

41 004. 27

République Algérienne Démocratique et Populaire

Ministère de l'enseignement supérieur et de la recherche scientifique

Université de 8 Mai 1945 - Guelma -

Faculté des Mathématiques, d'Informatique et des Sciences de la matière

Département d'Informatique



Mémoire de fin d'études Master

Filière : Informatique

Option : Système informatique

Thème :

Modélisation de l'arrière-plan dans les caméras amovibles

Encadré Par :

Dr. Brahim Farou

Présenté par :

Younes ASSASLA

Juin 2018

Remerciement

Je remercie ALLAH le tout puissant de m'avoir donné la santé et la volonté d'entamer et de terminer ce mémoire.

Tout d'abord, ce travail ne serait pas aussi riche et n'aurait pas pu avoir le jour sans l'aide et l'encadrement de Mr Brahim FAROU, je le remercie pour la qualité de son encadrement exceptionnel, pour sa patience, sa rigueur et sa disponibilité durant la préparation de ce mémoire.

Mes remerciements s'adressent également à tous mes professeurs pour leurs générosités et la grande patience dont ils ont su faire preuve malgré leurs charges académiques et professionnelles.

Dédicace

Je dédie ce modeste travail à :

*A mes parents .Aucun hommage ne pourrait être à la
hauteur de l'amour Dont ils ne cessent de me combler. Que
dieu leur procure bonne santé et longue vie.*

*A ma femme que j'aime beaucoup et qui m'a soutenu tout
au long de ce projet, et mes chers enfants Noure et Anes.*

et bien sur à mes frères et mes sœurs.

A toute ma famille, et mes amis,

*Et à tous ceux qui ont contribué de près ou de loin pour
que ce projet soit possible, je vous dis merci.*

Résumé

L'extraction de l'arrière-plan est une approche courante dans les systèmes de traitement de la vidéo pour détecter les objets en mouvement dans des séquences vidéo prises avec une caméra statique. De nombreux chercheurs ont essayé d'étendre ces méthodes de modélisation pour l'adapter sur les caméras Pan-tilt-zoom (PTZ). (Une caméra panoramique). Notre travail décrit un système qui permet l'extraction de l'arrière-plan et détecter les objets en mouvement dans des séquences vidéo prise avec une caméra PTZ. Ce système est composé de deux modules séparés, le premier module permet de construire les différentes images panoramiques basées sur les points caractéristiques de chaque image, et le second module permet la soustraction de fond en utilisant le modèle de mélange de gaussiennes. Les résultats obtenus sur des bases personnelles ont donné des résultats satisfaisants malgré les difficultés rencontrées pour la normalisation de cette dernière.

Mots-Clés : Vidéo surveillance, Modélisation de l'arrière-plan, GMM, détection de mouvement, image panoramique.

Table des matières

Résumé.	i
Table des matières.	iv
Table des figures.	vi
Liste des tableaux.	vii
Introduction Générale.	1
1 Les systèmes de vidéo surveillance	3
1.1 Introduction	3
1.2 La vidéo surveillance	4
1.2.1 Présentation	4
1.2.2 Domaine d'application	4
1.3 Les fichiers vidéos	6
1.3.1 Présentation	6
1.3.2 Paramètres d'une vidéo	6
1.3.3 Les caractéristiques d'une vidéo	7
1.4 Traitement et analyse d'image	9
1.4.1 Définition d'une image	9
1.4.2 Acquisition d'une image	10
1.4.3 Dimension	10
1.4.4 Résolution	10
1.4.5 Bruit	10
1.4.6 Histogramme	11

1.4.7	Luminance	11
1.4.8	Contraste	11
1.4.9	Les espaces de couleur	12
1.4.10	Le passage d'un espace à un autre	14
1.5	Les images panoramiques	15
1.5.1	Dispositifs et logiciels utilisés	15
1.5.2	Acquisition	17
1.5.3	Traitement	18
1.5.4	Assemblage	19
1.5.5	Immersion et visualisation	21
1.6	État de l'Art	22
1.7	Conclusion	25
2	Modélisation de l'arrière-plan	26
2.1	Introduction	26
2.2	Acquisition de la vidéo	26
2.3	Fragmentation et Prétraitement	27
2.4	Construction du champ de couverture	28
2.4.1	Présentation	28
2.4.2	Détection des points d'intérêts	29
2.4.3	La mise en correspondance	30
2.4.4	Estimation de la matrice de transformation (la matrice homographie)	31
2.4.5	L'algorithme RANSAC	31
2.4.6	Détermination des dimensions de l'image panoramique	32
2.4.7	Remplissage de l'image panoramique	32
2.5	Modélisation et l'extraction de l'arrière-plan	34
2.5.1	Les mixture de gaussiennes	34
2.5.2	L'image de référence	35
2.5.3	Initialisation	35
2.5.4	Mise à jour des paramètres	35
2.5.5	L'algorithme K-means	36
2.5.6	Application des opérations morphomathématiques	39

TABLE DES MATIÈRES

2.6	Conclusion	39
3	Tests et résultats	40
3.1	Introduction	40
3.2	Outils utilisés	40
3.2.1	Java	40
3.2.2	Eclipse	41
3.2.3	OpenCV	41
3.2.4	Jama	42
3.3	Base de tests	42
3.4	Tests et résultats	44
3.4.1	Résultat de l'algorithme de fragmentation	44
3.4.2	Résultat de la construction de l'image panoramique (l'image de référence) .	45
3.4.3	Résultat de l'algorithme de mise à jour d'une image panoramique	46
3.4.4	Résultat de l'algorithme d'extraction de l'arrière- plan (GMM)	47
3.4.5	Affichage du résultat sur une séquence d'images (sur la même base)	48
3.4.6	4.6. Affichage du résultat sur une autre base (au niveau de maison de culture 'Mohamed chafai')	49
3.5	Conclusion	51
	Conclusion Générale.	53
	Bibliographie.	54

Table des figures

1.1	Exemple de caméras de surveillance [STT]	5
1.2	Les caractéristiques d'un fichier vidéo.	7
1.3	Image Entrelacée [sdadAeP18]	8
1.4	Profondeur de couleur.	9
1.5	Les trois composant de l'espace RGB.	12
1.6	L'espace XYZ.	12
1.7	L'espace HSV.	13
1.8	L'espace Yuv.	14
1.9	Objectif fisheye NIKON FC-E8 Focale 0.21, angle de vue 183 [Rob09].	16
1.10	Capture d'une image panoramique en utilisant un miroir sphérique [Rob09].	16
1.11	Projection cylindrique de la scène [Rob09].	16
1.12	Principe de la projection cylindrique avec une visualisation par développement [Rob09].	20
1.13	Projection sur un icosaèdre tronqué [Rob09].	20
1.14	Changement de l'emplacement d'un pixel dans une image panoramique.	23
2.1	L'architecture du système.	27
2.2	Détection des points d'intérêts.	29
2.3	La mise en correspondance.	30
2.4	Remplissage de l'image panoramique.	33
2.5	Assemblage et rectification.	33
3.1	Vidéo fragmentée.	44
3.2	Construction de l'image panoramique	45
3.3	Mettre à jour l'image panoramique	46

TABLE DES FIGURES

3.4	L'extraction de l'arrière-plan	47
3.5	Résultat sur une séquence d'images	48
3.6	Résultat du matching sur une autre vidéo	49
3.7	Résultat de la détection sur une autre vidéo	50

Liste des tableaux

3.1 Description de la base de tests.	43
--	----

Introduction Générale

Avec la généralisation de l'utilisation d'images numériques, la détection du mouvement dans les vidéos s'est révélée être un outil indispensable pour des applications aussi diverses que la compression et l'indexation vidéo, les applications propres à la robotique (détection d'obstacles, guidage de missile) ou encore celles liées à la télésurveillance (intrusion, trafic routier ou poursuites de cibles).

Plusieurs de ces applications nécessitent un champ de vision large telles que la construction de cartes aériennes ou satellites ou encore la vidéo surveillance. Cependant, de nombreux systèmes conventionnels de capture d'images sont limités par leur champ de vision, souvent plus petit que celui de l'humain. La création d'images panoramiques, qui sont des images à champ visuel très large, est un moyen permettant de compenser ce champ visuel limité. Les images panoramiques sont tout d'abord apparues dans le domaine de la peinture, à la fin du 18^e siècle, dans le but de donner l'illusion de regarder une scène réelle, avant de gagner le domaine de la photographie.

Dans ce mémoire, nous allons réaliser un système de détection des objets en mouvement dans une séquence vidéo filmée par une caméra mobile.

Le but de ce travail est d'exploiter les caractéristiques des images panoramiques pour élargir le champ de vision des caméras et rendre possible l'application des algorithmes destinés aux caméras statiques pour les caméras en mouvement linéaire. Tout d'abord, nous allons proposer une méthode pour la création des images panoramique pour offrir une vision plus large pour les systèmes de vidéo surveillance et par conséquent minimiser le nombre de caméras nécessaires pour la surveillance d'un lieu public ou privé. Cette démarche nous a permis de passer d'un état dynamique de la caméra vers un état statique permettant d'appliquer les anciennes techniques de détection de mouvement sur les caméras en mouvement. L'image de référence ainsi créée sera mise à jour à jour régulièrement afin de permettre aux algorithmes de modélisation de l'arrière-plan de retrouver les pixels correspondant aux objets mobiles du premier plan.

Le manuscrit est organisé en trois chapitres, précédé par une introduction générale et suivi par une conclusion générale et quelques perspectives.

Nous allons donc, dans le premier chapitre, faire une petite présentation sur la vidéo surveillance, puis nous allons présenter des généralités sur la vidéo et leurs caractéristiques. Nous allons ensuite aborder l'image en traitant les mêmes points, puis nous allons terminer ce chapitre par un aperçu détaillé sur les images panoramiques.

Dans le deuxième chapitre, nous allons présenter l'architecture générale de notre système, qui se divise en plusieurs parties. Chaque partie est un processus qui permet de réaliser une tâche spécifique relative à un traitement de la vidéo surveillance. Nous allons d'abord aborder la fragmentation de la séquence vidéo filmée par une caméra mobile, ensuite, nous sélectionnons les images couvrant toute la scène pour construire l'image panoramique qui représente le modèle statique de la scène complet. Puis nous détaillons les différentes étapes d'extraction de l'arrière-plan et la détection des objets dans cette scène étendue.

Le troisième chapitre est dédié aux tests. Nous allons d'abord définir l'environnement de travail, le matériel sélectionné ainsi que les environnements de développement utilisés dans cette application. Nous allons ensuite présenter quelques résultats qualitatifs des algorithmes développés sur l'ensemble des vidéos filmées par une caméra mobile.

Nous allons terminer ce manuscrit par une conclusion générale et quelques perspectives relatives à l'utilisation des caméras en mouvement dans les systèmes de vidéo surveillance.

Chapitre 1

Les systèmes de vidéo surveillance

1.1 Introduction

Il est manifeste aujourd'hui que la surveillance des lieux et des personnes se généralise. Actuellement, que ce soit dans des zones du domaine public ou chez des particuliers, si cette tendance se poursuit, il est bien évident que les besoins d'une analyse automatique des séquences vidéo obtenues seront de plus en plus importants. Dans un tel contexte, on peut distinguer les systèmes de vidéo surveillance.

Ces systèmes sont été apparus pour la première fois durant la deuxième guerre mondiale par les Allemands pour des raisons de sûreté pour observer le lancement de leurs missiles. Depuis, le secteur civil s'est intéressé à ce principe afin de prévenir le terrorisme et plus généralement la criminalité ,ces systèmes sont été industrialisés en permettant de contrôler plusieurs lieux sur une batterie d'écrans.

La première génération des systèmes de vidéo surveillance existait entre les années 60 jusqu'aux années 80 et ils étaient basés sur les traitements suivants : l'acquisition, la transmission et le traitement de l'image. Le rôle principal de ce système est d'augmenter la vision de l'être humain par la transmission des vidéos de plusieurs caméras qui surveillent plusieurs endroits dans une seule chambre et sur un seul moniteur. Le problème majeur de ce système est : la fragilité des observateurs humains qui ne peuvent malheureusement pas rester figer avec une très grande concentration sur le moniteur pendant une longue période.[FAR16]

La deuxième génération est venue pour pallier un peu les problèmes de la première génération en utilisant des systèmes hybrides basés à la fois sur le signal analogique et numérique. Cette

hybridation a permis d'introduire quelques opérations qui jouaient un peu le rôle d'un assistant aux agents de surveillance [FAR16].

La troisième génération, apparue depuis les années 2000, est totalement basée sur un système digital allant de la caméra, soit fixe ou mobile, vers les systèmes de traitement de l'information.

Cette troisième génération a été conçue pour permettre un déclenchement des alarmes en temps réel et une recherche efficace des informations hors ligne. Et c'est pour cette raison que le défi des systèmes de troisième génération était d'offrir des systèmes intelligents capables de générer en temps réel des alarmes traitant des événements très complexes et de sauvegarder ces derniers pour une éventuelle recherche d'information. La construction de tels systèmes nécessite des algorithmes rapides robustes et efficaces pour la détection, la classification, l'analyse et le suivi des activités des objets en mouvement. [FAR16]

Tous les systèmes de surveillance cités ci-dessus fonctionnent avec des caméras stationnaires, de nombreux chercheurs ont essayé d'étendre ces systèmes pour les adapter sur les caméras Pan-tilt-zoom (PTZ). L'utilisation de ce type caméra permet de contrôler interactivement la prise de vue. Le pilotage de la caméra, suivant deux axes de rotation, permet de sécuriser un champ de vue étendu avec une résolution importante et permet de générer un panorama plus ou moins complet de la scène.

1.2 La vidéo surveillance

1.2.1 Présentation

La vidéo surveillance est un système de caméras et de transmission d'images, disposé dans un endroit public ou privé pour le surveiller à distance ; il s'agit donc d'un type de télésurveillance. Les images obtenues avec ce système peuvent être traitées automatiquement et/ou visionnées puis archivées ou détruites. La surveillance a pour but de contrôler les conditions de respect de la sécurité, de la sûreté ou de l'exécution d'une procédure particulière. [wik]

1.2.2 Domaine d'application

Les domaines d'application de la vidéo surveillance sont variés et peuvent être publics ou privés. La vidéo surveillance offre aux agents responsables de la surveillance des informations de



FIGURE 1.1 – Exemple de caméras de surveillance [STT] .

haut niveau pour simplifier la prise de décision dans un temps très réduit et pour offrir, en cas d'accident, une plateforme de recherche effective des vidéos enregistrés pour un traitement hors ligne.

