



Université ABBES LAGHROUR Khenchela
Faculté des Sciences et de la Technologie
Département de Génie Industriel
جامعة عباس لغرور خنشلة
كلية العلوم والتكنولوجيا
قسم الهندسة الصناعية



N° Série :

Mémoire de fin d'étude

Présenté pour l'obtention du diplôme de Master

Filière : Automatique

Spécialité : Automatique et Informatique Industriel

THEME

**Détection et poursuite d'un objet
dans une séquence de vidéo**

*Réalisées par : - Merah Salsabil
- Chabbi Bilal*

Devant Le Jury :

<i>Président:</i>	<i>Dr. Mokhtari Khalil</i>
<i>Examinatrice:</i>	<i>Dr. Aouragh Salima</i>
<i>Rapporteur:</i>	<i>Dr. Friha Souad</i>

Promotion 2021/2022

Remerciements

Tout d'abord nous remercions Allah de tous nos cœurs de nous avoir donné le courage et la patience qui nous ont permis d'accomplir ce modeste travail.

Nous remercions tout d'abord la respectueuse M^{me} Friha Souad qui a voulu bien nous encadrer, de nous avoir orientés et guidés et d'avoir éclairé nos chemins et surtout d'avoir été là pour nous soutenir.

Nous remercions aussi monsieur Boumahrez pour ses conseils.

Notre remerciements vont également aux membres de jury d'avoir accepté d'examiner et d'évaluer ce travail :

Dr. Khalil Mokhtari: Président

Mme Aouragh Salima: Examinatrice

Nous remercions tous nos amis et collègues pour leur soutien moral et tous les moments sympathiques qui vont rester gravés dans nos mémoires.

Enfin, à nos très chers parents et frères, vos encouragements, vos prières et recommandations nous ont accompagnés tout au long de nos études.

Ce modeste travail est le témoignage de notre reconnaissance et de notre profonde affection. Nous vous aimons énormément

Dédicace

Louange à Allah, maître de l'univers.

Paix et salut sur notre Prophète Mohammed

Je dédie ce modeste travail :

En signe de respect et de reconnaissance à mes très chers parents pour les valeurs qu'ils m'ont

Transmises

Mes petites nièces, le bonheur de ma vie

Mes frères et sœurs et leurs marguerites qu'Allah les bénisse

Le Chrysanthème roi de l'automne

Mes violettes et leurs pétales pour l'espoir qu'elles m'ont donné

Aux Bangtan sonyeondan

A tous les miens.....

Résumé :

Au cours de ces dernières décennies, la détection et le suivi d'objets ont attiré beaucoup d'intérêt en raison de leurs diverses applications dans la vie humaine, en particulier la vidéosurveillance et la robotique. Le suivi d'objet est l'estimation de la localisation d'un objet et la détermination de sa trajectoire au cours du temps dans une séquence vidéo. Bien que de nombreux algorithmes de suivi aient été développés ces dernières années, il demeure un problème non résolu à cause du nombre élevé de facteurs environnementaux.

Le but de ce travail est de détecter et de suivre les objets mobiles dans des séquences vidéo. L'outil de base que nous avons utilisé pour la détection des objets mobiles est la technique de soustraction de l'arrière-plan, c'est une technique simple qui permet d'extraire l'avant-plan (Foreground). Le filtre de Kalman est utilisé pour le suivi car il peut être appliqué dans le cas des scènes complexes ou simples à la fois.

Les tests effectués sur plusieurs vidéos ont donnés de bons résultats en général qui peuvent être améliorés dans la perspective de la continuation du présent travail.

Mot clé : Détection, suivi, filtre de Kalman, vidéosurveillance.

Abstract :

In recent decades, object detection and tracking has attracted a lot of interest because of its diverse applications in human life, especially video surveillance and robotics. Object tracking is the estimation of the location of an object and the determination of its trajectory over time in a video sequence. Although, many object tracking algorithms have been developed in recent years, it remains an unresolved problem because of the high number of environmental factors.

The goal of this work is to detect and track moving object in video sequences .The basic tool that we used for the detecting of moving object is the technique of subtraction of the background , it is a simple way to allow extracting the foreground. We used Kalman filter for tracking moving objects because it can be applicable to both simple and complex scenes

Tests carried out on several videos have given good results in general. They could be improved in the perspective of the continuation of the present work.

Key words : Detection, tracking, Kalman filter, video surveillance.

Sommaire

Liste des figures

Introduction général

Chapitre I : Notion de base sur les images et les vidéos

1. Introduction.....	01
2. L'image et sa représentation.....	03
2.1 Types d'images.....	03
2.2 Caractéristiques de l'image	04
2.2.1 Pixel.....	04
2.2.2 Poids de l'image.....	04
2.2.3 Bruit	05
2.2.4 Transparence	05
2.2.5 Luminance	05
2.2.6 Contraste.....	05
2.2.7 Histogramme	05
2.3 Les différents formats d'images	06
2.3.1 Format vectorielle	06
2.3.2 Format matricielle	07
3. Frame	07
4. Objet	08
5. La détection de l'objet	08
6. La vidéo	08
6.1 Définition	08
6.2 Les fondamentaux physiques et techniques	08
6.2.1 La vision	08
6.2.2 La couleur	08
6.2.3 Le signal	09
6.2.4 Affichage vidéo	10
6.2.5 Le son.....	10
7. Types de vidéo	10
7.1 La vidéo analogique	10

7.2 La vidéo numérique	11
7.2.1 La fréquence d'échantillonnage	11
7.2.2 La quantification de l'échantillonnage	11
8. Composition des fichiers vidéo	11
8.1 Le conteneur	11
8.2 Le(s) conteneur(s)	11
9. Les formats de fichiers vidéo	12
9.1 Les codecs vidéo	12
9.2 Les Conteneurs vidéo	12
10. Hiérarchie des données vidéo	13
10.1 Séquence vidéo (Vidéo Séquence)	13
10.2 Groupe d'images (Group of Picture)	13
10.3 Image (Picture)	13
10.4 Tranche (Slice)	13
10.5 Macroblocs (Macroblocks)	13
10.6 Blocs (Block)	14
11. Représentation d'une séquence vidéo	14
11.1 La résolution en luminance	14
11.2 La résolution spatiale.....	14
11.3 La résolution temporelle.....	14
12. Conclusion.....	14
Chapitre II : détection et suivi d'objet	
1. Introduction	16
2. Définition d'un objet vidéo	00
3. Détection d'objets	16
4. les méthodes de détection	17
4.1 La différence temporelle	17
4.2 Flux optique	18
4.3 Soustraction de l'arrière-plan	18
4.4 Elimination des ombres	20
5. Le suivi des objets	20
5.1 Domaines d'application	21

5.2 Les défis du suivi d'objet	22
6. les méthodes de suivi.....	23
6.1 Suivi de silhouette.....	23
6.2 Suivi de noyaux.....	23
6.3 Suivi de points.....	24
7. Le filtre de Kalman[.....	24
7.1 Principe du filtre de Kalman.....	24
7.2 Suivi de mouvement par filtre de Kalman.....	25
8. Conclusion.....	26
Chapitre III : Résultats de détection d'un objet en mouvement	
1.Introduction	28
2. Environnement du travail	28
2.1 Environnement matériel ou Hardware:	28
2.2 Environnement logiciel.....	28
3. Architecture du système.....	29
3.1 Acquisition des vidéos.....	29
3.2 Prétraitement	29
3.3 Détection des objets en mouvement	30
3.4 Suivi avec filtre kalman	30
4.Tests et Résultats.....	32
4.1 A partir de la webcam.....	32
4.2 Vidéo téléchargée (You tube)	34
4.3 Vidéo capturée sur téléphone.....	36
5. Discussion des résultats.....	38
6. Conclusion.....	38
CONCLUSION GÉNÉRALE.....	39
Bibliographie.....	40

Liste de figures :

Chapitre I :

Figure I.1 : Image couleur, son repère et un extrait de pixels.	03
Figure I.2 : Exemple d'image en niveaux de gris.....	06
Figure I.3 : codage de couleur en RVB.....	08
Figure I.4 : Schéma de principe d'un Capture Tri-CCD (charged coupled device).....	09
Figure I.5 : Hiérarchie des données dans le flux vidéo.	14

Chapitre II :

Figure II.1 Exemple de détection du mouvement avec la méthode de la soustraction d'images consécutives.....	17
Figure II.2 Le flux optique, (a et b) Région de la bouche sur deux images consécutives, (c) Champ de vecteur de vitesses.	18
Figure II.3 présentation de La méthode de soustraction de fond.	19
Figure II.4 – Illustration du suivi par la détection basée sur la classification des SVM . (a) montre la carte de score de visage / non-visage classification; (b) affiche la zone de recherche et de la région de contexte pour; (c) trace l'hyperplan de classification.	21
Figure II.5– Quelques applications de suivi d'objets.....	22
Figure II.6 – Quelques difficultés de suivi d'objet.....	23
Figure II.7 le cycle du filtre de Kalman.....	25

Chapitre III :

Figure III.1 Architecture général du système.....	29
Figure III.2 Objet mobile détecté	30
Figure III.3 Architecture du filtre de Kalman.....	31
Figure III.4 Résultat 1	32
Figure III.5 Résultat 2.....	32
Figure III.6 Résultat 3.....	33
Figure III.7 Résultat 4.....	33
Figure III.8 Résultat 1.....	34
Figure III.9 Résultat 2.....	34
Figure III.10 Résultat 3.....	35
Figure III.11 Résultat 4.....	35
Figure III.12 Résultat 1.....	36
Figure III.13 Résultat 2.....	36
Figure III.14 Résultat 3.....	37
Figure III.15 Résultat 4.....	37

Introduction général :

La vision par ordinateur est à la base de tout système de vision artificielle. C'est une branche de l'intelligence artificielle dont l'objectif est de permettre à une machine d'analyser, traiter et comprendre une ou plusieurs images prises par un système d'acquisition (caméras, etc.).

Le suivi d'objets est l'un des sujets de recherche les plus importants et les plus difficiles en vision par ordinateur et a été appliqué avec succès dans un large éventail d'applications réelles, comme la vidéosurveillance et la robotique.

