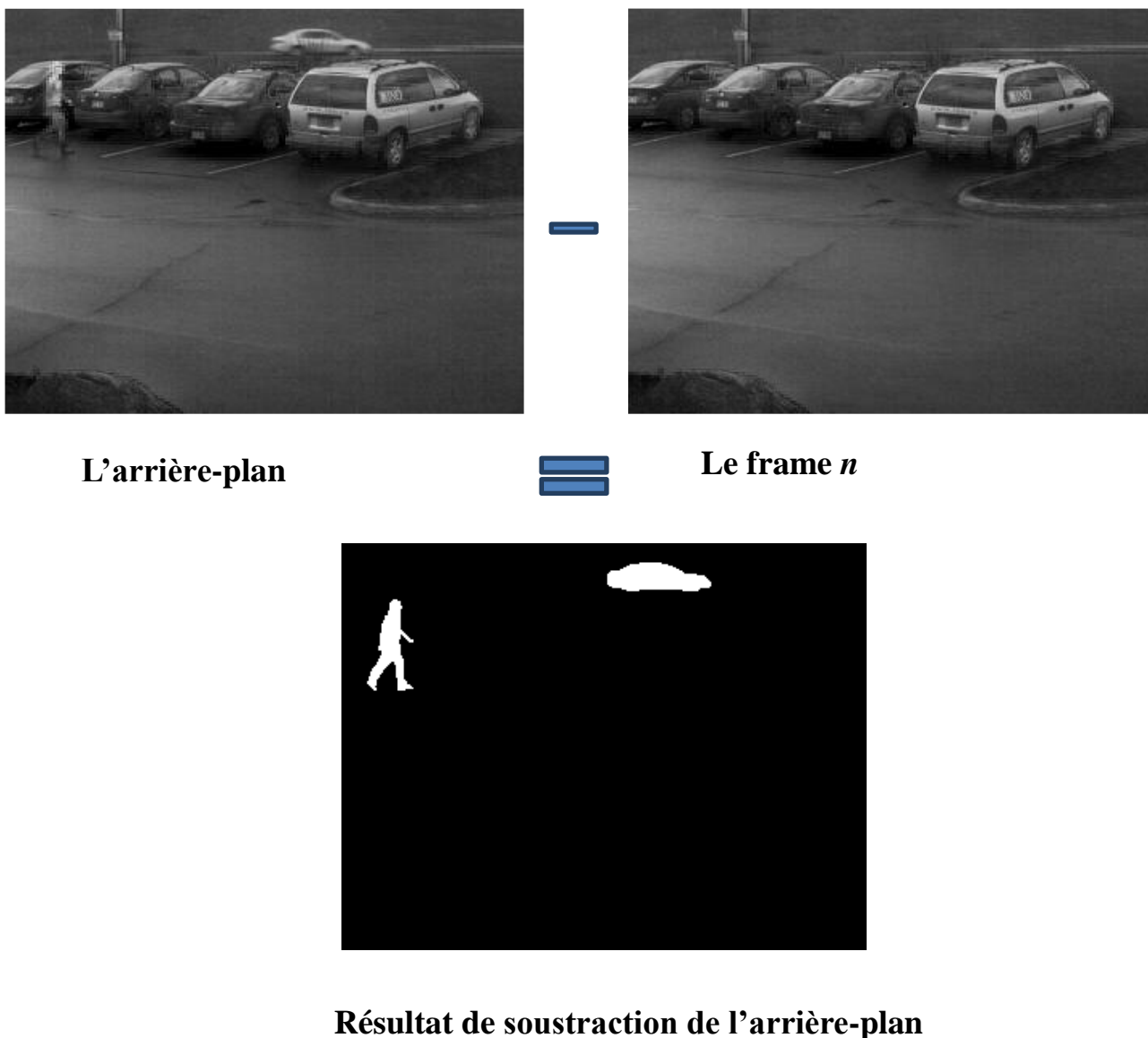
Pour chaque frame, On choisit, la moyenne des trois images qui le précède directement comme image de fond. [3]. Le résultat est montré dans la figure 4.9.



**Figure. 4.9.** Arrière-plan

### 6.3. Résultats de soustraction de fond :

L'image obtenue de l'étape précédente, est ensuite soustraite de l'image en question, pour donner le résultat de détection. Après avoir convertit l'image couleur au niveau de gris nous faisons une binarisation de l'image. Le résultat est une image noir et blanc. Cette binarisation est faite après avoir choisi un seuil.