1.2.2.1 La sécurité publique et commerciale

- La surveillance des lieux publics (banques, les magasins de stockage, les aéroports, les musées, les stations de service) et les propriétés privées et les parkings, elle est utilisée principalement pour la prévention de la délinquance.
- La surveillance de transport privé ou collectif pour la détection des accidents.
- La surveillance des maisons, des magasins et des forêts pour la détection des incendies.
- Les contrôles d'accès aux endroits sécurisés.

1.2.2.2 Exploitation des données

La vidéo surveillance nous permet de :

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

1.2.2.3 Le trafic autoroutier

- Mesurer la vitesse des véhicules dans les autoroutes.

- Détecter les infractions des feux rouges et le dépassement de la ligne continu.

1.2.2.4 Sécurité militaire

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

1.3 Les fichiers vidéos

Le système de vidéo surveillance nous génère des fichiers de type vidéo; que nous allons étudier dans ce qui suit.

1.3.1 Présentation

La vidéo regroupe l'ensemble des techniques permettant l'enregistrement ainsi que la restitution d'images animées, accompagnées ou non de son, sur un support électronique et non de type photochimique.

Le flux visuel comporte une séquence d'images fixes qui selon l'axe temporel apparaissent animées à une fréquence de 24 à 30 images par seconde.

1.3.2 Paramètres d'une vidéo

1.3.2.1 La résolution en luminance

elle détermine le nombre de couleurs possibles pour représenter un pixel (généralement 8 bits pour les images en niveau de gris et 24 bits pour les images couleur).

1.3.2.2 La résolution spatiale

définis le nombre de lignes et de colonnes de la matrice de pixels.

1.3.2.3 La résolution temporelle

est le nombre d'images par seconde.

1.3.3 Les caractéristiques d'une vidéo

On représente dans le schéma suivant les principales caractéristiques d'un fichier vidéo :



FIGURE 1.2 – Les caractéristiques d'un fichier vidéo.

1.3.3.1 La cadence

C'est le nombre d'images par seconde, sur l'écran c'est toujours une suite d'images fixes, qui donne l'illusion du mouvement (cinéma : 24 images/seconde c'est la norme internationale, Télévision : en France 25 images/seconde et 29.97 images/seconde aux Etats Unis)[FAR16].

1.3.3.2 Définition de l'image

Une image numérique est formée de pixels, sa qualité va dépendre, entre autres du nombre de pixels qui la composent.

1.3.3.3 Le Format d'une image

Les formats populaires des images d'une vidéo sont rectilignes, donc ils ont une largeur et une hauteur. La largeur et la hauteur sont généralement mesurées en pixel, si par exemple une vidéo est de 320 pixels de largeur et 240 pixels de hauteur il est dit avoir des dimensions de 320X240 pixel , et dispose d'un ratio de 4 :3.

1.3.3.4 Entrelacement vidéo

La fréquence de 25 images par seconde est suffisante pour la perception de l'œil humain, mais sur un écran cathodique, l'image avait tendance à scintiller. Les ingénieurs ont trouvé qu'en affichant la même image en deux parties, d'abord les lignes impaires puis les lignes paires, le scintillement disparaît grâce à la rémanence des tubes cathodiques. Par conséquent le signal vidéo TV est entrelacé : 50 demi-images par seconde en PAL (50i). Tout fonctionnait parfaitement bien puisque les caméras filmaient en mode entrelacé, le montage s'effectuait à l'aide d'un moniteur de contrôle (écran cathodique) où l'image entrelacée s'affichait correctement et on se servait du support DVD (exclusivement en mode entrelacé) pour la diffusion [sdadAeP18].

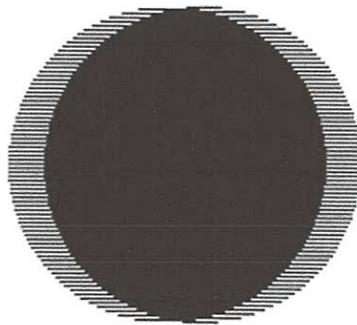


FIGURE 1.3 – Image Entrelacée [sdadAeP18]

1.3.3.5 Profondeur de couleur et débits

Une image numérique utilise plus ou moins de mémoire selon le codage des informations de couleur qu'elle possède. C'est ce que l'on nomme le codage de couleurs ou profondeur des couleurs, exprimé en bit par pixel (bpp) : 1, 4, 8, 16 bits... En connaissant le nombre de pixels d'une image et la mémoire nécessaire à l'affichage d'un pixel, il est possible de définir exactement le poids que va utiliser le fichier image sur le disque dur (ou l'espace mémoire requis en RAM pour réaliser un calcul sur cette image).

1.3.3.6 Codec

Un codec est un dispositif matériel ou logiciel permettant de mettre en œuvre l'encodage ou le décodage d'un flux de données numérique, en vue d'une transmission ou d'un stockage. Certains codecs intègrent également une fonction de compression ou encore de chiffrement des données.

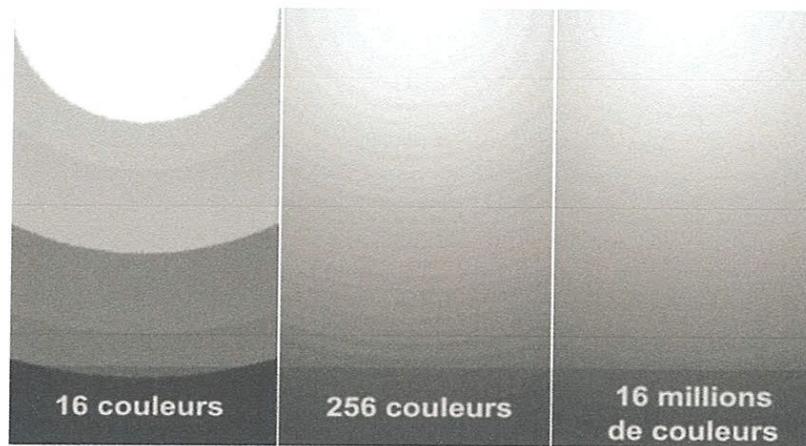


FIGURE 1.4 – Profondeur de couleur.

Selon le type d'encodage utilisé, on obtient un format spécifique, ces formats sont des types vidéos que l'on peut créer via un logiciel, à des fins de stockage pour être visionné, diffusé et partagé, etc. Les formats vidéo actuels sont regroupés en deux sous-catégories : les formats vidéos conçus pour les systèmes de diffusion vidéo/TV et ceux conçus pour les images d'ordinateur. Les exigences techniques diffèrent d'une catégorie à l'autre.

1.4 Traitement et analyse d'image

Le traitement d'images est un domaine très vaste qui a connu, et qui connaît encore, un développement important depuis quelques dizaines d'années.

On désigne par traitement d'images numériques l'ensemble des techniques permettant de modifier une image numérique afin d'améliorer ou d'en extraire des informations.

Nous allons donner un aperçu sur les différentes techniques connues dans ce domaine.

1.4.1 Définition d'une image

Une image numérique est une image dont la surface est divisée en éléments de tailles fixes. Appelés pixels, ayant chacun comme caractéristique un niveau de gris ou de couleurs prélevé à l'emplacement correspondant dans l'image réelle, ou calculé à partir d'une description interne de la scène à représenter.

1.4.2 Acquisition d'une image

L'acquisition d'images constitue une étape importante dans la chaîne de conception et de production d'images. Avant de manipuler l'image par un système informatique il est nécessaire de lui faire subir une transformation qui la rendra lisible et manipulable par ce système. Pour effectuer la transformation de cet objet externe (l'image d'origine) à sa représentation interne (dans l'unité de traitement) on doit passer par la procédure de numérisation. Ces systèmes de saisie, dénommés optiques, peuvent être classés en deux catégories principales :

- les caméras numériques.
- et les scanners.

1.4.3 Dimension

On désigne par une dimension la taille d'une image qui se présente sous forme de matrice telle que le nombre de lignes de cette matrice multiplié par le nombre de colonnes nous donne le nombre total de pixels de cette image. Chaque pixel a une valeur numérique représentative des intensités lumineuses (pixels).

1.4.4 Résolution

C'est la clarté ou la finesse de détails atteinte par un moniteur ou une imprimante dans la production d'images. Sur les moniteurs d'ordinateurs, la résolution est exprimée en nombre de pixels par unité de mesure (pouce ou centimètre). On utilise aussi le mot résolution pour désigner le nombre total de pixels affichables horizontalement ou verticalement sur un moniteur ; plus grand est ce nombre, meilleure est la résolution .

1.4.5 Bruit

Un bruit (parasite) dans une image est considéré comme un phénomène de brusque variation de l'intensité d'un pixel par rapport à ses voisins, il provient de l'éclairage des dispositifs optiques et électroniques du capteur.

1.4.6 Histogramme

L'histogramme des niveaux de gris ou des couleurs d'une image est une fonction qui donne la fréquence d'apparition de chaque niveau de gris (couleur) dans l'image. Il permet de donner un grand nombre d'informations sur la distribution des niveaux de gris (couleur) et de voir entre quelles bornes est répartie la majorité des niveaux de gris (couleur) dans le cas d'une image trop claire ou d'une image trop foncée [BH16].

Il peut être utilisé pour améliorer la qualité d'une image (Rehaussement d'image) en introduisant quelques modifications, pour pouvoir extraire les informations utiles de celle-ci.

1.4.7 Luminance

C'est le degré de luminosité des points de l'image. Elle est définie aussi comme étant le quotient de l'intensité lumineuse d'une surface par l'aire apparente de cette surface, pour un observateur lointain, le mot luminance est substitué au mot brillance, qui correspond à l'éclat d'un objet. Une bonne luminance se caractérise par [BH16] :

- Des images lumineuses (brillantes).
- Un bon contraste : il faut éviter les images où la gamme de contraste tend vers le blanc ou le noir, ces images entraînent des pertes de détails dans les zones sombres ou lumineuses
- L'absence de parasites.

1.4.8 Contraste

C'est l'opposition marquée entre deux régions d'une image, plus précisément entre les régions sombres et les régions claires de cette image. Le contraste est défini en fonction des luminances de deux zones d'images. Si L_1 et L_2 sont les degrés de luminosité respectivement de deux zones voisines A_1 et A_2 d'une image, le contraste C est défini par le rapport [BH16] :

$$C = \frac{L_1 - L_2}{L_1 + L_2} \quad (1.1)$$

1.4.9 Les espaces de couleur

1.4.9.1 L'espace RGB

L'espace RGB est l'espace vectoriel engendré par les trois composantes primaires (Rouge, Vert, Bleu).

L'ensemble des couleurs produites se représente comme l'intérieur d'un cube :

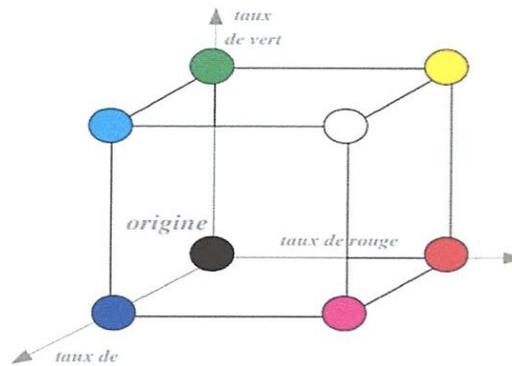


FIGURE 1.5 – Les trois composant de l'espace RGB.

1.4.9.2 L'espace XYZ

L'espace XYZ (CIE 1931) est défini à partir d'une transformation linéaire sur l'espace RGB telle que toutes les couleurs du spectre visible soient contenues dans le triangle XYZ. Les coordonnées XYZ des couleurs naturelles ont ainsi toujours des valeurs positives.

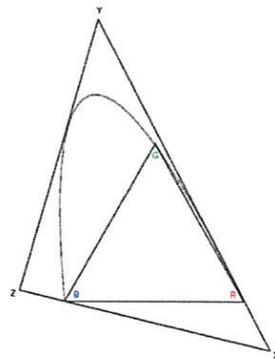


FIGURE 1.6 – L'espace XYZ.

1.4.9.3 L'espace HSV

Le principe de l'espace HSV est de caractériser les couleurs de façon plus intuitive, conformément à la perception naturelle des couleurs, en termes de (teinte,saturation,valeur) :

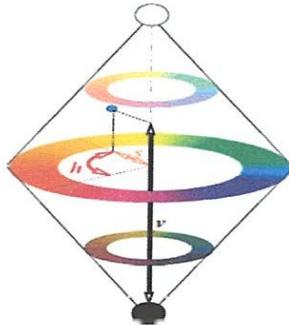


FIGURE 1.7 – L'espace HSV.

Teinte : intuitivement, c'est le nom qu'on utilisera pour désigner la couleur, 'vert', 'mauve', 'orange', etc. Idéalement associé à une longueur d'onde, donc à une position sur le cercle de Newton.

Saturation : c'est le taux de pureté de la couleur, qui doit varier entre la pureté maximale (couleur éclatante) et l'achromatisme (niveau de gris).

Valeur : c'est la mesure de l'intensité lumineuse de la couleur, qui doit varier entre le noir absolu et le blanc.

1.4.9.4 L'espace Yuv

Le principe de l'espace Yuv est de représenter les couleurs en utilisant une composante de luminance Y, et 2 composantes de chrominance (u,v) correspond aux composantes : bleue et rouge dans les coordonnées chromatiques réduites :

$$C_b + C_r + C_g = C_{te} \quad (1.2)$$

$$C_g = C_{te} - C_b - C_r \quad (1.3)$$

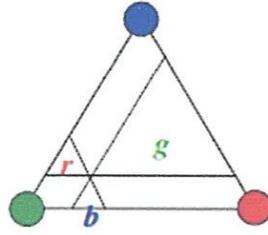


FIGURE 1.8 – L'espace Yuv.