Le suivi d'objet est un domaine de recherche dont les premiers travaux remontent à la fin des années 80 et dont les grands progrès ont été réalisés ces dernières années. De nos jours, le suivi d'objets dans une séquence vidéo est classé parmi les sujets de recherche les plus actifs. Le but central du suivi est d'estimer au fil du temps la localisation de l'objet cible dans chacune des images d'une séquence vidéo.

Le suivi d'objets reste une tâche complexe et difficile en raison de nombreux défis liés aux limitations des capteurs de vision (faible cadence d'image, basse résolution, faible dynamique par pixel, distorsions des couleurs, bruit, etc.), aux objets (objets non rigides, nombre d'objets variant au fil de temps, occultations entre objets, petites tailles d'objets, etc.), aux exigences des scénarios applicatifs (fonctionnement en temps réel, haute fiabilité du système, etc.) et à l'environnement (variation d'éclairage, occultations causées par l'environnement, etc.).

De plus, la prolifération de données vidéo et de nouveaux dispositifs d'acquisition de données ont suscité un grand intérêt pour la construction d'algorithmes de suivi plus intelligents.

De nombreux algorithmes de suivi d'objet ont été proposés dans la littérature pour faire face à ces défis et pour assurer une bonne qualité de suivi, certains utilisent des modèles générateurs tandis que d'autres utilisent des modèles discriminants

Cependant, Il n'existe aucune méthode de suivi unique qui peut être appliquée avec succès à tous les scénarios. Le processus du suivi d'objets peut s'exprimer en fonction de la détection des objets. [13]

Nous avons étudiés quatre méthodes de détection des objets en mouvement dans une scène vidéo et trois méthodes de suivi .

Notre mémoire est décomposé en trois chapitres :

Chapitre I : Nous parlons sur des notions sur traitement d'image

Chapitre II : Nous présentons les méthodes de détection et suivi

Chapitre III : conclut ce mémoire en présentant l'ensemble des résultats expérimentaux obtenus pour la détection et le suivi d'objets en mouvement.

Ce travail est terminé par une conclusion générale et une bibliographe.

Chapitre I

Notion de base sur les images et les vidéos

1. Introduction

Le traitement des vidéo est une science récente qui a pour but d'offrir aux spécialistes de différents domaines, comme au grand public, des outils de manipulation de ces données digitales issues du monde réel. Dans ce chapitre, nous allons présenter les différentes phases de la formation de vidéo ainsi que sa définition, de ces différents types, de ces caractéristiques et de ces formats.

2. L'image et sa représentation

L'image est définie comme étant une fonction $f(x, y)$ à deux dimensions, où x et y sont les coordonnées spatiales, et l'amplitude à tous points $f(x, y)$ correspondant à l'intensité ou au niveau de gris. Lorsque les points (x, y) et l'amplitude sont discrétisés, on parle d'image numérique ou digitale. Dans ce dernier cas la fonction f est remplacée par la lettre I et le couple (x, y) par le couple (i, j) [1]

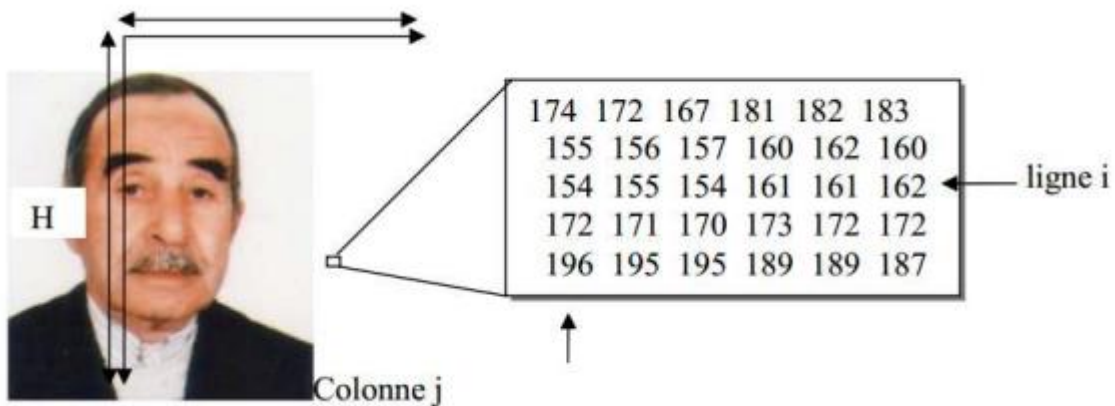


Figure I.1 : Image son repère et un extrait de pixels. [2]

L'image numérique, désignée aussi par le terme scène, possède un repère comme indiqué Figure I.1, il est différent de celui d'une fonction mathématique. Elle a une hauteur (H) et une largeur (W). [2]

2.1 Types d'images

Les images peuvent être en plusieurs types :

➤ Image binaire

Une image binaire est une matrice rectangulaire dont les éléments valent 0 ou 1. Lorsque l'on visualise une telle image, les zéros sont représentés par des noirs et les uns par des blancs. [3]

➤ Image en niveau de gris

Chaque pixel est un niveau de gris, allant de 0 (noir) à 255 (blanc). Cet intervalle de valeur signifie que chaque pixel est codé sur huit bits (un octet). 256 niveaux de gris suffisent pour la reconnaissance de la plus part des objets d'une scène. [15]

➤ Codage d'une image en niveaux de gris

Si on code chaque pixel sur 2 bits on aura 4 possibilités (noir, gris foncé, gris clair, blanc). L'image codée sera très peu nuancée. En général on code chaque pixel sur 8 bits = 1 octet. On a alors 256 possibilités (on dit 256 niveaux de gris).

L'image de 10 000 pixels codée occupe alors 10 000 octets en mémoire. Exemple d'image en 72 pixels par pouce (environ 30 pixels par cm), codée en 256 niveaux de gris.

Cette image de 303 x 303 pixels occupe 303 x 303 = 91809 octets puisque chaque pixel occupe 1 octet en mémoire. [15]

➤ Image couleur (RGB)

Chaque pixel possède une couleur décrite par la quantité de rouge (R), vert (G) et bleu (B). Chacune de ces trois composantes est codée sur l'intervalle [0, 255], ce qui donne $256^3 = 16\,777\,216$ couleurs possibles. Il faut 24 bits pour coder un pixel. [15]

➤ Codage d'une image en couleurs 8 bits

Dans ce cas on attache une palette de 256 couleurs à l'image. Ces 256 couleurs sont choisies parmi les 16 millions de couleurs de la palette RVB. Pour chaque image le programme recherche les 256 couleurs les plus pertinentes.

Chaque code (de 0 à 255) désigne une couleur. L'image occupe 3 fois moins de place en mémoire qu'avec un codage 24 bits. L'image est moins nuancée : sa qualité est bonne mais moindre. [15]

➤ Codage d'une image en couleurs 24 bits

Il existe plusieurs modes de codage de la couleur. Le plus utilisé est le codage Rouge, Vert, Bleu (RVB). Chaque couleur est codée sur 1 octet = 8 bits. Chaque pixel sur 3 octets c'est à dire 24 bits : le rouge de 0 à 255, le vert de 0 à 255, le Bleu de 0 à 255.

Le principe repose sur la synthèse additive des couleurs : on peut obtenir une couleur quelconque par addition de ces 3 couleurs primaires en proportions convenables.

On obtient ainsi $256 \times 256 \times 256 = 16\,777\,216$ (plus de 16 millions de couleurs différentes). [15]

2.2 Caractéristiques de l'image

2.2.1 Pixel

Une image est constituée d'un ensemble de points appelés pixels (voir l'extrait de la figure I.1). Le pixel (Picture élément) représente ainsi le plus petit élément constitutif d'une image numérique. La quantité d'information que véhicule chaque pixel donne des nuances entre images monochromes et images couleurs. Pour les images 3D le «pixel» est alors appelé un voxel, et représente un volume élémentaire. Des exemples d'images de ce type se rencontrent dans les images médicales. Les images tomographiques axiales sont ainsi des images construites à partir de plusieurs radiographies faites sous des angle de vue différents. [15]

2.2.2 Poids de l'image

C'est la taille de l'image. Etant donné que cette dernière est représentée sous forme d'une matrice dont les valeurs représentent l'intensité (pixels), le nombre de colonne (W) multiplié par le nombre de ligne (H) donne le nombre total de pixels dans l'image. Pour une image de 640x480 en couleur:

- Nombre de pixel = $640 \times 480 = 307200$
- Poids de chaque pixel = 3 octets

➤ Le poids de l'image = $307200 \times 3 = 921600$ octets = 900 Ko. [15]

2.2.3 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. [4]

2.2.4 Transparence

La transparence est une caractéristique définissant le niveau d'opacité des éléments de L'image, c'est la possibilité de voir à travers l'image des éléments graphiques situés derrière celle-ci. [15]

2.2.5 Luminance

C'est le degré de luminosité des points de l'image. Elle est définie ainsi comme étant le quotient de l'intensité lumineuse d'une surface par l'aire apparente de cette surface. [5]

2.2.6 Contraste

Est une propriété intrinsèque d'une image qui désigne et quantifie la différence entre les parties claires et foncées d'une image (elle différencie les couleurs claires des couleurs foncées). En photographie on le définit le contraste comme la différence entre la densité la plus forte et la plus faible d'une image. Le contrôle du contraste est un élément important de la pratique photographique. Le contraste final de l'image dépend à la fois du sujet, de la nature et du traitement du négatif et du positif. [15]

2.2.7 Histogramme

Un histogramme est un graphique statistique permettant de représenter la distribution des intensités des pixels d'une image. Il fournit diverses informations comme les statistiques d'ordre (moyenne, variance,...), l'entropie, et peut permettre d'isoler des objets. [15]