**Figure 4.10.** Possibilité de détection d'objet dans une scène filmée par le système fusionnée

#### 6.4. Détection

La détection est un signal d'une présence d'un objet dans un espace bien défini. Pour garantir cette opération, il faut définir un seuil de détection.

Si la différence entre les deux images est inférieure au seuil, il n'y a pas d'objet détecté,

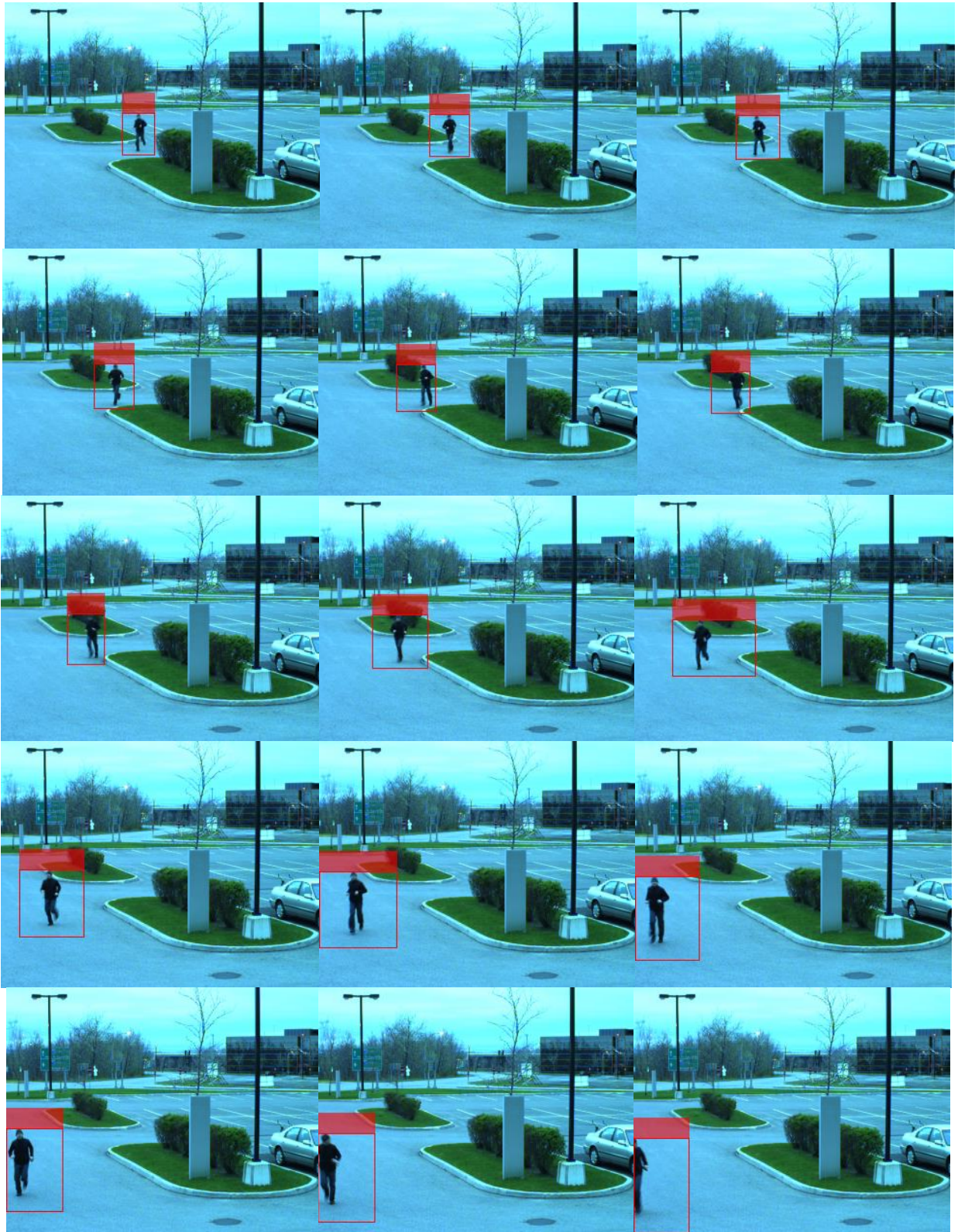
Si la différence est supérieure au seuil, ceci signifie la présence de l'objet.

## **7. suivi**

Dans notre travail, nous avons utilisé le filtre de Kalman pour suivre le mouvement détecté.

Nous appliquons le filtre sur trois scénarios différents.

La figure 4.11 montre quelques frames de la vidéo de suivi que nous avons obtenue. Nous remarquons que nous avons pu suivre la personne qui est en mouvement dans la scène surveillée.