1.4.10 Le passage d'un espace à un autre

- Le passage de RGB a XYZ :

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 2,769 & 1,7518 & 1,1300 \\ 1,000 & 4,5907 & 0,0601 \\ 0,000 & 0,0565 & 5,5943 \end{pmatrix} \cdot \begin{pmatrix} R \\ G \\ B \end{pmatrix} \quad (1.4)$$

- Le passage de RGB a HSV :

$$v = \frac{r + g + b}{3} \quad (1.5)$$

$$s = 1 - \frac{3\min(r, g, b)}{r + g + b} \quad (1.6)$$

$$h = \begin{cases} \theta & \text{si } b \leq g \\ 2\pi - \theta & \text{si } b > g \end{cases} \quad (1.7)$$

$$\text{avec : } \theta = \arccos\left(\frac{(r - g) + (r - b)}{2\sqrt{(r - g)^2 (r - b) (g - b)}}\right) \quad (1.8)$$

- Le passage de RGB a Yuv :

$$Y = 0,2989r + 0,5866g + 0,1145b \quad (1.9)$$

$$u = 0,5647(b - Y) = -0,1687r - 0,3312g + 0,5000b \quad (1.10)$$

$$v = 0,7132(r - Y) = 0,5000r - 0,4183g - 0,0817b \quad (1.11)$$

1.5 Les images panoramiques

L'utilisation des images panoramiques est largement répandue dans des domaines tels que la surveillance, la vision par ordinateur, la robotique et la réalité virtuelle. Elles sont également utilisées dans des applications commerciales comme le divertissement, la TV interactive, l'immobilier et le tourisme virtuel. Ces 15 dernières années, les systèmes panoramiques de formation d'image ont sensiblement progressé.

1.5.1 Dispositifs et logiciels utilisés

Non seulement les professionnels peuvent créer et afficher des panoramas, mais il existe une grande quantité de logiciels disponibles, dont certains gratuits. Chacun d'entre nous, avec un ordinateur et simple appareil photo numérique, peut créer des panoramas. Même les grandes compagnies sont présentes sur ce marché comme Apple avec son logiciel QuickTime2 ou Canon avec le logiciel PhotoStitch3 qui permettent un accès facile à la création d'images panoramiques. Il existe également plusieurs méthodes pour capturer des panoramas. La solution la plus simple consiste à utiliser un appareil photo classique monté sur un trépied et en faisant tourner manuellement l'appareil photo autour de son centre optique. Les professionnels auront plutôt tendance à utiliser une caméra motorisée de type PTZ (Pan, Tilt, Zoom). Le principe est exactement le même, à la différence que les rotations de la caméra autour de son centre optique sont pilotées et contrôlées par un ordinateur. Ceci permet de déterminer avec plus de précision les conditions de la prise de vue. L'un des inconvénients de cette approche, qu'elle soit manuelle ou pilotée, est qu'il faut du temps pour parcourir l'ensemble de la scène. La représentation panoramique ne peut pas être réalisée en temps réel. À un instant donné, une seule fraction de la scène est vue par la caméra. Des solutions existent pour capturer l'ensemble de la scène en temps réel. La première consiste à utiliser des équipements photographiques appelés omnidirectionnels. Ils peuvent être dioptriques ou catadioptriques [Rob09].

1.5.1.1 Les équipements dioptriques

Ils sont composés d'un capteur d'image et d'éléments simplement réfléchissants (objectif grand angle ou fish-eyes par exemple).



FIGURE 1.9 – Objectif Fisheye NIKON FC-E8 Focale 0.21, angle de vue 183 [Rob09].

1.5.1.2 Les équipements catadioptriques

dans ces équipements des éléments réfractifs (miroir sphérique ou parabolique) vont être ajoutés. Cette première solution permet effectivement de capturer l'ensemble de la scène en temps réel. Cependant, la résolution obtenue est généralement faible, même en utilisant un capteur à haute résolution [Rob09].



FIGURE 1.10 – Capture d'une image panoramique en utilisant un miroir sphérique [Rob09].



FIGURE 1.11 – Projection cylindrique de la scène [Rob09].

La construction et l'exploitation d'une image panoramique nécessitent 4 grandes étapes :

L'acquisition, la projection des prises de vue dans l'image panoramique, l'amélioration du rendu et la visualisation. Nous allons présenter rapidement ces quatre points [Rob09].

1.5.2 Acquisition

La formation d'image panoramique est un exercice particulier qui requiert un nombre d'images différent selon la technologie employée et le but final de l'application. Dans tous les cas, la méthode qui doit être utilisée pour la construction de l'image panoramique est fonction des paramètres suivants : résolution, couverture visuelle et temps d'acquisition. L'utilisation de caméra omnidirectionnelle est fréquemment utilisée dans la robotique. Le but ici est d'obtenir une vue la plus large possible dans le périmètre immédiat du capteur. Les caméras omnidirectionnelles sont à privilégier lorsque le temps d'acquisition est inférieur à la seconde et que la résolution ou plus précisément la profondeur de champ ne sont pas très importantes. Avec ces modèles de caméras, sur le plan horizontal, le champ couvert peut aller sans grande difficulté jusqu'à 360. Par contre, sur le plan vertical, le champ couvert dépasse plus rarement 90. Avec une caméra unique en rotation autour de son centre optique le champ couvert peut aller jusqu'à 360 sur le plan horizontal et 180 sur le plan vertical. Il dépend essentiellement de la conception mécanique de la mise en rotation de la caméra. La résolution peut être très importante en fonction de la distance focale de l'objectif. Par contre le temps d'acquisition est alors très long et dépend finalement du nombre d'images nécessaires pour une couverture optimale de la scène à observer. Une solution intermédiaire consiste à utiliser ce que l'on appelle traditionnellement un « bouquet » de caméra. Il s'agit d'un ensemble de caméras installées autour d'un axe. Le bouquet de caméra est un compromis intéressant puisqu'il permet d'augmenter la résolution tout en gardant l'aspect temps réel de l'acquisition. Même si le champ couvert peut être de 360 sur le plan horizontal et 180 sur le plan vertical, des contraintes mécaniques limitent généralement celui-ci. De plus cette solution est relativement coûteuse. Outre le nombre plus important de caméras, cette solution nécessite des ressources en matériels plus importantes (carte d'acquisition, processeur, mémoire ...). Le système GeoView 30004 de la société iMove utilise 6 appareils photo, quatre sur le plan horizontal, un vers le bas et un vers le haut [Rob09].

1.5.3 Traitement

Les images capturées par les dispositifs panoramiques vont devoir subir un certain nombre de traitements avant de pouvoir être exploitées. Dans le cas des dispositifs omnidirectionnels, l'image est très déformée. Dans les autres dispositifs, une image ne contient qu'une partie de la scène. La représentation complète de la scène nécessite donc plusieurs images qui vont être projetées sur une image panoramique. Les images ne sont donc pas forcément acquises au même moment ou avec le même capteur. Si les informations de prise de vue ne sont pas précises ou si elles ne sont pas connues, une étape de mise en correspondance sera également nécessaire. Dans le cas où une seule caméra est utilisée, les différentes images formant le panorama ne sont pas prises en même temps. Les conditions d'illumination peuvent être différentes. De même, la caméra peut disposer d'une correction automatique la luminosité de façon à ajuster le gain du capteur. Les images devront alors subir un traitement pour supprimer les effets de couture [Rob09].

1.5.3.1 Mise en correspondance

La mise en correspondance est un problème difficile qui continue à mobiliser la communauté scientifique. Plusieurs solutions sont proposées dans la littérature et de nombreux articles proposent une synthèse de ces méthodes. Le principe de base consiste à mettre en correspondance des zones de l'image en fonction de leurs propriétés radiométriques ou géométriques. Les méthodes sont généralement classées en deux catégories : denses ou éparées. Les méthodes denses cherchent à estimer les paramètres de transformation en exploitant l'intensité de l'ensemble des pixels contenu dans la zone de recouvrement. Les méthodes éparées n'utilisent pas tous les points, mais seulement quelques points particuliers (coins, extrema locaux) ou des primitives géométriques (lignes, cercles...). Dans le cas où il n'y a pas de déformation des images, les méthodes denses sont réputées plus précises, mais elles sont souvent moins rapides, moins robustes aux changements de luminosité et à la présence d'objet en mouvements [Rob09].

1.5.3.2 Correction du gain

Le contrôle de la luminosité n'est pas un problème aussi simple qu'il y paraît. Si la luminosité est relativement constante dans toutes les directions alors il suffit de fixer manuellement le gain de la caméra pendant toute la durée de la prise de vue. Par contre, il ne faut pas que cette luminosité évolue pendant cette prise de vue (lors du passage d'un nuage par exemple). Cependant, même si la

luminosité générale n'évolue pas au cours de la prise de vue, si les écarts de luminosité dans la scène sont trop importants, les zones sombres apparaîtront sous-exposées dans la mosaïque d'images. À l'inverse, d'autres zones beaucoup plus lumineuses risquent de se trouver sur-exposées. La solution pourrait être de laisser la caméra gérer automatiquement le gain en fonction de la luminosité de la portion de scène visée. Cette solution offre bien sûr l'avantage que chaque image est acquise en optimisant la dynamique du capteur. Le problème est qu'une même portion de la scène prise avec deux gains différents n'a plus la même valeur de luminosité. L'exemple suivant est un cas d'école. La caméra est placée à l'intérieur d'une pièce. Cette pièce est éclairée par une lumière artificielle et par une lumière naturelle à travers une fenêtre. La luminosité à l'intérieur est faible par rapport à celle de l'extérieur. Si le gain de la caméra est fixé manuellement pour obtenir une luminosité correcte à l'intérieur de la pièce, la fenêtre apparaîtra sur-exposée et il ne sera pas possible de visualiser l'extérieur. Dans le cas contraire, si le gain est piloté automatiquement par la caméra, alors l'extérieur devient visible, mais le phénomène de « couture » apparaît [Rob09].

1.5.4 Assemblage

La mise en correspondance des images est utilisée pour déterminer avec la meilleure précision possible les paramètres de la prise de vue. Cette information permet de projeter correctement l'image dans le panorama. Dans le cas général, une homographie est calculée à partir de surfaces planes contenues dans la zone de recouvrement des deux images. Nous verrons que dans le cas particulier des caméras PTZ où le centre de projection est confondu avec le centre des rotations, tous les points de la zone de recouvrement peuvent être utilisés pour calculer l'homographie, même s'ils n'appartiennent pas à des surfaces planes. Avant toute chose, nous devons définir le modèle de représentation que nous allons utiliser. Lorsqu'il s'agit de fusionner quelques images de façon à obtenir une image globale plus large, il suffit de prendre une image de référence et de projeter toutes les images dans le plan cette image de référence. Lorsque l'angle solide devient trop important, les déformations inhérentes à la projection sont importantes. Au-delà d'une certaine limite, l'homographie ne peut plus être calculée où du moins ne permet pas une projection conforme. Il est donc nécessaire d'utiliser d'autres modes de projection. Nous classons ces projections en deux catégories. Les projections planes et la projection sur polyèdre. Les projections planes permettent de visualiser l'ensemble de la scène sur une seule image, mais elles entraînent des déformations importantes. Parmi les projections planes, la plus utilisée est la projection cylindrique [Rob09].

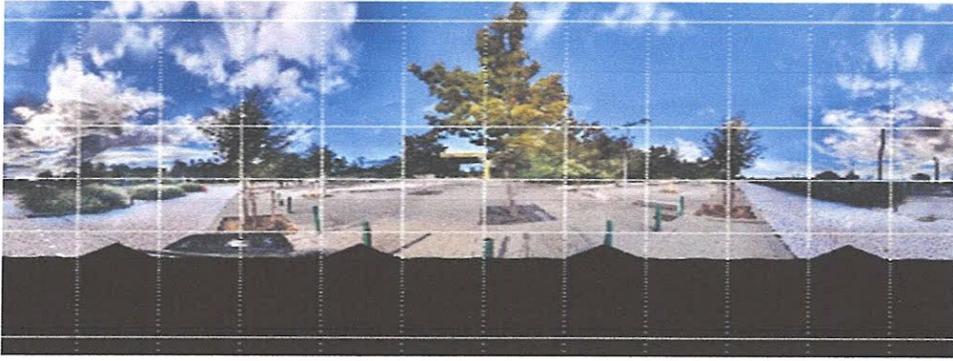


FIGURE 1.12 – Principe de la projection cylindrique avec une visualisation par développement [Rob09].

Un autre mode de représentation est la projection de la scène sur un polyèdre. Ces projections ne permettent pas la représentation de la scène en une seule image, mais elles limitent les déformations ainsi que le coût du stockage. Le principe est de projeter les images sur les faces du polyèdre. Un autre avantage de ce type de représentation est que la projection d'une image sur une face, se calcule à partir d'une homographie. Habituellement, le polyèdre utilisé est le cube. Nous lui préférons l'icosaèdre tronqué pour deux raisons. La première est qu'à résolution identique, l'espace mémoire est moins important dans le cas de l'icosaèdre tronqué. Le deuxième est que les angles sont moins aigus (plus « plats ») que sur le cube, ce qui limite les défauts lors du rendu [Rob09].



FIGURE 1.13 – Projection sur un icosaèdre tronqué [Rob09].

1.5.5 Immersion et visualisation

La visualisation est souvent l'étape ultime du mosaïquage. Bien que les techniques de base soient connues depuis très longtemps, ce n'est que récemment que les mosaïques d'images servent à autre chose qu'à la visualisation. Cependant, cela reste pour beaucoup, le seul intérêt de ces constructions. Comme nous venons de le voir, la projection d'une image sur un cylindre permet de visualiser l'image à partir d'une seule vue avec une limitation au niveau des pôles, mais surtout en déformant la scène. Une fois le mosaïquage réalisé, plusieurs logiciels permettent de visualiser une portion de la scène en redressant l'image dans les mêmes conditions que si elle avait été acquise par la caméra. Il est alors possible de se déplacer dans la scène et d'effectuer des changements de focale afin d'augmenter ou diminuer l'angle de visualisation. Le logiciel le plus répandu est QuickTimeVr. En interne, il utilise une structure de donnée basée sur la projection sur un cube. S'il est relativement facile de réaliser un fichier compatible avec QuickTimeVr, l'étape la plus délicate est de construire correctement les faces du cube en supprimant tous les défauts que nous avons évoqués précédemment. L'inconvénient de la plupart des logiciels est qu'ils ne gèrent pas l'aspect multi-résolution ou alors indirectement. Ce que nous entendons par « gestion de la multi résolution » est la possibilité de ne pas stocker l'ensemble du panorama avec la même résolution. En effet, l'ensemble de la scène n'est pas forcément d'un intérêt égal. Lorsque nous réalisons un panorama, nous aimerions avoir plus de détail sur certaines zones de l'image et moins sur d'autres (le ciel par exemple). Avec la plupart des logiciels disponibles, c'est le niveau de détail le plus fin qui conditionne la résolution du panorama tout entier. Le temps de calcul et surtout la taille du panorama ne sont pas optimisés. Cela peut poser un problème lorsque l'on transmet un panorama sur le réseau. C'est la raison pour laquelle, certains logiciels gèrent la multi-résolution en utilisant plusieurs panoramas complets de la même scène avec des niveaux de résolutions différents. Lors d'une transmission sur Internet, cette solution permet rapidement d'obtenir une première représentation de la scène à basse résolution de façon à ce que l'utilisateur puisse naviguer. Les résolutions plus élevées arrivant ensuite au gré du débit de la connexion, mais uniquement pour les zones observées. [Rob09]

1.6 État de l'Art

La détection et l'analyse des objets en mouvement est actuellement l'un des domaines de recherches les plus actifs dans les systèmes de traitement de la vidéo, l'extraction de l'arrière-plan représente souvent la première tâche dans ces systèmes. La performance de l'extraction de l'arrière-plan dépend principalement des algorithmes de modélisation de cet arrière-plan.