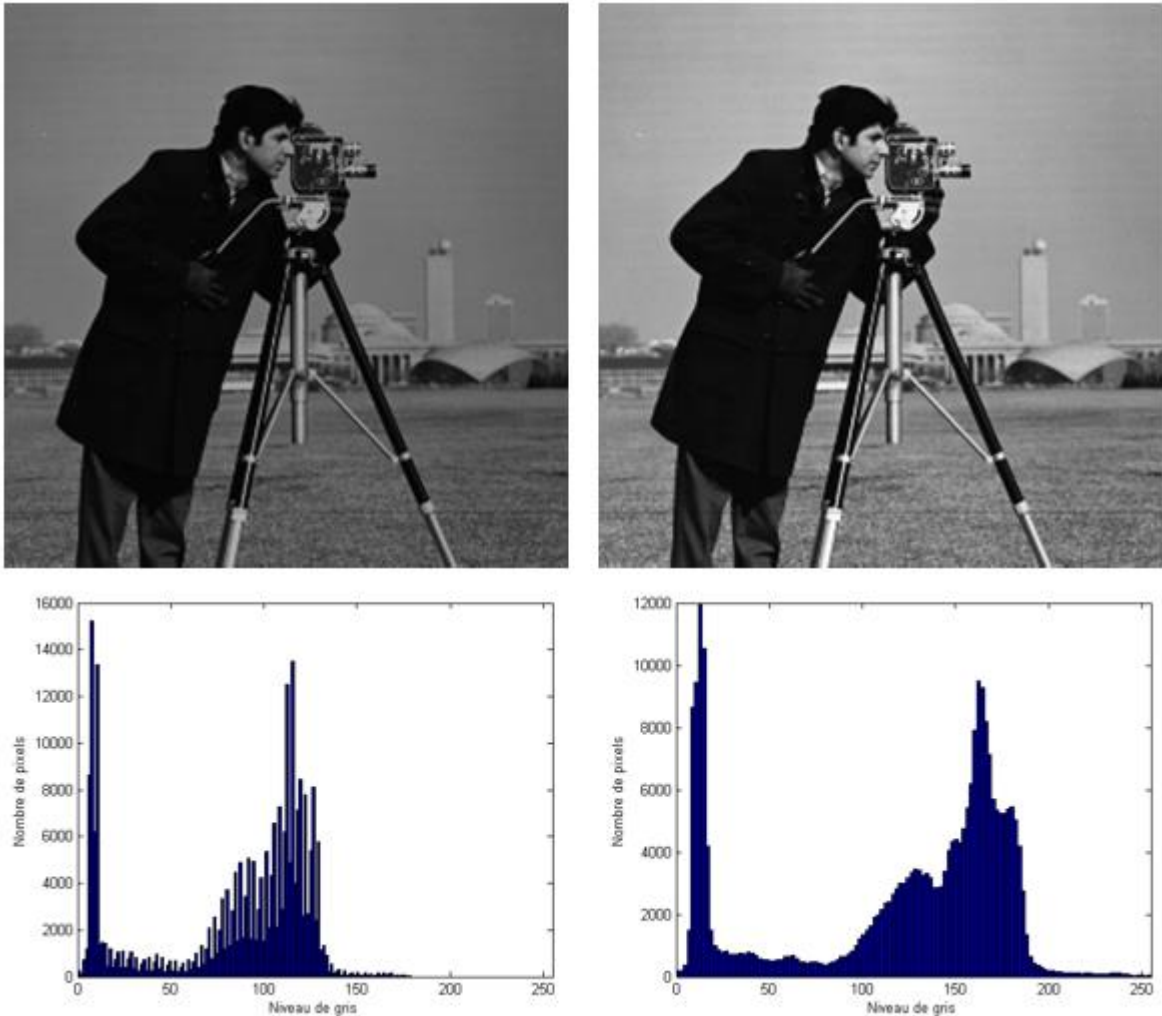


Figure I.2 : Exemple d'image en niveaux de gris. [15]

Pour les images en couleur plusieurs histogrammes sont nécessaires.

Par exemple pour une image codée en RGB:

- un histogramme représentant la distribution de la luminance
- trois histogrammes représentant respectivement la distribution des valeurs respectives des composantes rouges, bleues et vertes. [15]

2.3 Les différents formats d'images

On peut classer les images en deux formats :

2.3.1 Format vectorielle

Dans une image vectorielle les données sont représentées par des formes géométriques simples qui sont décrites d'un point de vue mathématique.

Par exemple, un cercle est décrit par une information du type (cercle, position du centre, rayon). Ces images sont essentiellement utilisées pour réaliser des schémas ou des plans.[15]

2.3.2 Format matricielle

Une image matricielle est formée d'un tableau de points ou pixels. Plus la densité des points sont élevée, plus le nombre d'informations est grand et plus la résolution de l'image est élevée. Corrélativement la place occupée en mémoire et la durée de traitement seront d'autant plus grandes

Les images vues sur un écran de télévision ou une photographie sont des images matricielles. [15] On obtient également des images matricielles à l'aide d'un appareil photo numérique, d'une caméra vidéo numérique ou d'un scanner. Parmi ces formats on peut citer :

- **BMP (Bitmap) :**
Le format BMP est le format par défaut du logiciel Windows. C'est un format matriciel. Les images ne sont pas compressées. Son logiciel d'origine.
- **Le format EPS :**
Matriciel n'est pas très différent du EPS vectoriel. En fait seules les données contenues dans le fichier sont différentes. Ainsi un logiciel de retouche de photos tel que Photoshop permet l'importation, la modification et l'exportation de fichiers en format EPS.
- **GIF (GraphicalInterchange Format) :**
Le format GIF est un format qui a ouvert la voie à l'image sur le World Wide Web. C'est un format de compression qui n'accepte que les images en couleurs indexés codé sur 8 bits, C'est un format qui perd beaucoup de son marché suite à une bataille juridique concernant les droits d'utilisation sur Internet.
- **JPEG (Joint Photographique Experts Group) :**
Les images JPEG sont des images de 24 bits. C'est-à dire qu'elles peuvent afficher un spectre de 16 millions de couleurs. C'est la meilleure qualité d'images disponible. [15]

3. Frame

Image composants la vidéo, les photogrammes [15]

4. Objet

Les objets physiques sont les objets du monde réel qui apparaissent dans les scènes observées par les caméras. [6] Les objets physiques sont divisés en deux types : les objets de contexte et les objets mobiles.

- Les objets de contexte Sont des objets physiques qui sont habituellement statiques (p. ex. les murs). Dans le cas où ils ne sont pas statiques, leurs mouvements peuvent être prédits par les informations contextuelles p.ex. Les chaises, les portes sont des objets de contexte. [6]
- Les objets mobiles Sont des objets physiques qui peuvent être perçus dans les scènes par leurs mouvements[6]. Il est cependant difficile de prédire leurs mouvements P. ex. les personnes, les véhicules.
- Les objets dans les vidéos sont des objets mobiles. Ils sont habituellement détectés et suivis dans plusieurs frames.

5. La détection de l'objet

La détection de l'objet comprend la détection d'objets et la reconnaissance de formes dans le cadre d'une séquence vidéo. Un mécanisme de détection d'objet est nécessaire dans ne importe quel procédé de suivi soit dans chaque trame ou lorsque l'objet apparaît en premier dans la vidéo. [7]

6. La vidéo

6.1 Définition

La vidéo est une succession d'images animées défilant à une certaine cadence afin de créer une illusion de mouvement pour l'œil humain. Elle peut être analogique (signal continu d'intensité de luminance) ou numérique (suite de trames ou images). [8]

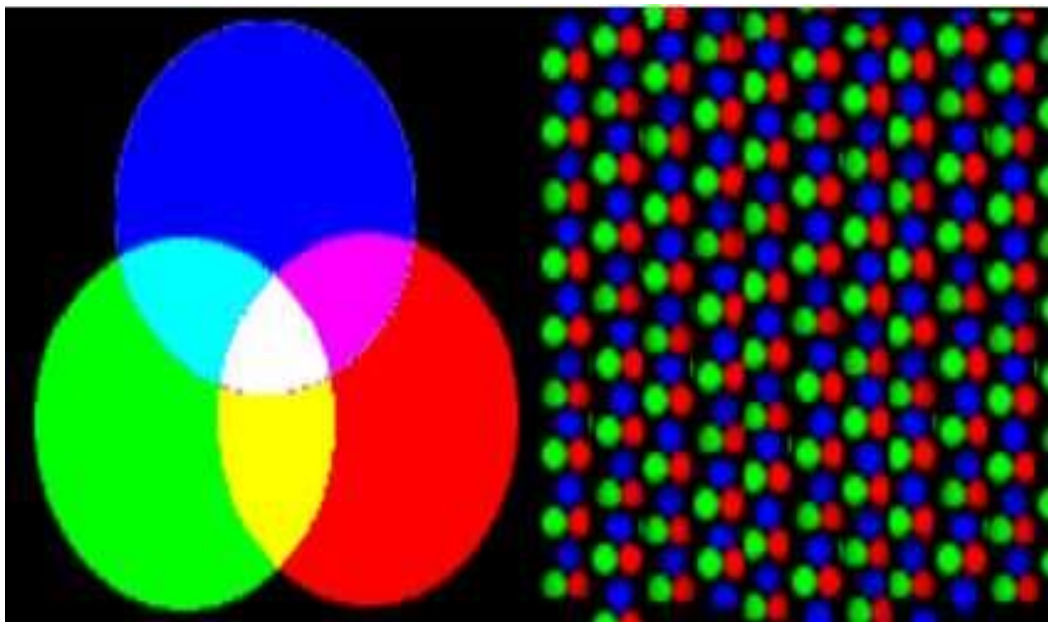
6.2 Les fondamentaux physiques et techniques

6.2.1 La vision

Dans le phénomène de la vision humaine, un flux lumineux composé de photons frappe l'œil. Dans l'œil, les bâtonnets vont réagir à l'intensité de la lumière (la luminance ou Y) et les cônes vont réagir à la couleur (la chrominance ou C). Le cerveau effectue la synthèse de ces informations pour composer une image. L'œil humain est davantage sensible à l'intensité lumineuse (Y) qu'à la couleur (C). [S12]

6.2.2 La couleur

La vidéo utilise ce qui est identifié comme la synthèse additive de la couleur. Dans le système colorimétrique additif, les trois couleurs primaires sont le rouge, le vert et le bleu (RVB ou RGB dans la terminologie anglophone). En combinant ces trois couleurs primaires, il est possible de reproduire tout le spectre visible par l'œil humain. L'écran de visualisation vidéo sera donc composé d'une série



de triades rouge-vert-bleu. L'activation de l'ensemble de ces triades formera l'image [S12]