**Figure 4.11.** Résultat de suivi dans un parking avec un seul objet mobile

Dans le cas où nous avons à traiter plusieurs objets mobiles distants l'un de l'autre, nous pouvons de même suivre leurs mouvements, comme le montre la figure 4.12.



Figure 4.12. Résultat de suivi dans un parking, avec deux objets mobiles



Sauf que dans certaines situations, notre système connaît des limites, comme par exemple dans le cas où les objets sont très proches l'un de l'autre, on les suit comme étant un seul mobile.

Mais une fois la distance entre eux augmente, le suivi devient parfait. Ceci est présenté par la figure 4.13 ci-dessous

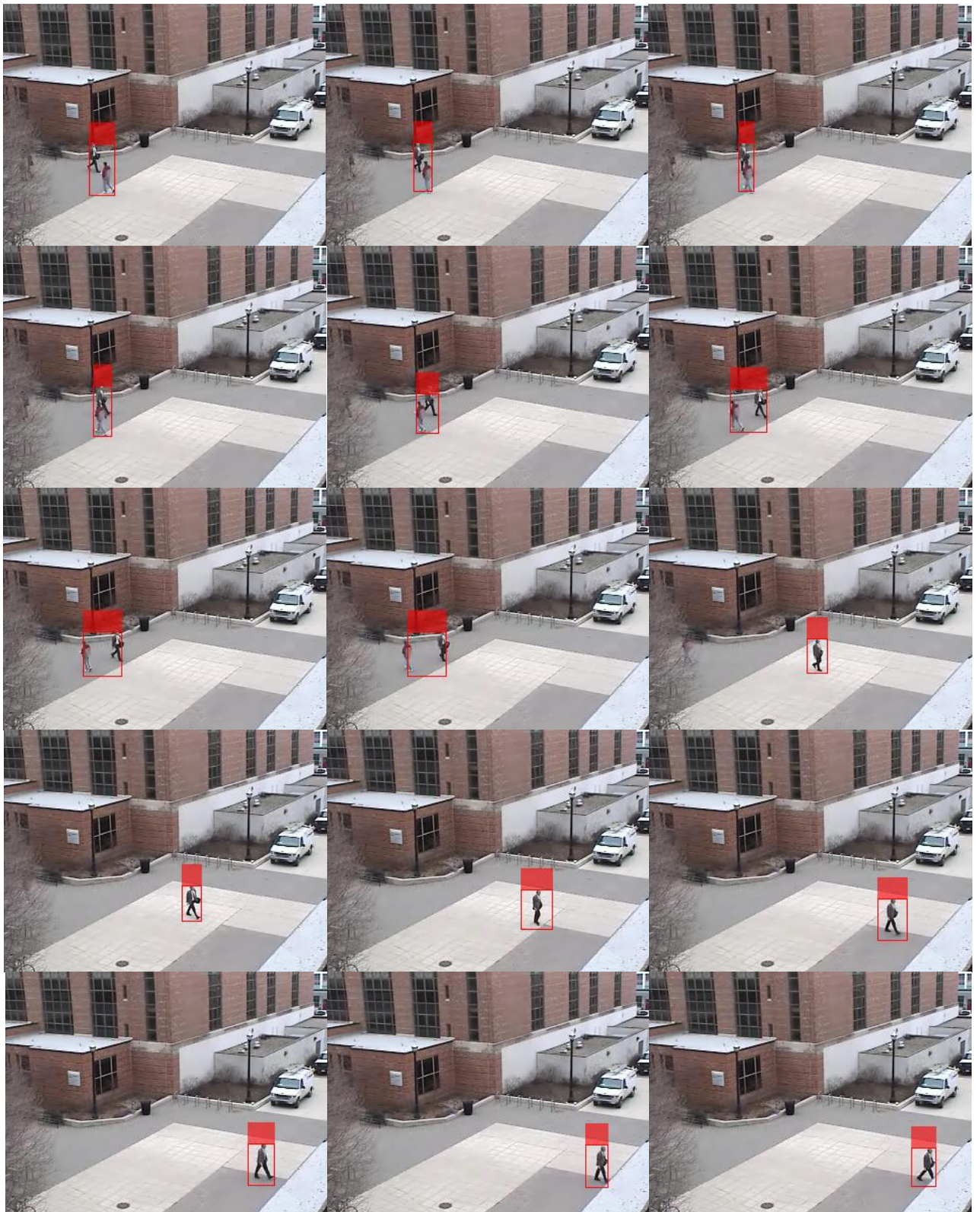


Figure 4.13. Résultat de suivi, avec deux objets mobiles proches

### **Conclusion**

Dans ce chapitre, nous abordons le problème de suivi de plusieurs personnes dans une séquence vidéo.