Dans la littérature, ces algorithmes de modélisation peuvent être classés en deux grandes catégories : celles basées sur les pixels et celles qui sont basées sur les blocs.

Dans la première catégorie, une fonction de distributions de probabilité est utilisée pour modéliser un pixel, les premières méthodes se sont orientées vers une modélisation par une gaussienne unique par pixel. Pour tenir compte de l'aspect multimodal, Friedman et Russell [FR97] sont parmi les premiers à proposer un modèle de mélange de gaussiennes qui étend le modèle seul gaussien en utilisant plus d'une distribution gaussienne par pixel pour améliorer l'extraction de l'arrière-plan dans les systèmes de surveillance de la circulation.

Puis ce modèle de mélange de gaussiennes est modifié par Stauffer et Grimson [SG99], qui utilise une approximation de K-means en ligne pour mettre à jour les paramètres du modèle de mélange de gaussiennes.

Dans la deuxième catégorie, ce sont des méthodes basées sur les blocs, ces méthodes utilisent principalement les caractéristiques des blocs indépendants ou légèrement chevauchés, Helkila et Pietikainen [HP06] ont proposé une méthode basée sur les patterns binaire local (LBP) pour créer un modèle d'arrière-plan. Les opérateurs LBP ont la capacité de tolérer les changements de luminosité et ont d'excellentes performances dans de nombreuses applications.

Les méthodes de modélisation de l'arrière-plan cités ci-dessus sont utilisées dans les systèmes qui fonctionnent avec des caméras fixe, de nombreux chercheurs ont essayé d'étendre ces méthodes de modélisation pour les adapter sur les caméras Pan-tilt-zoom (PTZ).

Les stratégies adoptées dans la littérature pour la modélisation de l'arrière-plan dans une vidéo issu d'une camera PTZ peuvent être divisées en trois grandes catégories : Frame-to-frame, frame-to-background et frame-to-global.

La méthode frame to frame utilise les informations de la zone de recouvrement entre un frame observé et ses frames précédentes. L'avantage de cette méthode est que leur temps d'apprentissage est réduit.

Kang et Paik [KPK⁺03] présentent un algorithme de génération d'arrière-plan adaptative qui utilise une méthode de transformation géométrique. Dans cet algorithme et lorsqu'une seule caméra stationnaire est utilisée, chaque emplacement d'un pixel fixe est invariant avec le temps. La figure 1.14 montre que l'emplacement d'un pixel qui est représenté par un rectangle noir change en fonction du mouvement de la caméra.

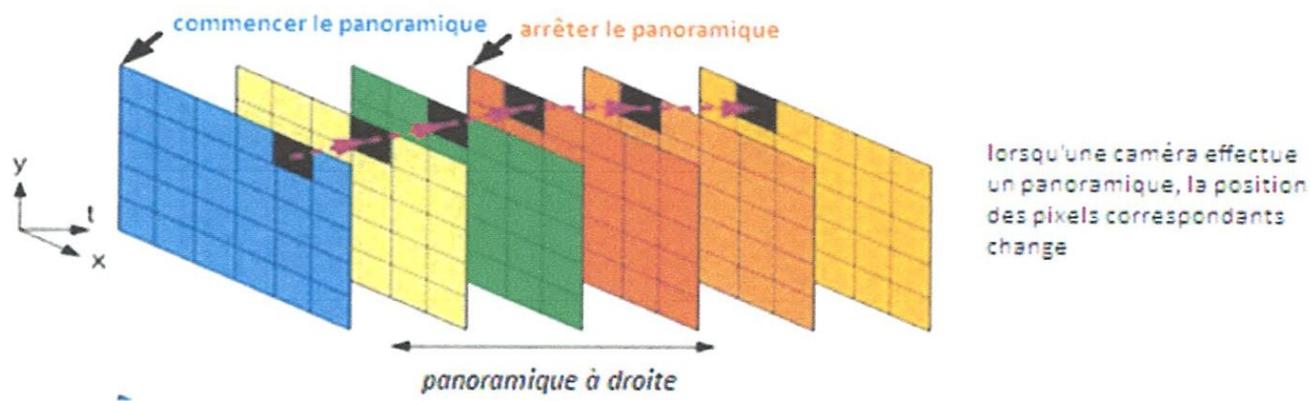


FIGURE 1.14 – Changement de l'emplacement d'un pixel dans une image panoramique.

Deux étapes sont utilisées dans cet algorithme pour enregistrer des images avec différents angles de panoramique et d'inclinaison, l'une consiste à utiliser l'algorithme à 8 points pour calculer la matrice homogène et l'auteur a utilisé l'angle de panoramique / inclinaison et une distance focale.

L'algorithme de 8 points nécessite au moins 4 points correspondants à partir de deux images, et ces points sont utilisés pour calculer la relation entre les images, qui est une matrice homogène. La première tâche de cette méthode consiste à choisir des angles de panoramique et d'inclinaison appropriés pour le suivi. Ces angles doivent être sélectionnés pour garantir un recouvrement d'au moins 50 pourcent en fonction du rapport de zoom, car nous devons suivre les objets en continu. Après avoir sélectionné ces angles, nous pouvons calculer la matrice homogène, qui décrit une relation entre une paire d'images qui ont des angles de panoramique et d'inclinaison différents. La matrice homogène calculée est utilisée pour projeter l'arrière-plan calculé dans la nouvelle image qui a des nouveaux angles de panoramique / inclinaison. Le système met à jour son arrière-plan, lequel est généré en utilisant un modèle gaussien, comme la caméra PTZ tourne. L'avantage de cette méthode est que nous n'avons pas besoin de connaître les paramètres internes de la caméra

PTZ, tels que la taille du capteur CCD (charge-coupled device) et la distance focale, mais nous devons calculer des matrices homogènes pour chaque combinaison possible d'angles panoramiques et d'inclinaisons. une autre limitation de ce système est qu'il est impossible de zoomer pour obtenir une vue de haute qualité de la cible.

Michelsoni et Forestidans [FMS⁺05], introduisent la méthode Frame to background , leur système est capable de : a) compenser les changements de fond dus au mouvement de la caméra dans les séquences vidéo acquises par une caméra PTZ tournant sur le centre des lentilles, b) détecter et suivre les objets mobiles dans des scènes réelles. La méthode proposée est basée sur les caractéristiques de l'image. Un algorithme de détection de changement image par image a été utilisé pour calculer la transformation de fond requise et pour modéliser le mouvement induit, ils ont utilisé une transformation translationnelle. Selon les auteurs les innovations de cette méthode sont : (a) estimer de manière robuste le déplacement intervenant entre deux trames consécutives et (b) l'accélération de la tâche pour la maintenance d'un ensemble fiable de fonctionnalités.

Les auteurs témoignent que leur méthode obtient une bonne estimation du déplacement démontré par le faible pourcentage d'erreurs réalisées. De plus, l'utilisation d'une carte de fonctionnalités pour la maintenance d'un ensemble de fonctionnalités fiables a permis au système de fonctionner en temps réel grâce à une plus faible complexité du processus d'extraction pour de nouvelles fonctionnalités. Ils affirment que le module de suivi a permis un suivi continu d'un objet à une vitesse modérée, même brièvement occulté par d'autres objets.

Les méthodes Frame to-global utilisent un panorama ou une caméra ajoutée pour fournir des informations spatiales globales pour la détection et le suivi de premier plan . Sinha et Pollefy [SPM04] utilisent un certain nombre de frames haute résolution à différentes échelles pour coudre un panorama pour chaque échelle et calibrer tous les panoramas ensemble afin d'obtenir des informations détaillées sur le FOV (field of view, en français « champ de vision ») de la caméra, mais leur méthode ne fournit pas un arrière-plan robuste pour la détection de premier plan puisque le panorama à chaque échelle peut inclure des objets mobiles. Chen et Yao [CYP⁺08] utilisent une caméra supplémentaire pour étendre le champ de vision dans leur système de surveillance, mais un petit objet est difficile à détecter dans les images à basse résolution lorsque la caméra supplémentaire capture tout l'environnement, en particulier dans une grande zone . Bien que ces méthodes fournissent des informations globales pour répondre au changement de rotation du PTZ, elles ne fonctionnent pas lorsque l'échelle du PTZ change rapidement.

1.7 Conclusion

Dans ce chapitre nous avons présenté un petit résumé sur la vidéo surveillance avec un petit historique ainsi que les applications de ce système, nous avons présenté aussi le document vidéo et ses caractéristiques avec les différents formats d'un vidéo numérique, et une brève présentation sur les images avec les différents espace de couleur , puis nous allons terminer ce chapitre par un aperçu détaillé sur les images panoramiques.

Chapitre 2

Modélisation de l'arrière-plan

2.1 Introduction

Ce chapitre constitue le cœur de notre système, nous allons aborder les différentes étapes nécessaires pour la détection des objets en mouvement dans un flux vidéo extrait à partir d'une caméra en mouvement qui suit une trajectoire linéaire. Il compose de plusieurs modules, Chaque module est un processus qui permet de réaliser une tâche spécifique. Nous allons d'abord aborder par la fragmentation de la vidéo, ensuite, nous sélectionnons les images couvrant toute la scène pour construire l'image panoramique qui représente le modèle statique de la scène complète. Puis nous détaillons les différentes étapes d'extraction de l'arrière-plan et la détection des objets dans cette scène étendue.

La figure 2.1 présente l'architecture de notre système :

2.2 Acquisition de la vidéo

Nous avons regroupé dans cette partie tous les traitements liés à l'acquisition d'une vidéo à partir d'une caméra mobile.

Généralement, le terme « acquisition » désigne la capture de la scène ou de l'environnement via une caméra (dans notre cas : une caméra mobile). La capture est effectuée a priori à cause de plusieurs contraintes techniques. L'aspect sécurité est un problème que nous avons rencontré lors de l'installation des caméras pour l'expérimentation.

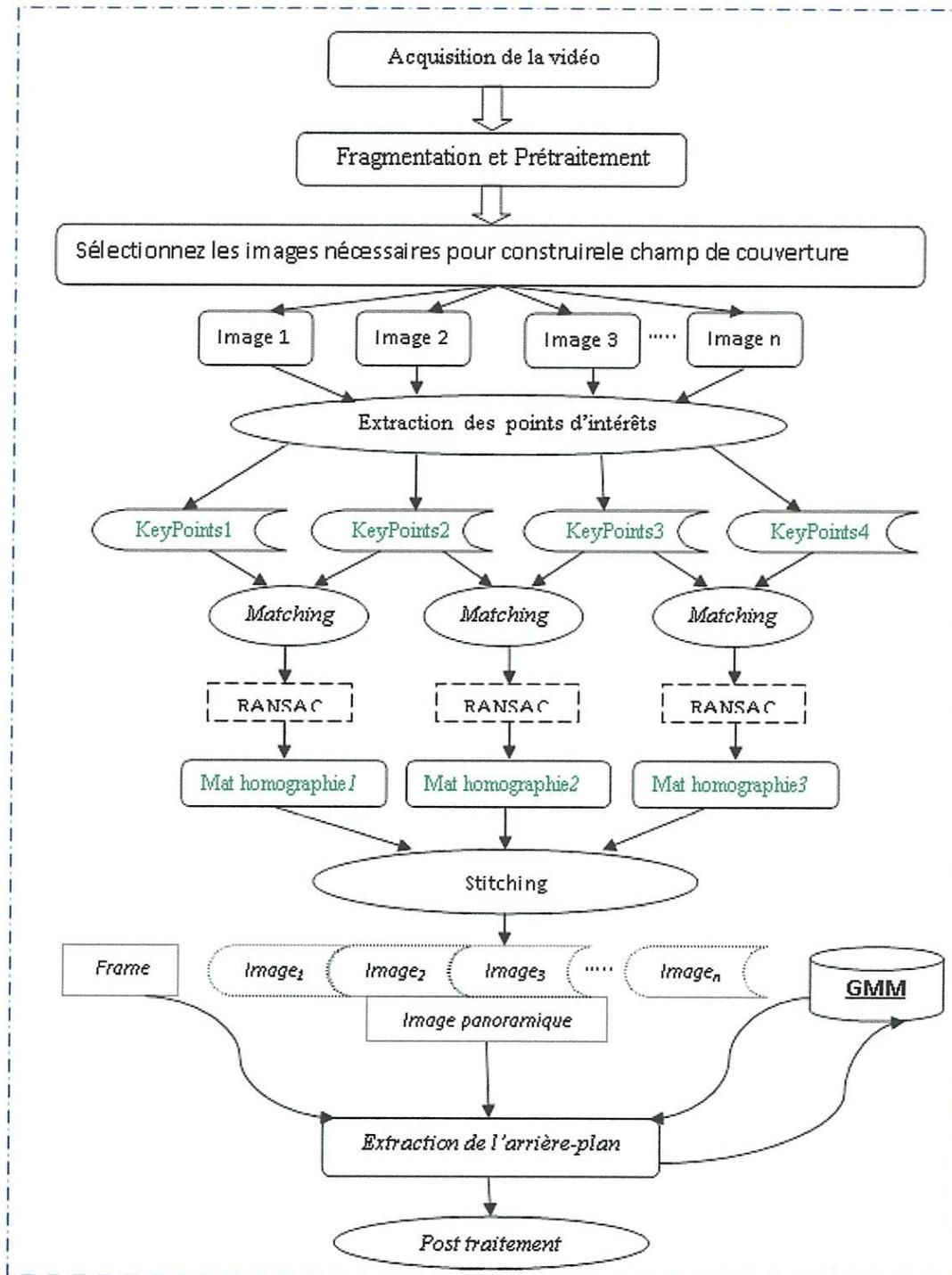


FIGURE 2.1 – L'architecture du système.