Figure I.3 : codage de couleur en RVB [15]

6.2.3 Le signal

En vidéo, c'est la caméra qui transforme l'information lumineuse (photons) en signal électrique (électrons). En vidéo analogique, l'intensité de ce signal électrique varie de façon continue. Le processus de transformation de l'information lumineuse en signal électrique est le suivant :

- L'élément optique de la caméra, l'objectif, sépare la lumière en trois composants le rouge, le vert, le bleu. Cette opération est réalisée en faisant passer le flux lumineux par une succession de filtres dichroïques réfléchissant certaines couleurs et en laissant passer d'autres.
- En vidéo professionnelle, les trois images sont projetées sur trois capteurs photo sensibles distincts formés chacun de centaines de milliers de point
Généralement entre 400 000 et 700 000. Ces capteurs sont nommés CCD (charged coupled device) ou dispositifs à transfert de charge. Les caméras domestiques ne sont généralement équipées que d'un seul capteur CCD [12]

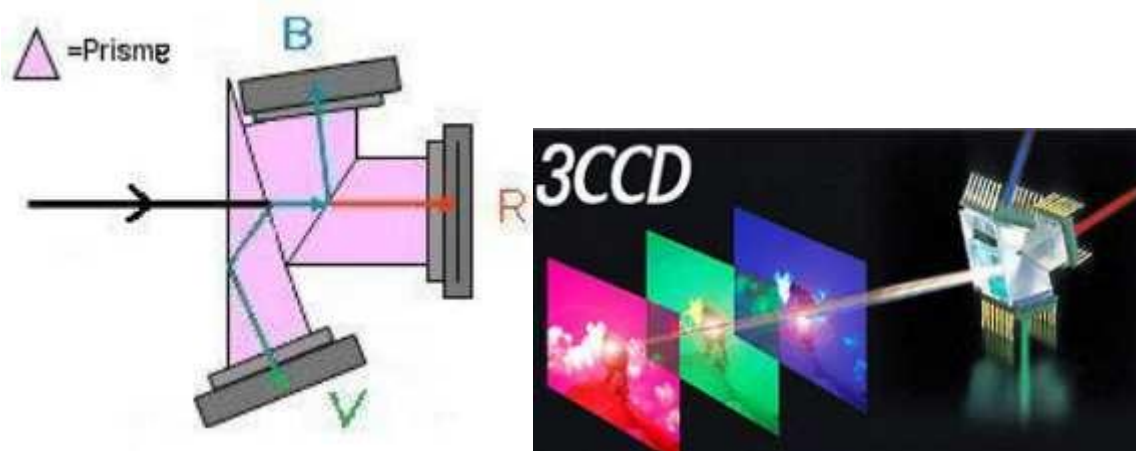


Figure II.4 : Schéma de principe d'un Capture Tri-CCD (charged coupled device) [S13]

- Pour chacun des points de chacun des capteurs, l'énergie lumineuse sera transformée en énergie électrique. Ainsi, à la sortie des capteurs, trois signaux électriques d'intensité variable, un signal pour chacune des trois composantes. La lumière blanche est formée par la somme des trois composantes RVB. Toutefois, elle n'est pas composée des trois couleurs primaires en quantité égale. Les propositions sont les suivantes :

$$100\% Y = 29,9\%(R) + 58,7\%(V) + 11,4\%(B)$$

- Une information supplémentaire devra être ajoutée à ces signaux. Comme nous l'avons dit, une valeur doit être établie pour chacun des points des capteurs. C'est cette valeur pour chacun des points de chacun des pixels qui permettra de recomposer l'image sur un écran de visualisation, chacun des points de la surface d'affichage pouvant alors prendre la même valeur qu'au moment de la capture par la caméra. La restitution de l'information pour chacun des pixels se fait par balayage ligne par ligne de la surface de l'écran de reproduction, les valeurs enregistrées pour chacun des pixels au moment de la capture étant transférées aux pixels de la surface de reproduction. Pour synchroniser ces balayages, un signal de fin de ligne est ajouté pour chacune des lignes. Un signal de fin d'image est également inséré. [15]
- Les caractéristiques d'un signal vidéo
Chaque fichier vidéo a des attributs qui décrivent ce qui constitue le signal vidéo. Ces caractéristiques comprennent:

- Châssis de taille: Il s'agit de la dimension de pixel de l'image
- Le ratio d'aspect: C'est le rapport entre la largeur et la hauteur
- Vitesse de défilement: C'est la vitesse à laquelle les images sont capturées et destinées à la lecture.
- Débit: Le taux de débit ou de données est la quantité de données utilisées pour décrire la partie audio ou vidéo du fichier. Il est généralement mesuré en unités par seconde et peut être en kilo-octets, méga-octets ou giga-octets par seconde. En général, plus la vitesse de transmission, meilleure est la qualité.
- Le taux d'échantillonnage audio: C'est à quelle fréquence le signal audio est échantillonné lors de la conversion d'une source analogique à un fichier numérique. [9]

6.2.4 Affichage vidéo

La technique de la composition d'image par balayage a une conséquence : contrairement à l'image cinéma, l'image vidéo ne s'affiche pas d'un seul coup, elle s'affiche ligne par ligne, en commençant par le haut à gauche pour finir en bas à droite. Afin d'assurer une uniformité dans la luminosité de cet affichage, en d'autres termes afin que l'intensité lumineuse des pixels du haut de l'écran ne commence pas à diminuer alors que ceux du bas viennent d'être activés, une technique spécifique a été développée. Il s'agit de la technique d'entrelacement. Un premier balayage est effectué sur les lignes impaires, puis un deuxième sur les lignes paires. Il faut donc deux passages pour recomposer une image complète. La fréquence d'entrelacement s'appuie sur la fréquence du réseau électrique. Comme en Amérique du Nord la fréquence des réseaux est de 60Hz, le système de balayage est donc de 60 demi-images, soit 30 images par seconde. De façon précise, le taux de rafraîchissement est plutôt de 29,97 images par seconde. La norme Nord-Américaine en télévision est le NTSC (National Television Standard Committee). L'image NTSC est formée de 525 lignes. De ces lignes, seulement 486 sont

utilisées pour afficher l'image. Les deux autres principales normes sont européennes. Il s'agit des normes PAL (Phase Alternation by Line) et SECAM (Séquentiel Couleur à Mémoire). L'image PAL ou SECAM est formée de 625 lignes dont 576 sont utilisées pour l'image. Comme la fréquence des réseaux électriques européens est de 50 Hz, le système de balayage est de 50 demi-images, soit 25 images par seconde. [15]

6.2.5 Le son

Le son fera l'objet d'un traitement distinct. Le signal sonore sera distinct du signal vidéo. Il sera également enregistré sur des pistes distinctes sur le ruban vidéo. [15]

7. Types de vidéo

On distingue deux grandes familles de systèmes vidéo : les systèmes vidéo analogiques et les systèmes vidéo numériques.

7.1 La vidéo analogique

La caméra balaye l'image bidimensionnelle qu'elle a devant elle par un faisceau d'électrons qui se déplace très rapidement de gauche à droite et plus lentement de haut en bas et produit une tension en fonction du temps. Elle enregistre ainsi l'intensité lumineuse, et à la fin du balayage, on a alors une trame. Le faisceau revient à l'origine pour recommencer. Le récepteur va recevoir cette intensité en fonction du temps, et pour reconstruire l'image, va répéter le processus de balayage. [10]

7.2 La vidéo numérique

Le processus de captation de l'image vidéo en mode numérique est essentiellement le même que pour la vidéo analogique. Un système optique sépare la lumière en trois composants. À la sortie toutefois, plus d'enregistrement d'un signal électrique mais plutôt l'enregistrement d'une valeur numérique définie pour chacune des trois couleurs de chacun des pixels (RVB). Le processus de transformation d'un signal électrique et valeurs numériques s'appelle la numérisation ou l'échantillonnage. Deux paramètres servent à échantillonner un signal électrique : la fréquence d'échantillonnage et la quantification de l'échantillonnage ou le nombre de bits utilisés pour le codage. [15]

7.2.1 La fréquence d'échantillonnage

Il existe un théorème appelé le théorème de Shannon. Ce théorème stipule que «l'information véhiculée par un signal dont le spectre est à support borné n'est pas modifiée par l'opération d'échantillonnage à condition que la fréquence d'échantillonnage soit au moins deux fois plus grande que la plus grande fréquence contenue dans le signal. La reconstitution du signal original peut être effectuée par un filtre passe-bas idéal de fréquence de coupure égale à la moitié de la fréquence d'échantillonnage.»

La bande passante d'un signal vidéo peut atteindre 6Mhz. Pour numériser ce signal, une fréquence d'échantillonnage de 13,5Mhz sera donc utilisée. En d'autres termes, le signal vidéo sera mesuré 13 500 000 fois par seconde. Comme nous savons que l'œil est sensible à la lumière (luminance) qu'à la couleur (chrominance), souvent, la fréquence d'échantillonnage de la chrominance sera moins importante. Dans les faits, l'on utilise un échantillonnage à 13,5Mhz pour Y (luminance) et 6,75 pour C (U et V). Cet échantillonnage est donc un échantillonnage à 16 bits par pixel. [15]

7.2.2 La quantification de l'échantillonnage Un certain nombre de bits serviront à coder chacune des mesures. Nous savons que l'œil humain discerne un maximum de 256 niveaux différents de luminance. Ainsi, un codage à 8 bits est suffisant pour reproduire l'ensemble des variations perceptibles (8 = 256 combinaisons possibles) Comme chacune des composantes couleurs de l'image sera quantifiée sur 8 bits (8 bits Rouge, 8 bits Vert, 8 bits Bleu) (224= 16 777 216 combinaisons possibles) un total de 24 bits (3 octets) seront utilisés pour coder les informations relatives à un seul pixel d'une image. [S12] [11]

8. Composition des fichiers vidéos

Un fichier vidéo se compose généralement de 2 éléments :

8.1 Le conteneur

Il correspond généralement au format du fichier. Son rôle est de rassembler et d'organiser dans un fichier, différents types de données (flux audio, vidéos, sous-titres, meta-données).