L'objectif principal de notre travail est de réaliser une application qui assure la détection de mouvement avec une caméra fixe. Nous avons défini les techniques de la vidéosurveillance ainsi que les méthodes de détection et suivi nous avons préparé les outils de traitement d'image utilisé dans notre travail. L'application développée est basé sur la méthode DWT pour la fusion soustraction de fond pour la détection filtre de kalman pour le suivi l'application développée est testée dans différentes situations de luminosité

On a obtenu des résultats acceptables, sauf dans le cas de plusieurs objets dans la scène aussi le temps de réponse est acceptable pour une application en temps réel.

## ***CONCLUSION GENERALE***

## **Conclusion Générale**

L'objectif principal de cette thèse était de suivre des objets dans le temps en utilisant une séquence vidéo. Pour cela, nous utilisons les données thermiques et couleurs.

Nous nous intéressons dans notre travail à la qualité de la fusion, pour avoir le maximum d'informations à partir des deux vidéos.

Dans le premier chapitre nous avons donné une étude panoramique sur la vidéosurveillance et ses générations, en s'intéressant à la vidéosurveillance intelligente et son architecture. Cette dernière, semble pouvoir résoudre beaucoup de problème de sécurité, mais a encore des limites et des difficultés

L'objectif de notre travail est d'exploiter des données thermiques et les données couleurs pour la détection et le suivi de personnes, donc la fusion de deux types de données est primordiale. Dans le deuxième chapitre nous donnons une étude bibliographique sur les méthodes de fusion, puis la détection et finalement le suivi.

Le formalisme mathématique des approches utilisées était présenté dans le troisième chapitre où nous avons montré d'abord un schéma synoptique de la solution proposée qui résume le travail, puis nous avons étudié la transformée en ondelette DWT qui a été utilisée pour la fusion. La soustraction de fond était présentée pour être utilisée pour la détection. Le suivi de personne est assuré par le filtre de Kalman, qui a été présenté dans le même chapitre.

Le dernier chapitre était consacré aux résultats, nous avons obtenu des résultats acceptables pour la fusion et la détection aussi pour le suivi sauf que nous avons trouvé une faible réponse pour suivre les personnes dans le cas de plusieurs objets dans la scène nous avons conclu aussi que le temps de réponse est acceptable pour une application en temps réel avec nos méthodes utilisées.

### **Perspectives**

Cette approche de suivre les objets peut encore être améliorée et donnera un parfait résultat. Dans ce travail, la fusion nous a donné un résultat mais il est essentiel pour les chercheurs d'améliorer le système de fusion pour la fusion de plus d'une image à partir du même capteur à une heure différente. Et de nouvelles techniques de fusion d'images doivent également être développées et ajoutées au système proposé. Et des chercheurs peuvent aussi utiliser d'autres solutions pour la détection d'objets multiples tout en implémentant sur un système réel



# ***BIBLIOGRAPHIES***

---

- [1] Al Najjar, M., Ghantous, M., & Bayoumi, M. (2014). Video Surveillance for Sensor Platforms. Springer.
- [2] Raty, T. D. (2010). Survey on contemporary remote surveillance systems for public safety. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 40(5), 493-515.
- [3] Valera, M., & Velastin, S. A. (2005, April). Intelligent distributed surveillance systems: a review. In Vision, Image and Signal Processing, IEE Proceedings-(Vol. 152, No. 2, pp. 192-204). IET.]
- [4] GOUAILLIER, V., & FLEURANT, A. (2009). La vidéosurveillance intelligente: promesses et défis. rapport de veille technologique et commerciale. Rapport technique, CRIM and Technopôle Défense et Sécurité
- [5] Luvison, B. (2010). Détection non supervisée d'évènements rares dans un flot vidéo: application à la surveillance d'espaces publics (Doctoral dissertation, Université Blaise Pascal-Clermont-Ferrand II).
- [6] Pop, S. (2008). Modèles de fusion et diffusion par équations aux dérivées partielles: application à la sismique azimutale (Doctoral dissertation, Bordeaux 1).
- [7] Kumar, S. (2013). Investigation and development of fusion techniques using night vision images through statistical and multi resolution approach.
- [8] Horaud, R., & Monga, O. (1995). Vision par ordinateur: outils fondamentaux (p. 426). Editions Hermès.
- [9] Fezzani, R. (2011). Approche parallèle pour l'estimation du flot optique par méthode variationnelle (Doctoral dissertation, Université Pierre et Marie Curie-Paris VI).
- [10] Marzat, J. (2008). Éstimation temps réel du Flot Optique. Master recherche adisibi, ENSEM, Nancy.
- [11] Dedeoglu, Y. (2004). Moving object detection, tracking and classification for smart video surveillance (Doctoral dissertation, bilkent university).
- [12] Ms. Sulochana T., M.Tech Scholar (VLSI Design And Embedded System), Depart of Electronics and Communication Engineering, Reva Institute of Technology and Management, Bangalore, Karnataka, India
- [13] CHAU Duc Phu « REPARATION DES TRAJECTOIRES DE PERSONNES SUIVIES BASEE SUR LE CLUSTERING DE POINTS