2.3 Fragmentation et Prétraitement

Nous avons appliqué dans cette partie une fragmentation qui permet de transformer la vidéo capturée en un ensemble de frames ou chaque frame représente une séquence de la vidéo. Ensuite,

un filtre médian est appliqué afin de corriger les erreurs de capture. Le filtre médian est réputé d'être très simple à implémenter, rapide et efficace pour rendre l'image uniforme.

La fragmentation se fait à l'aide de l'algorithme suivant :

NB-Image = (temps exploité par la vidéo) / (temps exploité par une image);

Tanque (i <= NB-Image) faire

Capturer l'image Ni;

Mettre l'image dans un buffer;

Convertir le buffer en une image;

Enregistrer l'image;

Fin Tq

2.4 Construction du champ de couverture

2.4.1 Présentation

Il existe plusieurs types de reconstruction d'une image panoramique à partir de plusieurs images. Les mosaïques les plus simples sont créées à partir d'un ensemble d'images dont les déplacements sont des translations « image-plan ». D'autres mosaïques sont créées en tournant la caméra autour d'un centre optique (c'est notre cas) à l'aide d'un dispositif spécial et créant une image panoramique qui représente la projection de la scène sur un cylindre.

Cette reconstruction passe par trois étapes, la première étape est l'utilisation d'un détecteur de points d'intérêts qui permet d'extraire sur une image, les coordonnées des points caractéristiques.

La deuxième étape est la mise en correspondance (Matching), les points ainsi localisés sont décrits à l'aide de descripteurs de vecteurs de caractéristiques qui permettent ensuite d'appareiller les points d'intérêts de plusieurs prises de vue distinctes d'une même scène.

Dans la troisième étape et après la mise en correspondance des vecteurs de caractéristique, nous obtenons la matrice de transformation en utilisant RANSAC "RANdom SAmple Consensus". Il s'agit d'une méthode itérative pour estimer les paramètres d'un modèle mathématique à partir

d'un ensemble de données observées. Sachant que la matrice de transformation est connue, les deux images peuvent être facilement alignées ensemble.

2.4.2 Détection des points d'intérêts

Si l'on souhaite créer une image panoramique à partir de deux images, la première étape consiste à extraire des points caractéristiques (features en anglais) dans les deux images. Le principe est d'extraire d'une image des points caractéristiques qui ont une certaine singularité et sont particulièrement porteuses d'information. Les points que l'on cherche à extraire peuvent être : coins, contours ou régions.

Pour trouver les points caractéristiques dans une image, on doit appliquer un algorithme de détection de points d'intérêts (détecteur).

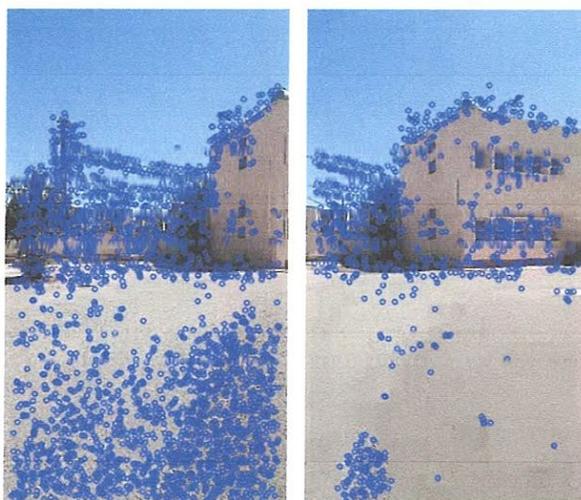


FIGURE 2.2 – Détection des points d'intérêts.

Divers facteurs peuvent en effet modifier ou perturber les résultats du détecteur des points d'intérêts :

- le bruit de l'image.
- les conditions d'illumination.
- la position de la caméra ou de l'appareil photo par rapport à l'objet.

Il existe plusieurs détecteurs des points caractéristiques, parmi lesquels :

- Le détecteur de Harris-Laplace
- Le détecteur de SIFT.

- Le détecteur de SURF.

Nous avons choisi le descripteur de SURF, ce descripteur est comme SIFT , peu sensible au changement d'intensité, de mise à l'échelle et de rotation, ce qui fait de lui un descripteur très robuste, Le principal avantage de ce descripteur par rapport à SIFT est que l'extraction des points d'intérêts se fait plus rapidement.

2.4.3 La mise en correspondance

On cherche à savoir quels sont les points de deux images différentes qui peuvent être mis en correspondance. Cette étape précède la fusion de deux images.

Pour savoir si deux points sont en correspondance, on utilise des scores. Les scores permettent en outre de trier les pixels candidats dans l'ordre de préférence, ce qui rend possibles des stratégies globales d'appariement des points des deux images.

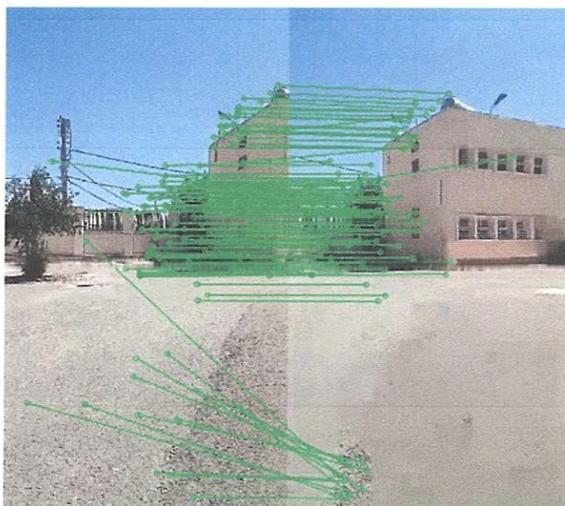


FIGURE 2.3 – La mise en correspondance.

La méthode la plus simple utilise la SSD (Sum of Square intensity Difference), c'est-à-dire la somme des différences d'intensité lumineuse des pixels situés au voisinage de chaque point au carré. Pour un point M_1 de la première image et un point M_2 de la seconde, cette quantité se calcule donc ainsi :

$$SSD = \sum_{i=-N}^N \sum_{j=-N}^N [I_1(M_1 + (i, j)) - I_2(M_2 + (i, j))]^2$$

Où :

- N caractérise la taille des voisinages autour des points que l'on considère.
- I_1 et I_2 représentent les fonctions d'intensité des deux images.

2.4.4 Estimation de la matrice de transformation (la matrice homographie)

Une homographie est une matrice 3×3 qui possède donc 9 éléments indépendants. Cette matrice est utilisée en coordonnées homogènes, elle est donc invariante par rapport au changement d'échelle et on peut alors la multiplier par une constante réelle non nulle sans changer ses effets sur une image.

Une homographie a donc 8 degrés de liberté et elle est déterminée par au minimum 4 points de correspondances de (qui ont chacune 2 degrés de liberté).

Pour calculer une homographie, il suffit de calculer la matrice \mathbf{H} telle que : $x'_i = Hx_i$ pour chaque couple de points : $x_i \longleftrightarrow x'_i$

où les x_i de l'image de départ correspondent aux x'_i de l'image d'arrivée.

Pour résoudre $x'_i = Hx_i$, il vaut mieux partir de $x'_i \wedge Hx_i = 0$ (\wedge correspond au produit vectoriel). En séparant les lignes de H , on obtient :

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} \wedge \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

2.4.5 L'algorithme RANSAC

Cet algorithme sert à consolider la méthode de mise en correspondance des points.

Nous avons remarqué que cette méthode pouvait associer des points qui ne correspondent pas en réalité (outliers). Plutôt que d'affiner la méthode, on se sert de l'homographie pour détecter ces points aberrants.

L'algorithme fonctionne de la façon suivante :

Répéter

- Sélectionnez quatre paires de points de caractéristiques (au hasard) parmi les couples de points détectés.
- On cherche l'homographie associée à ces points (exacts)

- On calcule une erreur associée à l'homographie et on regarde si elle minimise l'ensemble des erreurs trouvées pour les autres homographies

Jusqu'à (N fois)

2.4.6 Détermination des dimensions de l'image panoramique

La détermination des dimensions de l'image panoramique consiste à déterminer une zone de recouvrement entre toutes les images. Les images doivent tout d'abord être exprimées dans le même repère : celui de la première image. Pour cela, on applique à chaque image les homographies calculées à l'étape précédente. Une fois ces images obtenues, il nous a fallu trouver un moyen de déterminer une zone de recouvrement. Nous avons choisi de nous baser sur les coins des images.

L'idée consiste à déterminer quatre entiers : limH , limB , limG et limD qui représente respectivement la limite haute de la zone, la limite basse, la limite gauche et la limite droite.

La détermination de ces entiers se fait de la façon suivante :

- Pour trouver limH , on considère tous les coins supérieurs des images et on détermine celui qui a l'ordonnée maximale : limH est égal à cette ordonnée.
- Pour trouver limB , on considère tous les coins inférieurs des images et on détermine celui qui a l'ordonnée minimale : limB est égal à cette ordonnée.
- Pour trouver limD , on regarde les abscisses des deux coins de la dernière image. limD est égal au minimum des deux.
- Enfin, on a $\text{limG} = 0$.

Essaye de redessiner l'image!!

2.4.7 Remplissage de l'image panoramique

Une fois l'homographie H calculée, il nous suffit, pour transformer une image, d'appliquer cette homographie à chaque pixel de l'image, à travers les deux formules suivantes :

$$u_2 = \frac{m_0 \cdot u_1 + m_1 \cdot v_1 + m_2}{m_6 \cdot u_1 + m_7 \cdot v_1 + 1}, \quad v_2 = \frac{m_3 \cdot u_1 + m_4 \cdot v_1 + m_5}{m_6 \cdot u_1 + m_7 \cdot v_1 + 1}$$

où x_1 est un point de l'image défini par ses coordonnées homogènes $(u_1, v_1, 1)$ dans l'image 1, et x_2 sa projection homographique dans l'image 2 est défini par ses coordonnées homogènes $(u_2, v_2, 1)$.

Cette méthode peut toutefois poser certains problèmes : l'image transformée n'occupe pas forcément tout l'espace de l'image d'arrivée.

Comme on peut le voir sur l'image ci-dessous, certains points de l'image finale n'ont pas d'antécédents dans l'image de départ.



FIGURE 2.4 – Remplissage de l'image panoramique.

Une solution consiste à calculer $x_i = H^{-1}\hat{x}_i$ plutôt que : $\hat{x}_i = Hx_i$.

Pour cela, après avoir inversé H , il suffit de parcourir l'image d'arrivée, et calculer $x_i = H^{-1}\hat{x}_i$ et de voir si le pixel x_i appartient à l'image de départ. S'il est effectivement dans l'image de départ, le pixel \hat{x}_i de l'image finale prend la même couleur que le pixel x_i de l'image de départ. On obtient alors :



FIGURE 2.5 – Assemblage et rectification.

2.5 Modélisation et l'extraction de l'arrière-plan

La modélisation et l'extraction de l'arrière-plan sont un problème important dans les systèmes automatiques de vidéo surveillance. Plusieurs travaux de recherche ont été proposés afin d'améliorer le rendement de cette opération. Effectivement, une bonne méthode d'extraction engendre une simplification des traitements ultérieurs permettant ainsi un gain en temps d'exécution et en espace mémoire.

Les GMM sont parmi de ces méthodes les plus utilisées pour la modélisation et l'extraction de l'arrière-plan. Les travaux de Hedayati et al. [HZH10] ont montré que les GMM offrent un bon compromis entre qualité d'extraction et temps d'exécution par rapport aux autres méthodes. Les GMM sont performants dans le cas où la vidéo ne contient pas de variation locale où un changement de luminosité.

2.5.1 Les mixture de gaussiennes

Stauffer et Grimson [SG99] ont présenté une nouvelle modélisation de l'arrière-plan basée sur une mixture de plusieurs gaussiennes. Les MOG ont prouvé leurs efficacités face aux changements de luminosité, les mouvements répétitifs, l'encombrement, l'introduction et le retrait des objets de la scène et les mouvements lents des objets. La motivation sur laquelle est basée la proposition de ce modèle est qu'un modèle uni modal de l'arrière-plan ne pouvait gérer le bruit d'acquisition d'image, le changement de lumière et l'existence de multitude de surfaces pour un pixel particulier dans le même temps.

Une MoG est un modèle statistique qui suppose que les données proviennent d'une somme pondérée de plusieurs distributions de gaussiennes. Stauffer et Grimson ont présenté une méthode GMM adaptative pour modéliser un fond dynamique dans des séquences d'images. Si les K distributions de gaussiennes sont utilisés pour décrire l'historique d'un pixel, l'observation d'un pixel donné sera forcément dans l'un des K états à la fois. K détermine la multi modalité de l'arrière-plan et la sélection de K est généralement basée sur la mémoire disponible et de la puissance de calcul. Stauffer et Grimson ont proposé de fixer K entre 3 et 5 [SG99].

2.5.2 L'image de référence

L'image de référence est une image du fond. Elle contient les parties statiques de la scène. Il faut d'abord la construire puis la réactualiser au cours du temps afin qu'elle s'adapte aux évolutions de la scène, dans notre cas c'est la première image panoramique créée précédemment.

2.5.3 Initialisation

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

$$P(X_t) = \sum_{i=1}^k w_{i,t} \times \eta(X_t, \mu_{i,t}, \sigma_{i,t}) \quad (2.1)$$

Avec les paramètres des gaussiennes :

- K : Le nombre de gaussiennes associées à chaque pixel.
- $w_{i,t}$: Le poids calculé pour chaque gaussienne i à l'instant t .
- $\mu_{i,t}$: La moyenne de la i^{eme} gaussienne.
- $\sigma_{i,t}$: L'écart-type de la i^{eme} gaussienne.
- $\eta(X_t, \mu_{i,t}, \sigma_{i,t})$: La densité de probabilité d'une gaussienne qui est donnée par la formule suivante :

$$\eta(X_t, \mu_{i,t}, \sigma_{i,t}) = \frac{1}{(2\pi)^{\frac{D}{2}} |\Sigma|^{\frac{1}{2}}} \exp^{-\frac{1}{2}(X_t - \mu)^T \Sigma^{-1}(X_t - \mu)} \quad (2.2)$$

Où D exprime la dimension du vecteur X .