Exemple de conteneurs:

- L'Audio Video Interleave (.AVI), Quicktime (.MOV), Real Media (.RM), MP4
- FLV (Flash Vidéo), MKV (Matroska), le WMV (Windows Media Video), Chaque conteneur possède ses spécificités en termes de nombre de pistes acceptées pour la vidéo et l'audio, de codecs reconnus. [S15]

8.2 Le(s) contenu(s)

Ils se composent essentiellement de flux audios et/ou vidéos. Ceux-ci sont généralement compressés à l'aide d'un codec (algorithme de compression/décompression) comme le Divx, le H264, le mp3.

Exemple: un flux vidéo au format Divx peut être « encapsulé » dans un conteneur AVI ou QuickTime. [15]

9. Les formats de fichiers vidéo

_Décrit l'ordre et la structure de ces images. Les données du flux vidéo, qui peuvent être accompagnées de sons sous la forme de flux audio, sont très volumineuses : elles doivent impérativement être compressées (codées) à l'aide d'un codec pour être stockées (sur disque dur ou sur les supports d'enregistrement : CD, DVD) ou/et transmises (et donc être adaptées au débit des réseaux).

Les flux vidéo (et le(s) flux audio éventuellement associé(s)), une fois encodés, sont généralement encapsulés dans des fichiers conteneurs : ces derniers permettent, notamment, leur lecture simultanée. [15]

9.1 Les codecs vidéo

Il existe de nombreux codecs permettant de compresser et de décompresser une vidéo dans son conteneur.

- **MJPEG (Motion JPEG)**

Le codec vidéo MJPEG compresse la vidéo image par image, en utilisant la technologie JPEG appliquée à l'image fixe, et réunit ces images en mouvement et le son dans un même format de fichier.

Le MJPEG est le codec le plus utilisé pour les captures vidéo des ensembles cartes d'acquisition et logiciels d'édition vidéo. La conservation d'une bonne qualité d'image produit toutefois de gros fichiers.

Le format MJPEG est un format non normalisé. Les solutions M-JPEG, au début des années 90, ont été développées sans concertations par des fabricants, conduisant à MPEG des solutions propriétaires et à des fichiers très souvent incompatibles entre eux.

- **MPEG (Moving Picture Experts Group)**

Les formats MPEG sont des formats de compression avec pertes pour les séquences vidéo. Le groupe MPEG (Moving Picture Experts Group), est un groupe d'experts créé en 1988 et chargé du développement de normes internationales pour la compression, la décompression, le traitement et le codage de la vidéo, de l'audio et de leur combinaison, de façon à satisfaire un large panel d'applications. Les formats produits par MPEG sont ouverts, mais non libres

- **DivX (Digital Video Express)**

- Extension : .avi

- Codec vidéo propriétaire et fermé proposé par DivX Inc., conçu à partir de MPEG-4 part 2, ce dernier ayant été modifié afin d'y ajouter la possibilité de compresser le son au format MP3. Cela permet ainsi d'obtenir des vidéos compressées très peu volumineuses avec une perte de qualité raisonnable. Ainsi le format DivX permet de stocker un film complet de plusieurs heures sur un CD-ROM de 650 ou 700 Mo. [15]

9.2 Les Conteneurs vidéo

Les conteneurs les plus courants sont les suivants :

- ❖ **AVI (Audio Video Interleave)**
 - Extension : .avi
 - Format propriétaire et ouvert
 - Le format AVI, développé par Microsoft, très répandu et lisible sur tous les lecteurs vidéo, c'est le format d'encapsulation le plus populaire. Dans un fichier AVI, chaque piste audio et/ou vidéo peut théoriquement être compressée par n'importe quel codec.
- ❖ **RealMedia**
 - Extension : .rm, .ram, .rpm
 - Format développé par Real Networks.
 - Format propriétaire et fermé.
 - Lisible uniquement avec Realplayer.
 - Conteneur de flux audio (notamment : RealAudio) et vidéo (notamment : RealVideo)
- ❖ **MP4**
 - Extension : .mp4, .mp4a, .mp4v, .m4P
 - Conteneur officiel pour la norme MPEG-4. [15]

10. Hiérarchie des données vidéo

La hiérarchie des données vidéo est présentée dans la Figure I.5

10.1 Séquence vidéo (Video Sequence)

Qui commence par un en-tête de séquence, contient un ou plusieurs groupes d'images et s'achève par un code de fin de séquence.

10.2 Groupe d'images (Group of Pictures)

Qui regroupe un en-tête et une série d'une ou plusieurs images facilitant l'accès direct

10.3 Image (Picture)

Qui est l'unité élémentaire pour le codage de la séquence vidéo. Une image est un groupe de trois matrices rectangulaires qui représentent la luminance (Y) et la chrominance (Cb et Cr). Cette représentation YCbCr est équivalente à celle RGB (Red, Green, Blue). Elle lui est préférable, car l'œil étant plus sensible à la luminosité qu'à la chrominance, il n'est pas nécessaire de stocker autant d'informations dans les matrices Cb et Cr que dans la matrice Y, alors qu'en RGB, dimension deux fois plus petite que la matrice Y.

10.4 Tranche (Slice)

Macroblobs ordonnés de gauche à droite, puis de haut en bas. En cas d'erreur dans la tranche, le décodeur passe à la suivante. Plus il y a de tranches, meilleur est le traitement des erreurs, mais fait perdre de la place.

10.5 Macroblobs (Macroblocks)

C'est une matrice rectangulaire de dimension 2 et constituée de blocs.

10.6 Blocs (Block)

C'est un ensemble des valeurs de luminance et chrominance de 8 lignes de 8 pixels. [15]

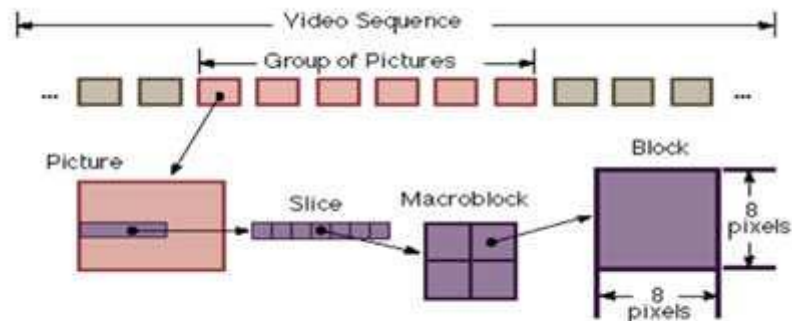


Figure I.5 : Hiérarchie des données dans le flux vidéo. [15]

11. Représentation d'une séquence vidéo

Une séquence vidéo brute est une suite d'images fixes, qui peut être caractérisée par trois principaux paramètres :

11.1 La résolution en luminance

Détermine le nombre de nuances ou de couleurs possibles pour un pixel. Celle-ci est généralement de 8 bits pour les niveaux de gris et de 24 bits pour les séquences en couleurs.

11.2 La résolution spatiale

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

11.3 La résolution temporelle

Est le nombre d'images par seconde. La valeur de ces trois paramètres détermine l'espace mémoire nécessaire pour stocker chaque image de la séquence. Cet espace mémoire est caractérisé par le débit, qui est le coût de stockage pour une seconde (capacité mémoire nécessaire pour stocker une seconde de vidéo). [12]

12. Coclusion

Dans ce chapitre, nous avons présentés une étude générale sur les caractéristiques fondamentales des vidéos dans le but de les améliorer ou d'extraire des informations pertinents.

Chapitre II

Détection et suivi d'objet

1. Introduction

La détection et le suivi d'objets sont parmi les problèmes les plus étudiés ces dernières années. Ils sont des tâches importantes et difficiles dans de nombreuses applications de vision par ordinateur telles que la robotique, la vidéosurveillance. La détection d'objet consiste à localiser l'objet dans chacune des trames d'une séquence vidéo. Le suivi d'objet est le processus de localisation spatiotemporelle d'un objet en mouvement au cours d'une séquence vidéo. Chaque méthode de suivi d'objet nécessite un mécanisme de détection d'objet, soit dans chaque trame ou lorsque l'objet apparaît d'abord dans la vidéo.

Le suivi d'objets dans une séquence d'images vidéo est un problème qui demande une extraction et un traitement d'informations provenant d'images complexes. Ce problème devient de plus en plus difficile si la contrainte temps réel est exigée. Il existe dans la littérature un grand nombre de méthodes de suivi. Ce nombre est dû, d'une part au nombre important de problèmes à régler et d'autre part à la diversité des types d'applications concernées par le suivi. Ainsi, chacune des méthodes peut traiter certains aspects et échoue sur d'autres.

2. Définition d'un objet vidéo

Intuitivement, un objet vidéo est défini par une forme, une texture et un mouvement (rigide ou non rigide). Cependant, la notion d'objet vidéo est beaucoup plus descriptive qu'une simple région, un objet vidéo peut par exemple être un modèle 3D.

La notion d'objet ne fait pas forcément référence à un objet du monde réel. En effet, dans le domaine de la vidéo, un objet n'est pas nécessairement un objet d'une scène 3D mais plutôt le résultat de l'analyse de la projection d'un monde 3D sur un plan. Ainsi, un objet vidéo est défini comme une région de la vidéo conforme à un modèle. On peut par exemple avoir pour modèle:

- un modèle de mouvement.
- un modèle d'objet physique

La notion d'objets est alors subjectivement définie par rapport aux modèles utilisés. On constate que pour pouvoir trouver des objets vidéo, il est nécessaire de proposer des modèles auxquels ces objets répondront [13]

3. Détection d'objets

La première étape dans un processus de suivi d'objets consiste en la détection des objets à suivre. Chaque méthode de suivi nécessite un mécanisme de détection d'objet, soit sur toutes les trames, soit lorsque l'objet apparaît pour la première fois dans la vidéo. L'approche commune de détection d'objets à suivre est d'utiliser l'information à partir d'une seule image initiale. Cependant, certaines méthodes de détection d'objet utilisent l'information temporelle calculée à partir d'une séquence d'images afin de réduire le nombre de fausses détections. Cette information temporelle est généralement sous la forme de différenciation de trame, qui met en évidence les régions qui changent entre les images [13].