- [14] Yasuda, K., Naemura, T., & Harashima, H. (2004). Thermo-key: Human region segmentation from video. *Computer Graphics and Applications, IEEE*, 24(1), 26-30.
- [15] Liu, Z. (2010). *Investigations on multi-sensor image system and its surveillance applications*. Universal-Publishers.
- [16] Shivsubramani Krishnamoorthy July, 2008 *Development of Image Fusion Techniques And Measurement Methods to Assess the Quality of the Fusion*
- [17] Mamataj, S., Das, B., & Rahaman, A. (2014). An Ease implementation of 4-bit Arithmetic Circuit for 8 Operation by using a new reversible COG gate. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, 3(1).
- [18] Örtten, B. B. (2005). *Moving object identification and event recognition in video surveillance systems* (Doctoral dissertation, MIDDLE EAST TECHNICAL UNIVERSITY).
- [19] Sen-Ching, S. C., & Kamath, C. (2004, January). Robust techniques for background subtraction in urban traffic video. In *Electronic Imaging 2004* (pp. 881-892). International Society for Optics and Photonics.

### **webographies**

- [20] [www.ferdinandpiette.com](http://www.ferdinandpiette.com)
- [21] [kalman filter | OpenCV opencv.org](http://kalmanfilter|OpenCV.opencv.org)
- [22] Banc IEEE OTCBVS WS Séries; J. Davis et V. Sharma
- [23] [www.vcipl.okstate.edu](http://www.vcipl.okstate.edu)
- [24] <http://www.ino.ca/en/video-analytics-dataset/>



## Résumé

Les techniques de suivi et d'analyse d'images sont de mieux en mieux maîtrisées et leur utilisation est de plus en plus répandue. La vidéosurveillance est utilisée comme un des moyens les plus incontournables pour la sécurité et la surveillance de territoire, des biens ou des personnes. Malheureusement des développements technique et scientifiques sont encore nécessaires afin de maîtriser le traitement, le stockage et la fusion des images, afin d'en extraire les informations utiles. Pour cela, il faudra surmonter plusieurs contraintes.

L'objectif de notre travail est de détecter et suivre un objet en mouvement dans une scène filmée par deux types de caméras, couleur et infrarouge. Afin de pouvoir résoudre le problème :

La première étape de notre travail consiste à acquérir deux vidéos, une avec une caméra couleur et l'autre avec une caméra infrarouge ensuite pour chaque frame des deux vidéos, on effectue le traitement suivant :

La fusion des deux images à l'aide de DWT. En utilisant l'image fusionnée précédente, la détection de la personne en mouvement dans la scène est effectuée par la soustraction de fond. La dernière étape consiste à suivre cette personne, en utilisant le filtre de Kalman.

**Mot clés :** Vidéosurveillance intelligente,Fusion DWT,Détection,Soustraction de fond,Suivi,Filtre de kalman

## Abstract

Monitoring techniques and image analysis are increasingly well mastered and their use is becoming more widespread. Intelligent video surveillance is used as one of the most essential means for security and surveillance area, goods or people. Unfortunately technical and scientific developments are needed in order to control the processing, storage and fusion of images, in order to extract useful information. To do so will require overcoming several constraints.

The aim of our work is to detect and track a moving object in a scene filmed by two types of cameras, color and infrared. In order to solve the problem:

The first step of our work is to acquire two videos, one with a color camera and the other with an infrared camera then for each frame of two videos, one carries out the following process:

The fusion of the two images using DWT. Using the above image fusion, the detection of the person moving in the scene is performed by background subtraction. The final step is to follow that person, using the Kalman filter.

**Key words:** Intelligent video surveillance,Fusin,DWT, detection,background subtraction,tracking,kalman filter