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

2.5.4 Mise à jour des paramètres

De nombreuses méthodes sont disponibles pour calculer les paramètres d'un mélange de gaussiennes. La méthode classique est l'algorithme EM (Expectation Maximisation), qui comporte plusieurs variantes. Le principe de ces algorithmes est de calculer des paramètres inconnus (la

moyenne, la variance et les poids des gaussiennes), à partir des données observées et incomplètes (l'histogramme temporel de chaque pixel). L'idée de ces algorithmes est d'affecter un poids à chaque gaussienne du mélange (pour chaque pixel). Le poids d'une gaussienne correspond au nombre d'images où le pixel prend une valeur correspondant à cette gaussienne.

2.5.5 L'algorithme K-means

Stauffer et Grimson ont programmé une approximation de l'algorithme EM afin de gagner le temps. Leur méthode comporte uniquement deux paramètres [SG99] :

α : la constante d'apprentissage.

T : déterminant la proportion de données correspondantes au fond.

Dans leur application, l'algorithme K-means est suivi d'un algorithme d'étiquetage en composantes connexes, pour regrouper les pixels du premier plan en objets, et d'un algorithme de suivi à hypothèses multiples servant à suivre les objets avec un ensemble de filtres de Kalman.

2.5.5.1 Principe de l'algorithme

L'algorithme K-means est une approximation de l'algorithme EM. Il traite directement la nouvelle valeur de chaque pixel pour mettre à jour les paramètres du mélange de gaussiennes du pixel.

2.5.5.2 Cas d'une seule dimension

On considère désormais un pixel, d'historique

$$\{X_1, X_2, \dots, X_t\} = \{I(x, y, i), 1 \leq i \leq t\}. \quad (2.3)$$

Pour chaque nouvelle valeur X_i de ce pixel, on cherche la gaussienne du mélange qui correspond au mieux à cette valeur, c'est-à-dire la i^{eme} gaussienne telle que :

$$|\mu_i - X_t| \leq 2.5 * \sigma_i \quad (2.4)$$

On obtient une correspondance avec un seuillage par pixel et par distributions, c'est-à-dire un seuillage adapté au mélange de gaussiennes de chaque pixel et où la correspondance avec une gaussienne dépend de la moyenne et de l'écart type de cette gaussienne. Afin d'éviter le risque de

double correspondance, on prend la gaussienne qui minimise le rapport :

$$\frac{|\mu_i - X_t|}{\sigma_i} \quad (2.5)$$

Ainsi, on est sûr de prendre la gaussienne γ qui correspond.

On met à jour sa moyenne et son écart-type (les paramètres des autres gaussiennes restent inchangés) selon les équations (2.6) et (2.7). Ces équations sont valables pour le cas où les paramètres sont des scalaires pour une vidéo en niveaux de gris.

Dans le cas d'un espace de couleur avec μ_t et X_t des vecteurs à trois composantes (R, V, B) et σ_t^2 qui devient \sum_t , la matrice de covariance de dimension 3×3 .

$$\mu_t = (1 - \rho) \mu_{t-1} + \rho X_t \quad (2.6)$$

$$\sigma_t^2 = (1 - \rho) \sigma_{t-1}^2 + \rho (X_t - \mu) T (X_t - \mu) \quad (2.7)$$

$$\rho = \alpha \eta (X_t / \mu_i, \sigma_i) \quad (2.8)$$

Où : $\eta (X_t / \mu_i, \sigma_i)$ est la distribution de la i^{eme} gaussienne dans le cas unidimensionnel. Dans les formules de mise à jour des paramètres μ et σ , la variable ρ est une deuxième constante du temps, dérivée de α .

Si aucune correspondance n'est trouvée, ce qui est le cas au début de l'algorithme, on remplace la gaussienne de poids le plus faible par une gaussienne de moyenne X_t , avec un écart-type élevé et un poids faible. Ensuite, on met à jour les poids de la façon suivante :

$$\omega_{i,t} = (1 - \alpha) \omega_{i,t-1} + \alpha M_{i,t}$$

Avec $M_{i,t} = 1$ pour la gaussienne qui correspond et 0 pour les autres. Enfin, on renormalise les poids en divisant chaque $\omega_{i,t}$ par la somme des poids.

2.5.5.3 Fonctionnement de l'algorithme K-means

L'initialisation des paramètres n'était pas détaillée chez Chen et al [CSZS02], ils parlent d'un poids faible et un écart-type élevé. Nous avons donc choisi d'initialiser les gaussiennes au fur et à mesure. On fait tout de même une première initialisation des gaussiennes pour éviter tout problème :

- $\omega_{i,t=0} = \frac{1}{N}$
- $\sigma_i = \frac{(0.001+random(256))}{255}$
- $\mu_i = \frac{(0.001+random(256))}{255}$

Ainsi, on initialise les gaussiennes au fur et à mesure, et si on dispose de 5 gaussiennes, il se peut que l'on utilise que 2 ou 3 la plupart du temps. Cet aspect de l'algorithme K-means est très intéressant. Lorsque toutes les gaussiennes sont utilisées et que la nouvelle valeur du pixel ne correspond à aucune des gaussiennes, on remplace la gaussienne de poids le plus faible par une gaussienne initialisée de la même façon que ci-dessus.

Un autre avantage de la mise à jour des paramètres de la gaussienne qui correspond à la nouvelle valeur du pixel est que si un objet s'arrête suffisamment longtemps pour être intégré dans le fond, la gaussienne décrivant l'état précédent du fond existe toujours jusqu'à ce qu'elle devienne la gaussienne du poids le plus faible. Si l'objet se remet en mouvement, on peut alors retrouver très rapidement l'état précédent du fond.

2.5.5.4 Construction du masque

L'étape finale de l'extraction de l'arrière-plan consiste simplement à faire la soustraction entre l'image courante de la vidéo et l'image de référence. L'image résultante de la différence est ensuite seuillée afin d'obtenir un masque binaire : $\forall (x, y) \in [largeur, hauteur]$

$$\begin{cases} si & |I(x, t) - I_{ref}(x, t)| < T \quad , alors \quad Masque(x, y, t) = 0 \\ si & |I(x, t) - I_{ref}(x, t)| > T \quad , alors \quad Masque(x, y, t) = 255 \end{cases} \quad (2.9)$$

Avec $I(x, t)$ l'image courante, $I_{ref}(x, t)$ l'image de référence et $Masque(x, y, t)$ l'image binaire obtenue par le seuillage T .

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

2.5.6 Application des opérations morphomathématiques

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

Dans le cas où nous appliquons ces opérations dans l'ordre inverse, qui est l'érosion suivie d'une dilatation, les régions de bruits isolés seront supprimées, mais cette fois, nous ne serions pas en mesure de fermer les trous qui se trouvent à l'intérieur des objets. Après avoir expérimenté différentes combinaisons de ces opérations, nous nous sommes fixé à la séquence suivante : deux niveaux d'érosion suivie par un niveau dilatation.

2.6 Conclusion

Dans ce chapitre, nous avons présenté l'architecture générale de notre système ainsi que les différents modules qui le composent. Nous avons pris le soin de détailler chaque processus afin de montrer notre apport à travers les idées implanté dans presque tous les niveaux du système proposé.

Chapitre 3

Tests et résultats

3.1 Introduction

Pour valider la méthode proposé, une étape d'analyse de teste est obligatoire, cet analyse nous permettra d'évaluer les avantages et les inconvénients de cette méthode, et détecter les points forts et les points faibles des algorithmes utilisés.

Afin d'expérimenter le système, nous allons d'abord implémenter les algorithmes et les modules proposé. Cette implémentation est réalisée à l'aide d'un langage de programmation très connu et une plateforme bien spécifique et des autres bibliothèques spécialisée, l'expérimentation nécessite également la présence d'une base de teste à travers lequel nous allons mesurer la performance du notre système.

3.2 Outils utilisés

3.2.1 Java

Java est un langage de programmation et une plate-forme informatique qui ont été créés par Sun Microsystems en 1995. Beaucoup d'applications et de sites Web ne fonctionnent pas si Java n'est pas installé et leur nombre ne cesse de croître chaque jour. Java est rapide, sécurisé et fiable. Des ordinateurs portables aux centres de données, des consoles de jeux aux superordinateurs scientifiques, des téléphones portables à Internet, la technologie Java est présente sur tous les fronts .

3.2.2 Eclipse

Eclipse est un IDE, Integrated Development Environment (EDI environnement de développement intégré en français), c'est-à-dire un logiciel qui simplifie la programmation en proposant un certain nombre de raccourcis et d'aide à la programmation. Il est développé par IBM, est gratuit et disponible pour la plupart des systèmes d'exploitation [Com17].

Au fur et à mesure que vous programmez, eclipse compile automatiquement le code que vous écrivez, en soulignant en rouge ou jaune les problèmes qu'il détecte. Il souligne en rouge les parties du programme qui ne compilent pas, et en jaune les parties qui compilent mais peuvent éventuellement poser problème (on dit qu'eclipse lève un avertissement). Pendant l'écriture du code, cela peut sembler un peu déroutant au début, puisque tant que la ligne de code n'est pas terminée (en gros jusqu'au point-virgule), eclipse indique une erreur dans le code [Com17].

3.2.3 OpenCV

OpenCV (Open Source Computer Vision Library) est une bibliothèque de logiciels de vision artificielle et d'apprentissage automatique. OpenCV a été construit pour fournir une infrastructure commune pour les applications de vision par ordinateur et pour accélérer l'utilisation de la perception de la machine dans les produits commerciaux. En tant que produit sous licence BSD, OpenCV facilite l'utilisation et la modification du code par les entreprises [tea18].

La bibliothèque dispose de plus de 2500 algorithmes optimisés, qui comprennent un ensemble complet d'algorithmes de vision artificielle et d'apprentissage automatique classiques et à la pointe de la technologie. Ces algorithmes peuvent être utilisés pour détecter et reconnaître des visages, identifier des objets, classer des actions humaines, suivre des mouvements de caméra, suivre des objets en mouvement, extraire des modèles 3D d'objets, produire des nuages de points 3D à partir de caméras stéréo image d'une scène entière, trouver des images similaires à partir d'une base de données, supprimer les yeux rouges des images prises en flash, suivre les mouvements des yeux, reconnaître les paysages et établir des marqueurs pour les incruster avec réalité augmentée, etc. OpenCV a plus de 47 000 personnes communauté et le nombre estimé de téléchargements dépassant 14 millions. La bibliothèque est largement utilisée dans les entreprises, les groupes de recherche et les organismes gouvernementaux [tea18].

3.2.4 Jama

JAMA est un paquet d'algèbre linéaire de base pour Java. Il fournit des classes au niveau de l'utilisateur pour construire et manipuler des matrices réelles et denses. Il est destiné à fournir une fonctionnalité suffisante pour les problèmes de routine, emballés de manière naturelle et compréhensible pour les non-experts. Il est destiné à servir de classe matricielle standard pour Java, et sera proposé en tant que tel au Java Grande Forum puis à Sun [oST12].

JAMA est basé sur une classe matricielle unique dans un cadre strictement orienté objet [oST12].

3.3 Base de tests

La non-présence d'une base de tests publique qui traite la problématique de la vidéo surveillance pour les caméras panoramique nous a obligés à créer notre propre base de tests. Pour cela, nous avons créé trois vidéos prises dans des conditions saléatoires afin de bien mesurer la performance de notre système dans différents contextes. Pour réaliser cette base nous avons utilisé une caméra de type Condor, avec 13 mégapixels qui tourne autour d'un axe. Le tableau 3.1 présente une description des vidéos utilisés dans les tests.

Nom du vidéo	Duré	Nombre de frame	Résolution	Description
VideoWilaya	1min et 13s	2210	368X656	Cette vidéo est prise a l'intérieur du siège de la wilaya de Guelma (espace de stationnement de véhicules). .
VideoMaisonDeCulture	1m et 6s	1939	368X656	Cette vidéo est prise face a la maison de culture ou il y a une route double voie, et un espace publique.
VideoDepartementInfo	1m	1500	368X656	Cette vidéo est prise a l'intérieur du département informatique.

TABLE 3.1 – Description de la base de tests.

3.4 Tests et résultats

3.4.1 Résultat de l'algorithme de fragmentation

Après l'exécution de la méthode de fragmentation qui permet de découper un vidéo en séquences d'images on obtient le résultat suivant :

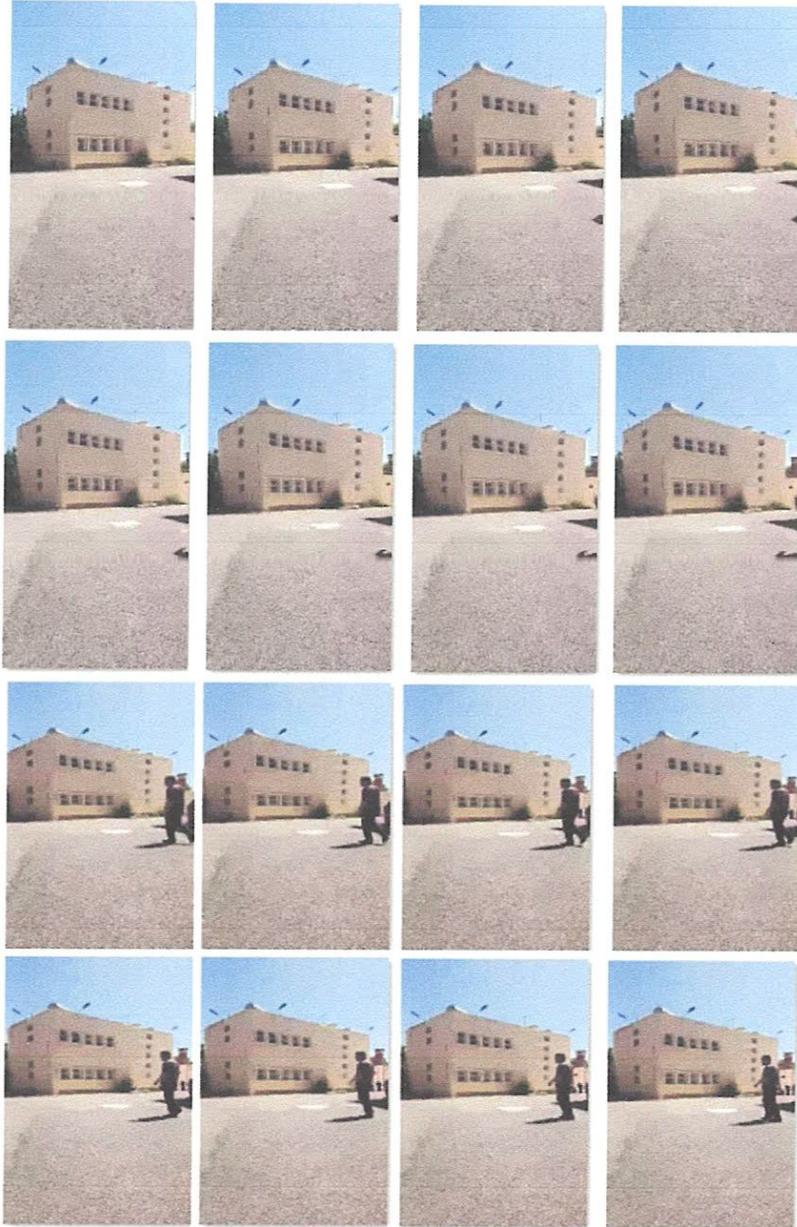


FIGURE 3.1 – Vidéo fragmentée.