4 . Les méthodes de détection [8]

4.1 La différence temporelle

La différence entre deux images consécutives représente une solution très intéressante. Comme son nom l'indique, elle consiste à soustraire une image acquise au temps T_n d'une autre au temps T_{n+k} , où k est habituellement égal à 1 (cf. Figure II.1). Ainsi, l'image résultante sera vide si aucun mouvement ne s'est produit pendant l'intervalle du temps observé car l'intensité et la couleur des pixels seront presque identiques. Par contre, si du mouvement a lieu dans le champ de vision, la différence d'images correspondant aux pixels frontières des objets en déplacement devraient changer drastiquement de valeurs, révélant alors la présence d'activité dans la scène. Cette technique nécessite très peu de ressources, car aucun modèle n'est nécessaire.

Cela implique qu'il n'y a pas de phase d'initialisation obligatoire avec une scène statique, ce qui procure une très grande flexibilité d'utilisation. De plus, une opération de soustraction d'images requiert très peu de puissance de calcul, lui conférant un avantage supplémentaire.

Par ailleurs, les résultats obtenus avec cette méthode ne sont pas aussi éloquentes que ceux générés en utilisant un modèle statistique de l'arrière-plan. En effet, certains traitements supplémentaires sont nécessaires afin de déterminer la zone en mouvement, car l'information disponible ne concerne que les contours des régions en déplacement (ce qui inclut également les zones intérieures d'un objet).



Figure II.1 Exemple de détection du mouvement avec la méthode de la soustraction d'images consécutives

Avantages :

- Adapte le changement de la scène.
- Détecte la région de mouvement. *

Inconvénients :

- Ne permet pas de détecter le mouvement dans les zones uniformes intérieures à l'objet.

- Ne fonctionne pas dans plusieurs cas, pour différentes raisons, telles que : la présence de bruit du capteur et les changements de luminosité de la scène qui modifient les intensités des pixels.

4.2 Flux optique

Le flux optique est un champ de vecteur des vitesses apparentes des objets de la scène sur le plan image (cf. Figure II.2). Il permet de déterminer le mouvement de chaque pixel appartenant à une image de la séquence d'images.

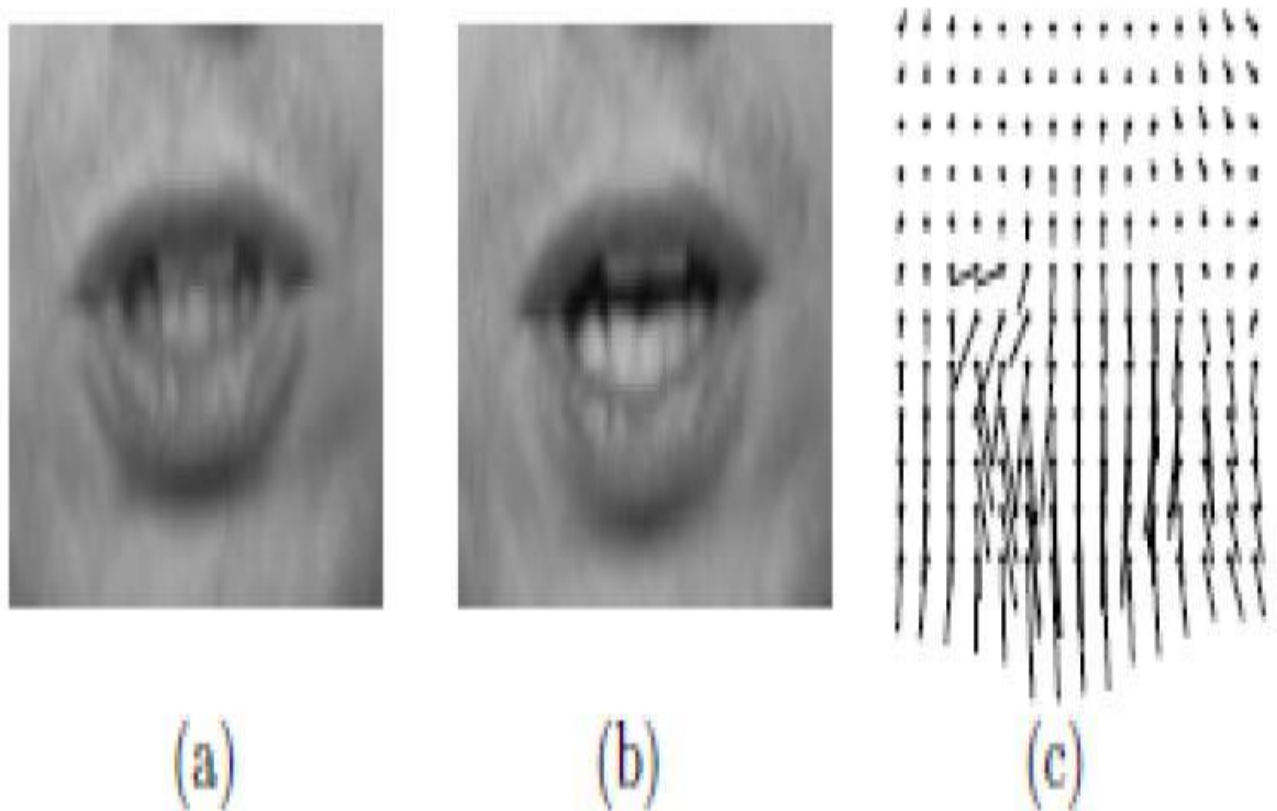


Figure II.2 Le flux optique, (a et b) Région de la bouche sur deux images consécutives, (c) Champ de vecteur de vitesses.

L'estimation du flux optique est basée sur l'hypothèse de l'intensité constante c'est-à-dire que l'intensité observée d'un objet ne change pas lorsque celui-ci est en mouvement. Cette méthode suppose donc, que le mouvement est instantané et faible d'une image à l'autre et que les changements observés entre deux images sont uniquement causés par le mouvement.

Avantages :

- Leurs résultats ne sont pas affectés par les changements soudains de l'éclairage.

Inconvénients :

- Somme importante de calculs à réaliser pour détecter du mouvement

4.3 Soustraction de l'arrière-plan

L'image de l'arrière-plan d'une scène contient les éléments statiques de cette scène, c'est la représentation de l'environnement dans lequel les objets évoluent. Cette image est considérée

comme une image de référence qui se soustrait de chaque image de la séquence afin d'extraire les objets en mouvement.

Cette méthode dans sa version simple, consiste à faire la différence des intensités des pixels $I(x,y, t)$ d'une image I prise à l'instant t , par rapport aux intensités des pixels $B(x,y)$ d'une image de référence d'arrière-plan B , où il n'y a pas de mouvement (cf. Figure II-3) [15].

Elle permet de détecter les pixels (x, y) ayant subi un changement d'intensité significatif. Ces pixels forment les objets en mouvement. Cela est modélisé par :

$$\text{Si } |I(x, y, t) - B(x, y)| > \tau$$

Alors le pixel (x, y) de I à l'instant t est en mouvement

Sinon le pixel (x, y) appartient à l'arrière-plan

Avec τ un seuil fixé manuellement.

C'est cette méthode que nous utilisons pour la partie détection.



Figure II.3 Présentation de La méthode de soustraction de fond.

Avantages

- Cette méthode détecte les objets s'ils sont immobiles
- Détecte l'objet complètement.

Inconvénients

- Le seuillage rend cette méthode peu robuste et les problèmes liés au bruit et aux changements de la luminosité demeurent présents dans cette méthode.
- La difficulté de choisir le meilleur seuil pour obtenir un bon résultat.

4.4 Elimination des ombres

Durant l'opération de segmentation des objets, les ombres générés par le déplacement d'objets peuvent créer d'importants problèmes, car le système est incapable de différencier un objet réel de son ombre. Ce changement d'intensité, provoque une mauvaise segmentation des objets en mouvement. Des chercheurs ont proposé des approches d'élimination d'ombres basées sur l'information couleur.

Les pixels de l'avant-plan, ayant des valeurs d'intensité différentes de ceux du fond, et des valeurs de couleurs normalisées plus proches des valeurs de fond, sont considérés faisant partie d'une zone ombrée (Equation (1)).

$$I_s(x, y) = \alpha I(x, y) \quad \alpha < 1 \quad (1)$$

Où $I(x, y)$ est la valeur de l'intensité au point (x, y) et l'indice s indique la valeur après détection d'ombre.

5. Le suivi des objets [13]

Le suivi d'objet, dans sa forme la plus simple, est l'estimation de la trajectoire d'un objet en mouvement dans le plan image. En d'autres termes, le suivi d'objet est l'estimation de la localisation de l'objet dans chacune des images d'une séquence vidéo et initialement détecté sur la première. Le procédé de localisation se base sur la reconnaissance de l'objet d'intérêt à partir d'un ensemble de caractéristiques visuelles telles que la couleur, la vitesse, la forme où un objet est une zone de l'image qui peut être modélisée par des contours, silhouettes, primitives géométriques tel un rectangle englobant l'objet d'intérêt, ou encore par des points d'intérêt.

Toutes les méthodes de suivi comportent deux couches techniques :

- la première permet de détecter l'ensemble des candidats potentiels (objets similaires à la cible suivie) dans chaque image de la séquence
- et la seconde effectue la mise en correspondance d'une image à l'autre d'un de ces candidats avec la cible afin de maintenir la cohérence du suivi au fil du temps.

Plusieurs méthodes de suivi d'objet ont été proposées. La plupart des méthodes qui permettent le suivi d'objets visent à identifier l'objet d'intérêt en utilisant la technique de la soustraction du fond pour identifier les objets en mouvement dans les séquences vidéo. Ces méthodes ne peuvent pas être appliquées au cas dans lequel la caméra se déplace à cause du changement de contexte. Les autres méthodes basées sur les paramètres de l'objet consistent à reconnaître l'objet par son modèle et estimer sa position par quelques techniques. Les paramètres à estimer peuvent être divers, mais comprennent principalement une composante géométrique, indiquant la position dans l'image du centre de l'objet. Ces méthodes de suivi nécessitent la détection seulement à l'initialisation du suivi quand un objet apparaît pour la première fois.

Après initialisation, la détection et le suivi s'effectuent conjointement. Récemment, le suivi d'objets visuels a été posé comme un problème de suivi par détection, comme illustré dans la figure II.4, où la modélisation statistique est effectuée dynamiquement pour supporter la détection d'objet.

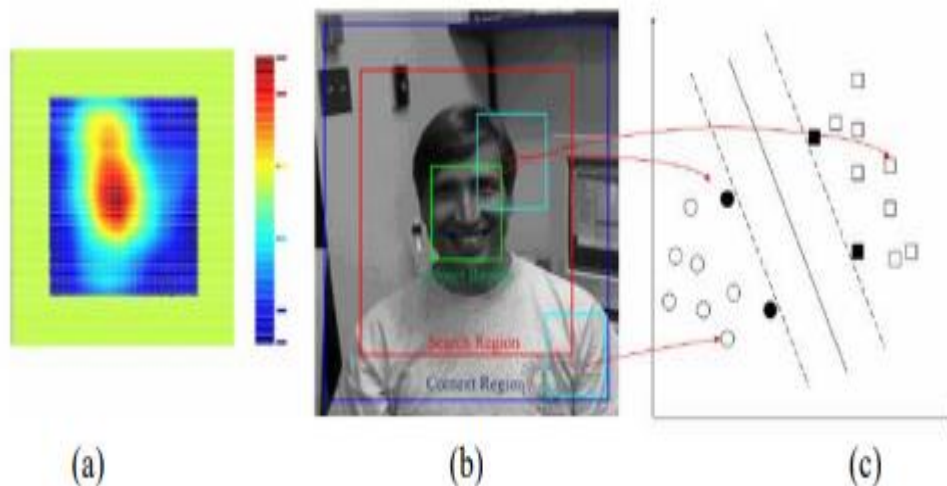


Figure II.4 – Illustration du suivi par la détection basée sur la classification des SVM . (a) montre la carte de score de visage / non-visage classification; (b) affiche la zone de recherche et de la région de contexte pour; (c) trace l'hyperplan de classification.

Les méthodes de suivi par détection sont devenues l'un des paradigmes les plus efficaces pour le suivi d'objets visuels, et ont obtenu des performances à la pointe de la technologie. Cela est dû, en partie, au succès des composants des algorithmes de détection d'objets.

5.1 Domaines d'application [13]

Le suivi d'objets dans des séquences vidéo a suscité un grand intérêt dans la dernière décennie en raison de la variété de ses domaines d'application, tels que :

- La vidéosurveillance (détection, suivi, reconnaissance du comportement de personnes, d'intrus),
- Le militaire (suivi des cibles ou guidage de missiles)
- La vidéoconférence (suivi des interlocuteurs),
- La gestion et l'analyse du trafic (le suivi d'une voiture ou des bords d'une route depuis une caméra embarquée sur un véhicule),
- La robotique : suivie d'obstacles pendant une phase d'évitement,
- Suivi d'indices visuels dans une tâche asservie par vision,
- Suivi d'un opérateur (corps, visage, main...) pour définir des modes d'interaction évolués entre l'homme et la machine,...etc.

- Imagerie médicale.

La figure II.5 illustre quelques exemples d'applications.

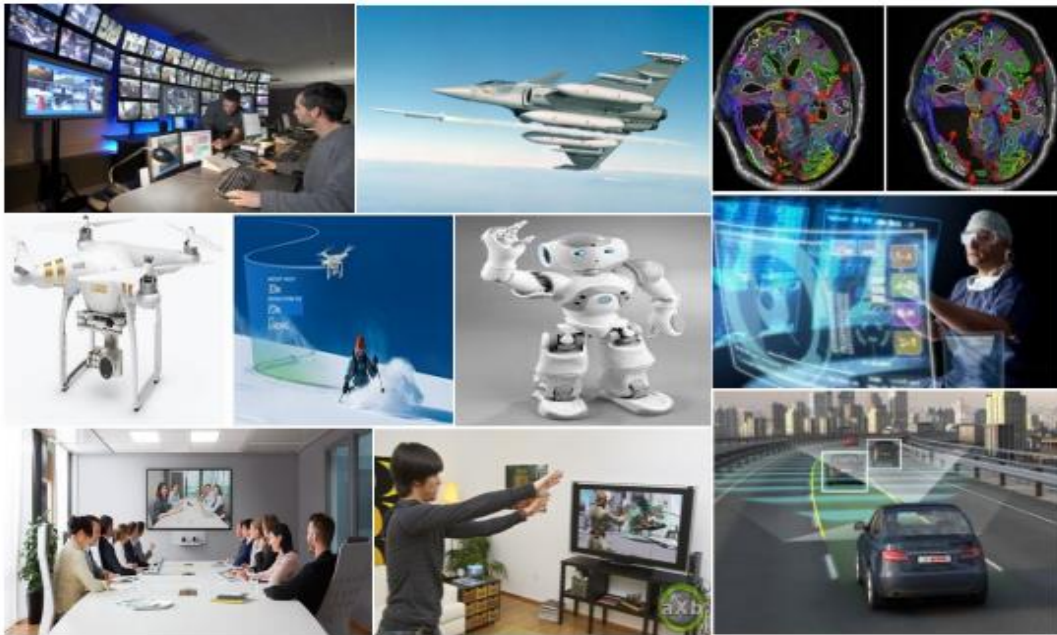


Figure II.5– Quelques applications de suivi d'objets

5.2 Les défis du suivi d'objet [13]

Bien que le suivi d'objets a été étudié depuis plusieurs décennies, et beaucoup de progrès ont été réalisés au cours des dernières années il reste un problème très difficile. Il n'existe aucune approche de suivi unique qui peut gérer avec succès tous les scénarios.

De nombreux facteurs influençant la performance de l'algorithme de suivi d'objets sont les suivantes (cf. Figure II.6) :

- Changements d'illumination ;
- Changements d'échelle ;
- Occultations partielles ou totales ;
- Similarité de couleur entre l'objet cible et le fond ;
- Mouvement de caméra ;
- Rotation dans le plan de l'image ;
- Objets non-rigides et/ou articulés ;
- Objet de petite taille ;
- Objet en mouvement rapide ;
- Présence de bruit dans les images ;
- Nécessité d'un suivi en temps réel.

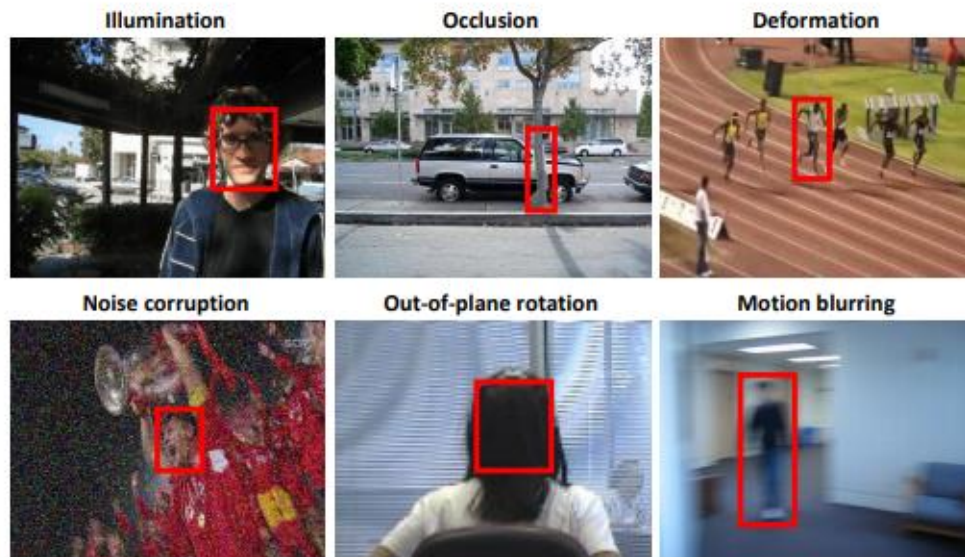


Figure II.6 – Quelques difficultés de suivi d'objet

6. les méthodes de suivi [14]

6.1 Suivi de silhouette

La représentation d'un tel objet par une silhouette permet de tenir compte précisément de la forme de l'objet. Le but des méthodes de suivi fondées sur l'utilisation de silhouettes est d'estimer la silhouette des objets d'intérêt pour chaque image de la vidéo. Cette approche est également connue sous le nom de segmentation d'image.

La segmentation d'images est une opération de traitement d'images qui a pour but de rassembler des pixels entre eux suivant des critères prédéfinis. Les pixels sont ainsi regroupés en régions qui constituent une partition de l'image. Il s'agit en général de séparer les objets du fond. Si le nombre de classes est égal à deux, elle est appelée aussi binarisation. Chez l'homme cette opération de séparer des objets dans une image se fait naturellement grâce à des connaissances de haut niveau fondées sur la compréhension des objets et de la scène, compréhension liée à l'apprentissage.

Mettre au point des algorithmes de segmentation utilisant une information sémantique pour chaque région de l'image est encore un des thèmes de recherche d'actualité en traitement d'images.

6.2 Suivi de noyaux

Les méthodes de suivi de noyaux représentent l'objet à suivre par un rectangle ou une ellipse. L'objet à suivre est représenté par un histogramme pondéré calculé sur une région contenant l'objet.

Le principal avantage de cette méthode est qu'elle permet d'estimer la position de l'objet en un nombre restreint d'itérations. Toutefois, dans le cas d'une scène contenant des objets ayant les mêmes distributions de couleurs et occupant des régions très proches, l'estimation sera ambiguë.

Pour y remédier on utilise une famille d'histogrammes élaborés à différentes échelles d'une image.

6.3 Suivi de points

L'objet est représenté par un point qui est son centroïde ou par un ensemble de points. Généralement, cette représentation par points convient pour suivre des objets qui occupent de petites régions dans une image. Dans le cas où, à chaque instant, toutes les cibles visibles peuvent être détectées de façon fiable et rapide, le problème de suivi peut être défini comme un problème d'appariement de détections entre images successives. Dans cette catégorie, les approches peuvent être déterministes ou probabilistes.