3.4.2 Résultat de la construction de l'image panoramique (l'image de référence)

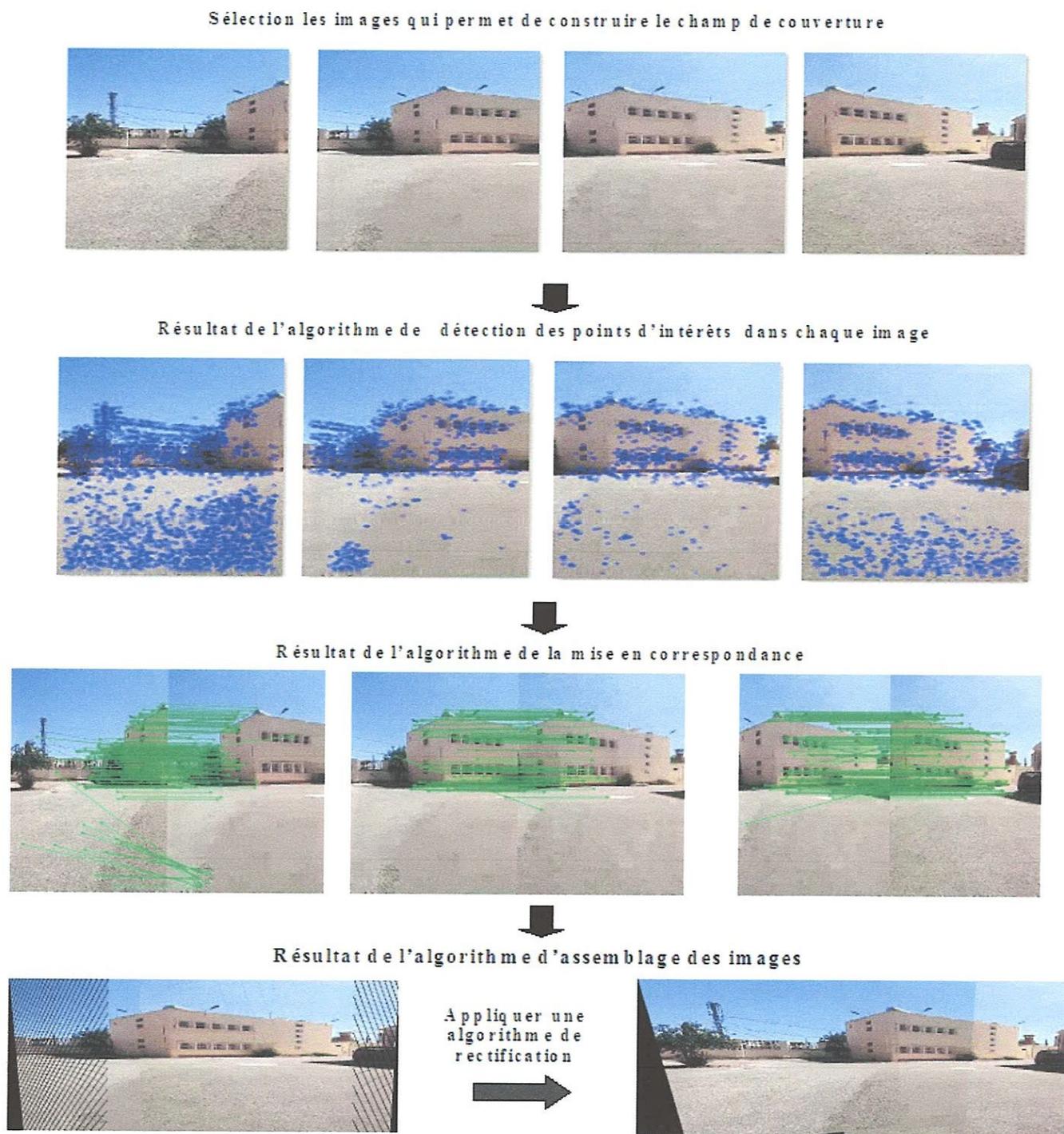


FIGURE 3.2 – Construction de l'image panoramique .

3.4.3 Résultat de l'algorithme de mise à jour d'une image panoramique

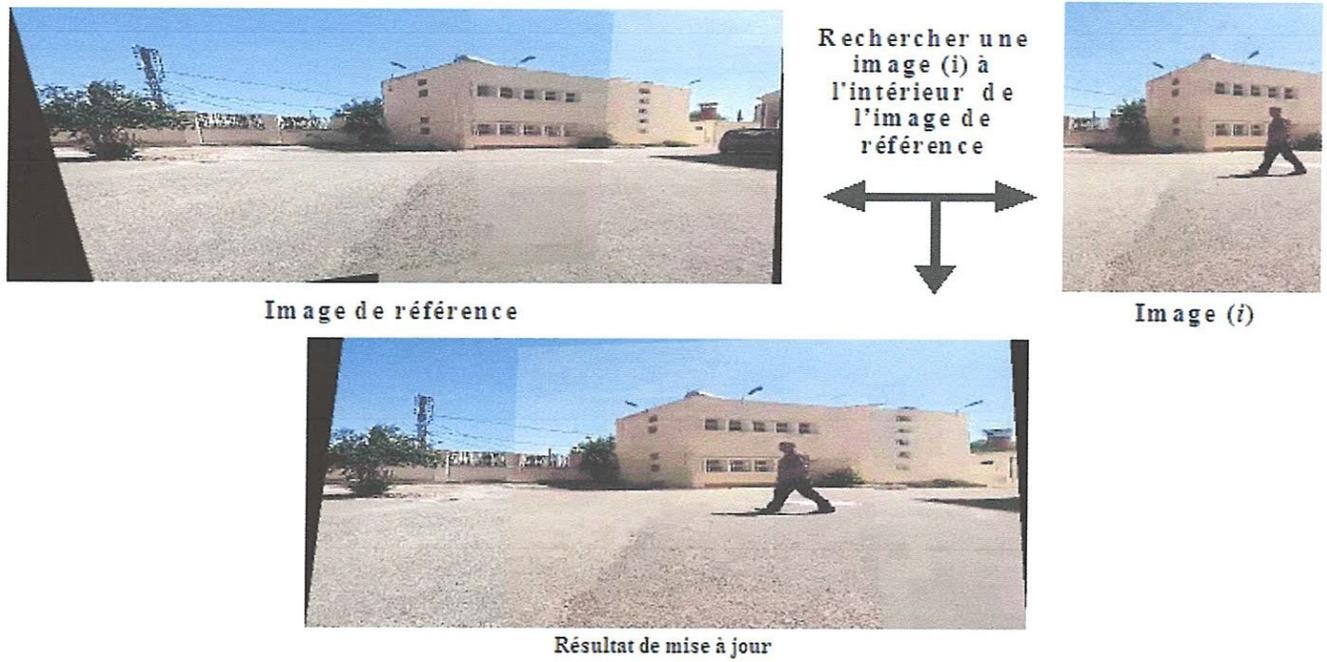


FIGURE 3.3 – Mettre à jour l'image panoramique .

3.4.4 Résultat de l'algorithme d'extraction de l'arrière-plan (GMM)

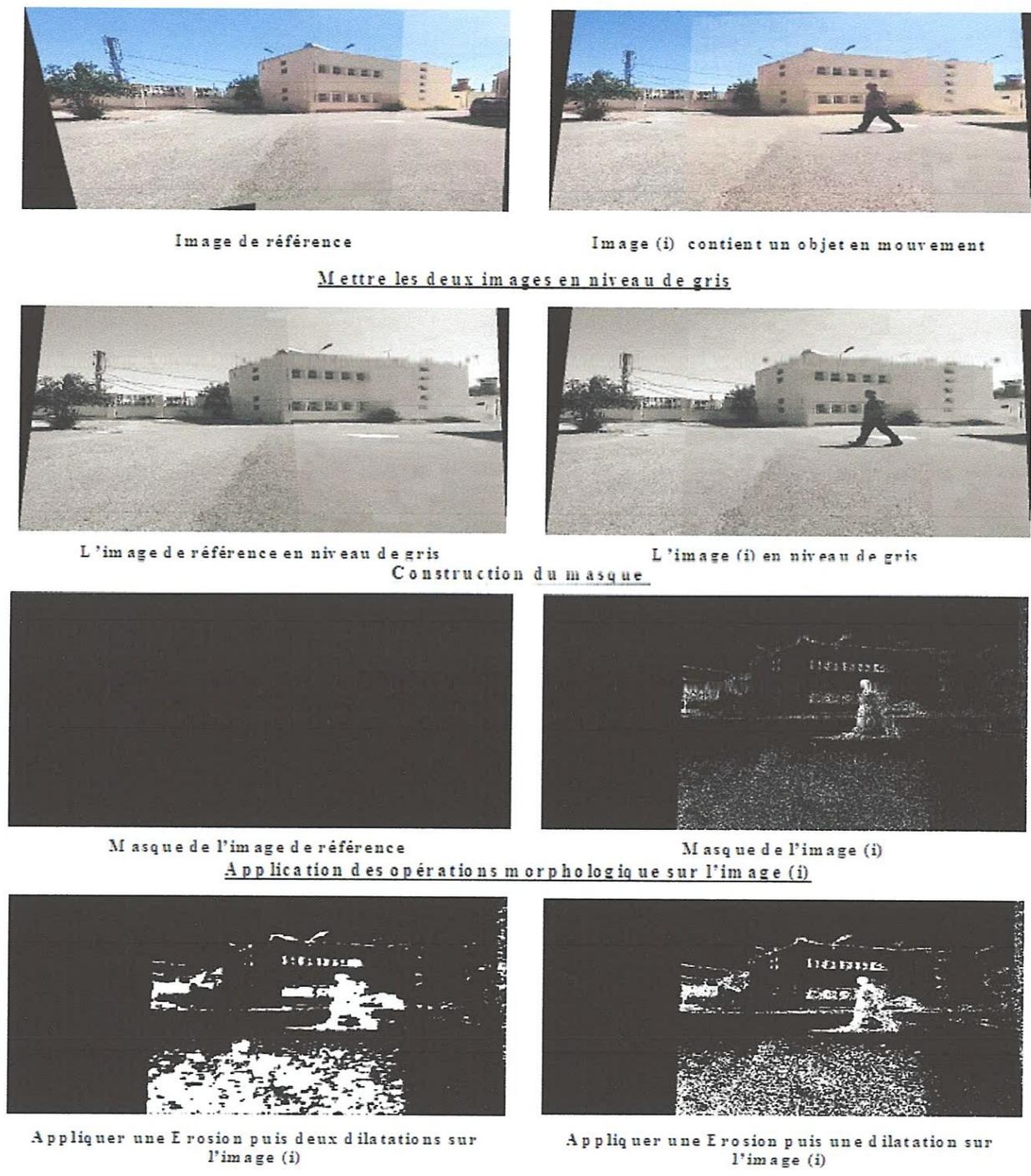


FIGURE 3.4 – L'extraction de l'arrière-plan .

3.4.5 Affichage du résultat sur une séquence d'images (sur la même base)

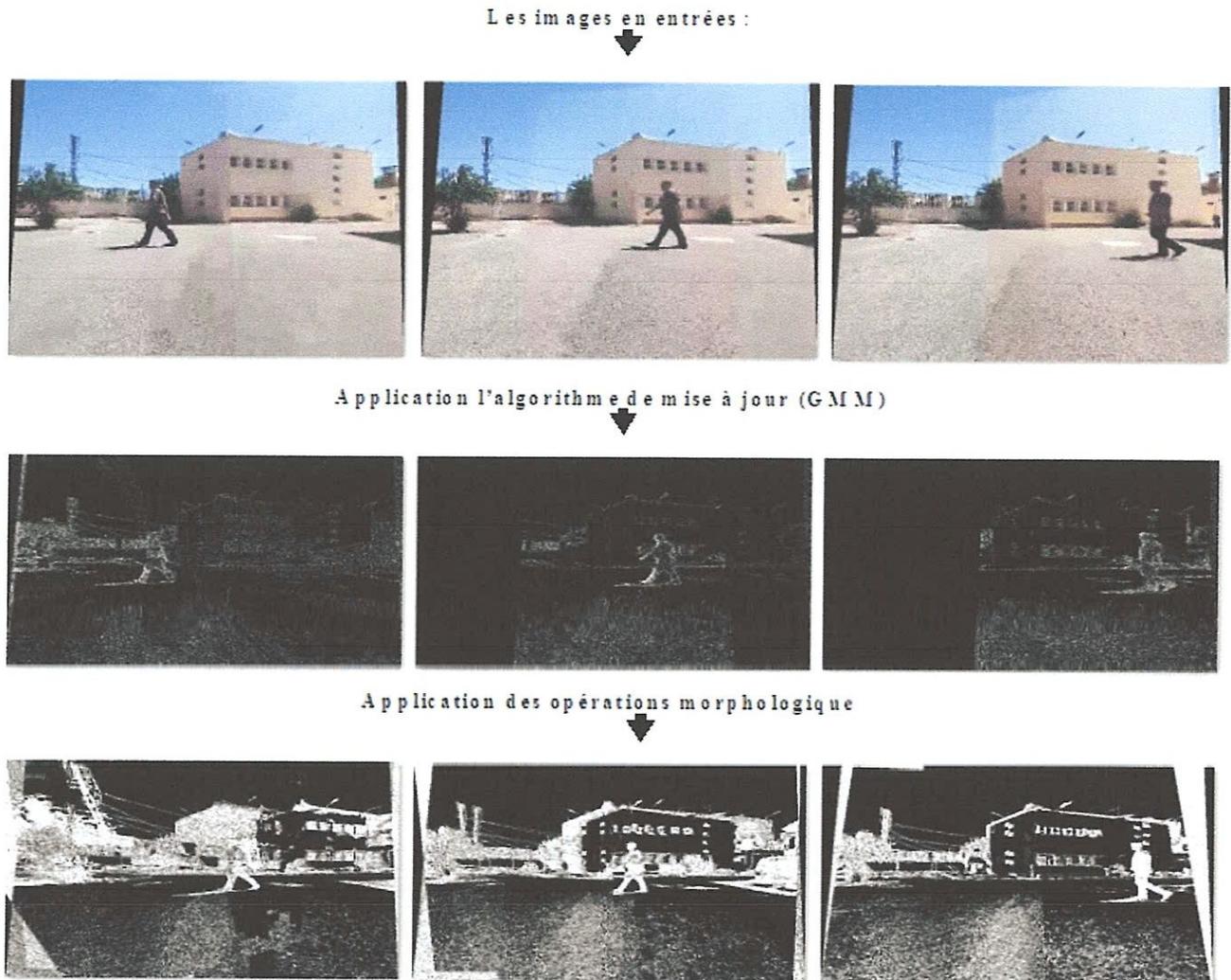


FIGURE 3.5 – Résultat sur une séquence d'images .

3.4.6 4.6. Affichage du résultat sur une autre base (au niveau de maison de culture 'Mohamed chafai')

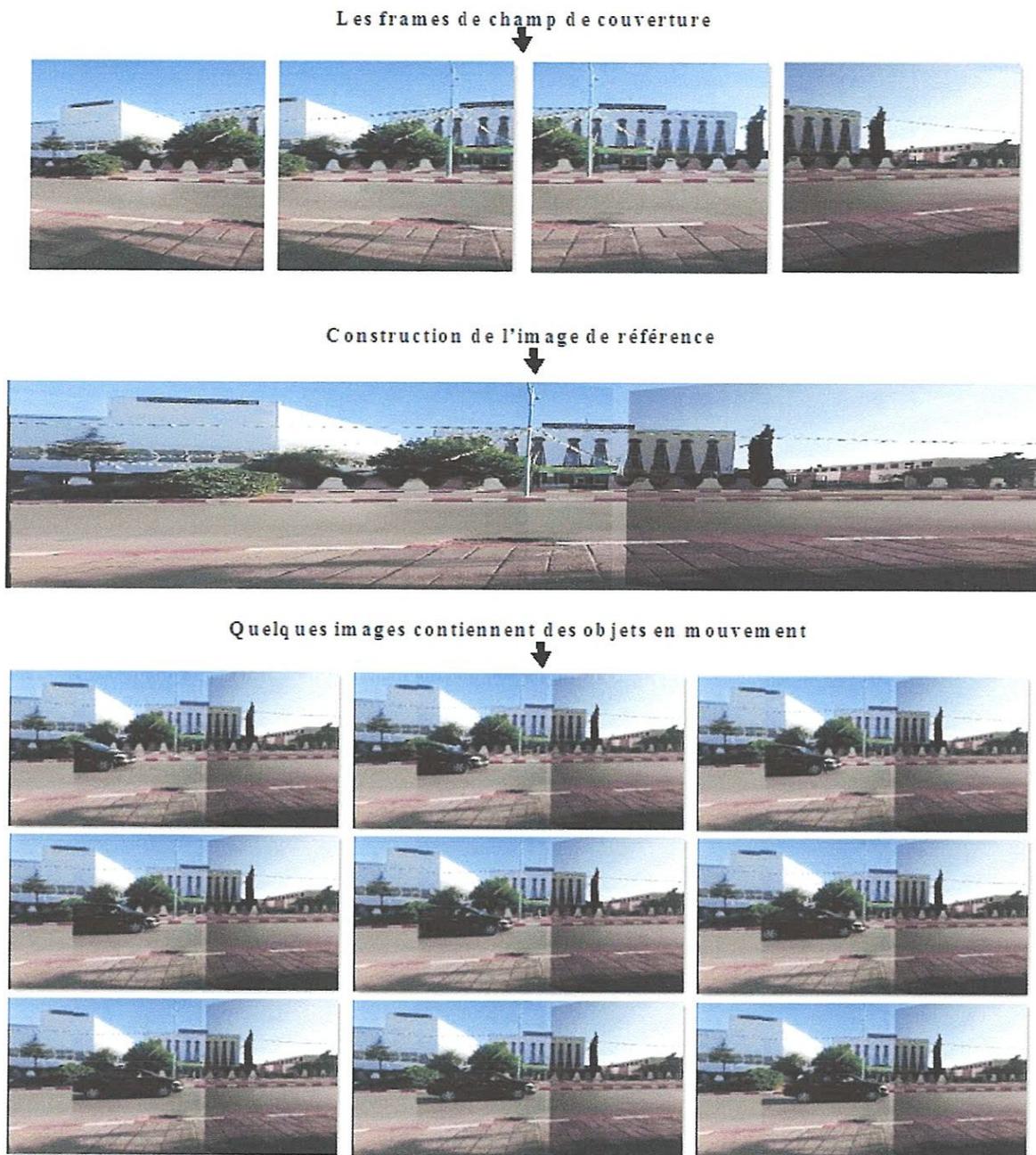


FIGURE 3.6 – Résultat du matching sur une autre vidéo .

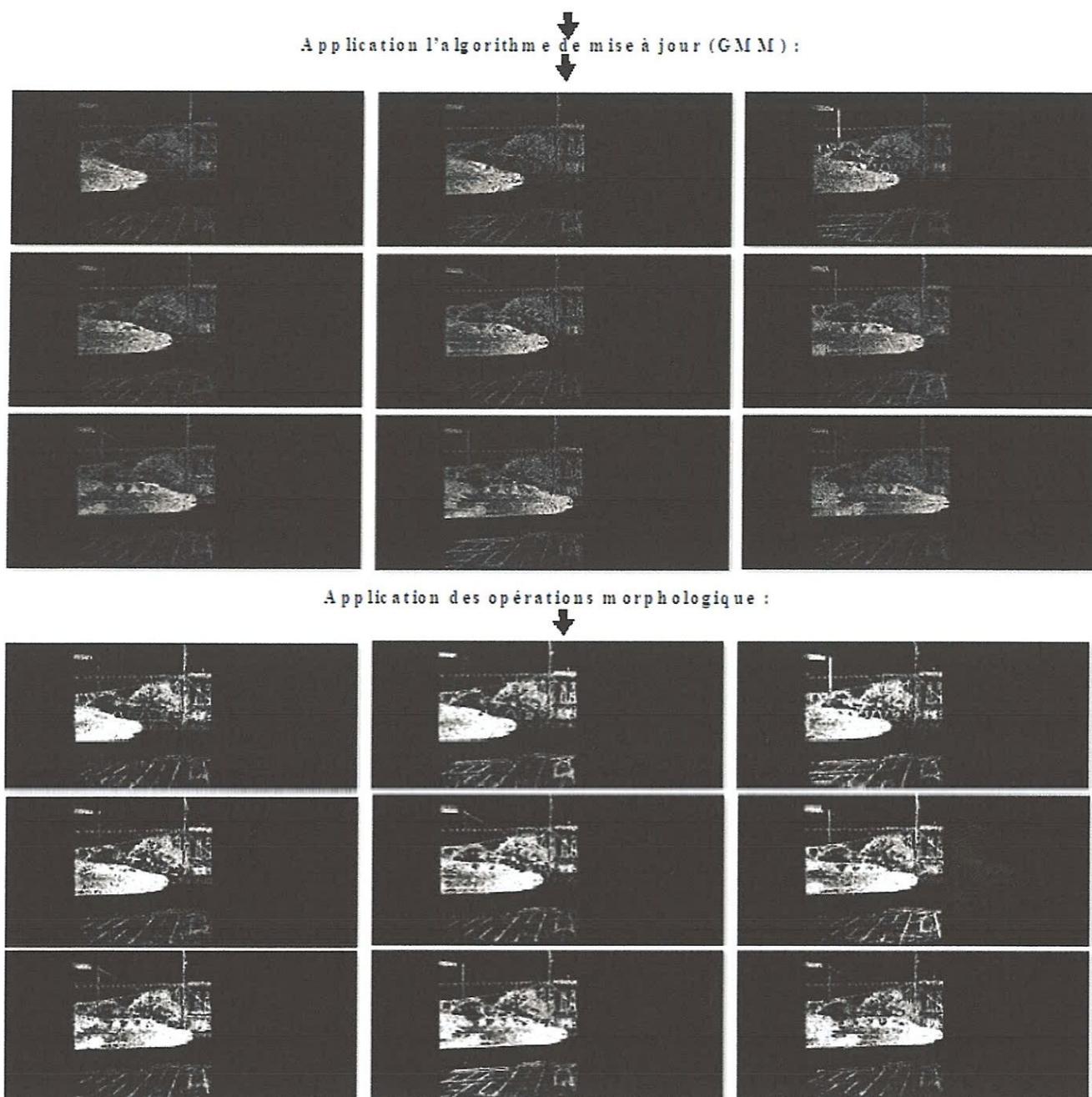


FIGURE 3.7 – Résultat de la détection sur une autre vidéo .

3.5 Conclusion

Dans ce chapitre nous avons présenté les différents résultats que nous avons obtenus en appliquant les méthodes de modélisation de l'arrière-plan dans le contexte d'une caméra mobile sur un axe rotatif. Nous avons eu beaucoup de problèmes pour tester notre système. Ces problèmes se situent dans la non-présence de base de tests publique. En effet, malgré la création de notre propre base, cette dernière n'est pas tout à fait adéquate à cause des problèmes relatifs au calibrage de la caméra, la nature de la caméra, le support de rotation de la caméra qui n'est rien d'autre qu'un petit bricolage fait maison. Malgré cela, le système développé a montré des signes d'efficacité encourageante dans une problématique traitée de manière partielle dans la littérature.

Conclusion Générale

Les idées soutenues dans ce mémoire ont une finalité pratique et un objectif précis, qui sont la conception et l'implémentation d'un système de modélisation et l'extraction de l'arrière-plan dans une séquence vidéo filmée par une caméra mobile (PTZ).

Pour atteindre cet objectif, nous avons effectué une extension des méthodes d'extraction de l'arrière-plan d'une vidéo filmée par une caméra fixe et adapté ces méthodes sur les caméras Pan-tilt-zoom (PTZ).

Les résultats obtenus sont largement acceptables, et encourageants, mais nous suggérons de continuer la recherche dans ce domaine pour améliorer davantage les résultats.

Comme perspective, le système pourrait être amélioré en augmentant sa vitesse d'exécution, en améliorant son résultat par le biais de nouvelles techniques, ou encore en facilitant son utilisation.

Le suivi des objets en mouvement reste un sujet de recherche ouvert en raison du manque de système de suivi satisfaisant, prenant en compte les distorsions intrinsèques et extrinsèques, la variation d'éclairage, occlusions, les bruis, et erreurs dans la correspondance et la classification d'objets ne sont qu'une partie des problèmes actuellement rencontrés dans la détection et le suivi des objets en mouvements .

Bibliographie

- [BH16] Sarra BENFRIHA and Asma HAMEL. Segmentation d image par coopération région-contours. 2016.
- [Com17] Europe / OSGi Community. Introduction à l'utilisation d'eclipse, 2017.
- [CSZS02] Shu-Ching Chen, Mei-Ling Shyu, Chengcui Zhang, and Jeff Strickrott. A multimedia data mining framework : Mining information from traffic video sequences. *Journal of Intelligent Information Systems*, 19(1) :61–77, 2002.
- [CYP⁺08] Chung-Hao Chen, Yi Yao, David Page, Bisma Abidi, Andreas Koschan, and Mongi Abidi. Heterogeneous fusion of omnidirectional and ptz cameras for multiple object tracking. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(8) :1052–1063, 2008.
- [FAR16] Brahim FAROU. Multimedia mining reconnaissance des formes dans une video, 2016.
- [FMS⁺05] Gian Luca Foresti, Christian Micheloni, Lauro Snidaro, Paolo Remagnino, and Tim Ellis. Active video-based surveillance system : the low-level image and video processing techniques needed for implementation. *IEEE Signal Processing Magazine*, 22(2) :25–37, 2005.
- [FR97] Nir Friedman and Stuart Russell. Image segmentation in video sequences : A probabilistic approach. In *Proceedings of the Thirteenth conference on Uncertainty in artificial intelligence*, pages 175–181. Morgan Kaufmann Publishers Inc., 1997.
- [HP06] Marko Heikkila and Matti Pietikainen. A texture-based method for modeling the background and detecting moving objects. *IEEE transactions on pattern analysis and machine intelligence*, 28(4) :657–662, 2006.

- [HZH10] M Hedayati, Wan Mimi Diyana Wan Zaki, and Aini Hussain. Real-time background subtraction for video surveillance : From research to reality. In *Signal Processing and Its Applications (CSPA), 2010 6th International Colloquium on*, pages 1–6. IEEE, 2010.
- [KPK⁺03] Sangkyu Kang, Joon-Ki Paik, Andreas Koschan, Besma R Abidi, and Mongi A Abidi. Real-time video tracking using ptz cameras. In *Sixth International Conference on Quality Control by Artificial Vision*, volume 5132, pages 103–112. International Society for Optics and Photonics, 2003.
- [oST12] National Institute of Standards and Technology. A java matrix package, 2012.
- [Rob09] Lionel Robinault. Mosaïque d images multi resolution et applications, 2009.
- [sdadAeP18] Ecole supérieure d art d Aix-en Provence. Atelier vidéo , entrelace ou progressif. quelle différence, 2018.
- [SG99] Chris Stauffer and W Eric L Grimson. Adaptive background mixture models for real-time tracking. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 2, pages 246–252. IEEE, 1999.
- [SPM04] Sudipta N Sinha, Marc Pollefeys, and Leonard McMillan. Camera network calibration from dynamic silhouettes. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 1, pages I–I. IEEE, 2004.
- [STT] STTRATEG. Sttrateg electricite generale ,camera video surveillance et telesurveillance.
- [tea18] OpenCV team. Open source computer vision library, 2018.
- [wik] wikipedia. /fr.wikipedia.org/wiki/vid