L'algorithme de suivi que nous utilisons dans le cadre de ce travail est basé sur le filtre de Kalman.

7. Le filtre de Kalman[14]

Le filtre de Kalman, du nom de son créateur Rudolph E. Kalman, mathématicien américain né en 1930, a été présenté en 1960. Il s'agit d'un estimateur optimal de processus aléatoires.

Il permet l'estimation de variables d'état d'un système (position/vitesse, par exemple, pour des systèmes mécaniques ou encore tension/intensité, par exemple, pour des systèmes électriques...) et il fournit aussi la variance de l'erreur de l'estimation. Son appellation « filtre » vient du fait qu'il permet, entre autre, d'estimer un signal bruité (bruit blanc), en séparant le signal de son bruit.

Le filtre de Kalman s'applique aux signaux non stationnaires, il résout le problème de filtrage linéaire et se présente sous la forme d'un ensemble d'équations différentielles ou récurrentes plus faciles à résoudre sur calculateur numérique en présence de condition initiales et d'entrées déterministe.

Ce filtre est utilisé en trajectographie pour estimer une trajectoire comme le cas de poursuite de cibles (avions, missiles, radars...), en navigation pour localiser un engin, ou encore en balistique pour estimer des conditions initiales inconnues.

Il est également employé lorsque l'on veut coupler des informations venant de sources différentes afin de retrouver, par exemple, une trajectoire.

Le Filtre de Kalman est une approche statistique, d'assimilation de données, dont le principe est de corriger la trajectoire du modèle en combinant les observations avec l'information fournie par le modèle de façon à minimiser l'erreur entre l'état vrai et l'état filtré.

7.1 Principe du filtre de Kalman [14]

Le filtre de Kalman donne l'estimation d'un processus en utilisant un système à retour sous la forme de mesures (bruitées). De cette manière, les équations du filtre de Kalman se regroupent en deux types d'équations :

- Les équations de prédiction (extrapolation).
- Les équations de mise à jour (correction).

Les équations d'extrapolation sont responsables de propager les estimées de l'état présent et les covariances d'erreur pour obtenir les estimées à priori de la prochaine étape.

Alors que les équations de correction sont responsables du retour, c'est à dire, introduire une nouvelle mesure avec l'estimée à priori afin d'obtenir une estimée à posteriori améliorée.

De ce fait, le filtre de Kalman ressemble finalement à un filtre prédicteur-correcteur pour la résolution des problèmes numériques (cf.figure III.7).

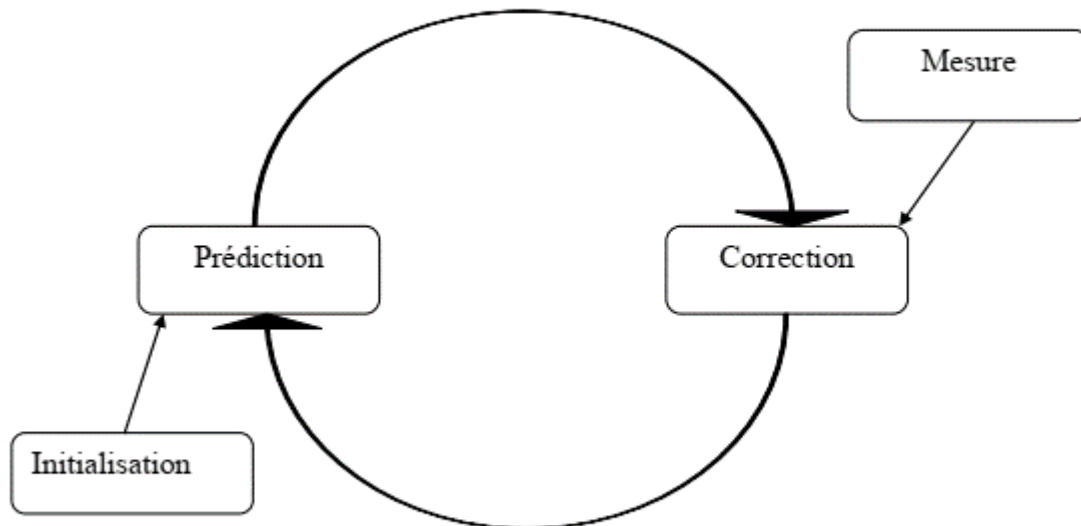


Figure II.7 le cycle du filtre de Kalman

F

7.2. Suivi de mouvement par filtre de Kalman [14]

Dans le suivi du mouvement, le filtre de Kalman est une méthode de prédiction qui a suscité l'intérêt de plusieurs chercheurs. Un filtre de Kalman est essentiellement un ensemble d'équations récursives, elles sont utilisées ensemble 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 de 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 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 de 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 prochaine trame. La position de l'objet est corrigée ensuite en considérant la prédiction et aussi l'observation.

8. Conclusion

Dans ce chapitre, nous avons présenté quelques algorithmes de détection et de suivi pour différentes situations tout en mettant le point sur leurs avantages et inconvénients.

Le choix d'un algorithme particulier dépend de l'application et de la nature de la vidéo traitée.

Le filtre de Kalman, utilisé dans le cadre de ce travail, est décrit et son principe pour le cas particulier du suivi de cible est explicité.

Chapitre III

Résultats de détection et poursuite des objet en mouvement

1. Introduction

Nous sommes intéressés dans notre projet de fin d'étude par la conception d'une application de détection de n'importe quel objet en mouvement dans une scène vidéo capturée et enregistrés par une webcam.

2. Environnement du travail

Dans cette section, nous présentons les environnements matériels et logiciels de notre travail.

2.1 Environnement matériel ou Hardware:

Le matériel utilisé est constitué d'un laptop dont les caractéristiques sont les suivantes :

Marque : HP Elite book 820 G3

Processeur Intel(R) Core(TM) i5-6300U CPU @ 2.40GHz 2.50 GHz

RAM installée 8.00 GB (7.88 GB usable)

Disque d'ur : 500 GB SSD

Nom de Webcam: 'HP HD Camera'

Résolutions disponibles : {'640x480' '320x180' '320x240' '424x240' '640x360'}

2.2 Environnement logiciel

Système d'exploitation : Windows 10 Pro- 64 bits

Notre application a été développée sous MATLAB(R2018a).

Nous avons choisi MATLAB pour les avantages suivants :

- Programmation infiniment plus rapide pour le calcul et pour l'affichage
- Une librairie très riche
- Code facile à comprendre et très lisible
- Possibilité d'inclure un programme en C/C++
- Langage interprété : Pas de compilation donc pas d'attente pour compiler
- Possibilité d'exécuter du code très lisible
- Une aide très bien faite

3. Architecture du système

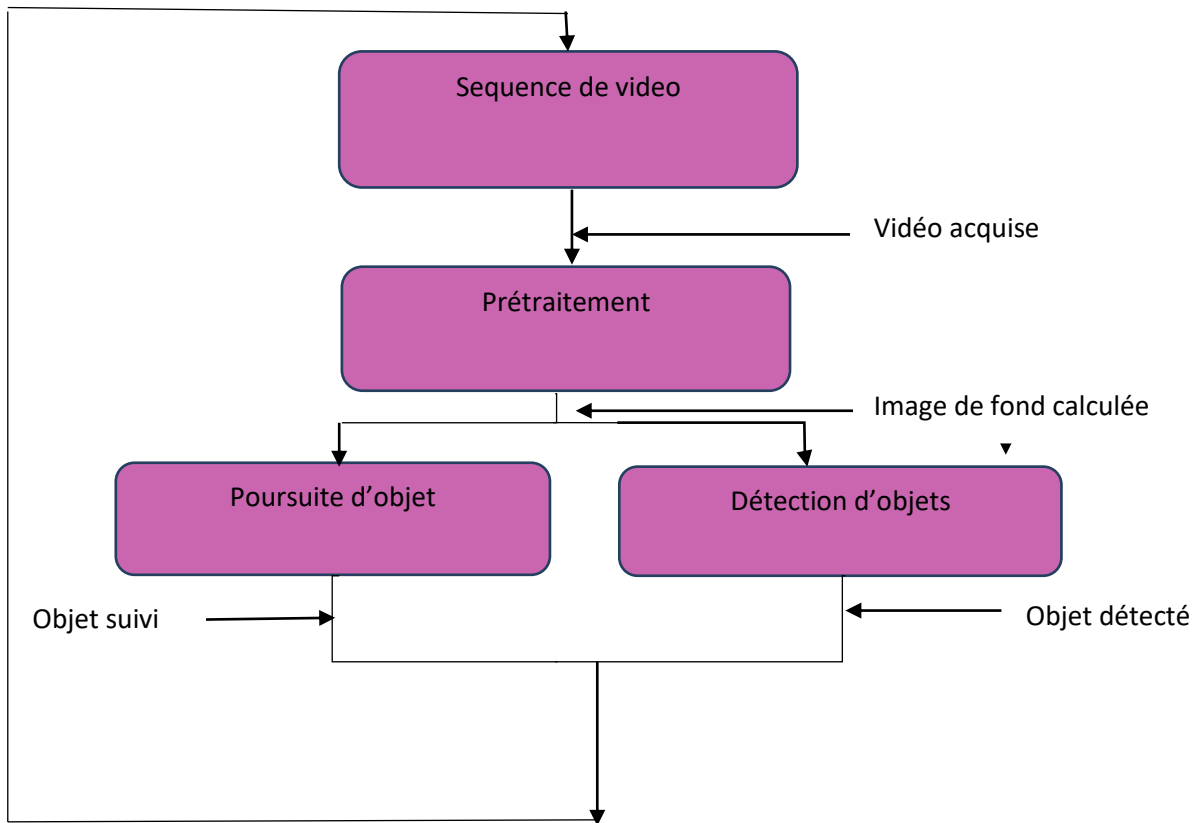


Figure III.1 Architecture général du système.

Les principales phases de notre système sont illustrées sur la figure III.1

3.1 Acquisition des vidéos

La vidéo utilisée pour la détection et le suivi de l'objet est acquise à l'aide d'une caméra webcam intégrée .

3.2 Prétraitement

Le prétraitement est composé de deux phases principales :

- Segmenter la vidéo en frames (images).
- Calcul de l'image de fond à l'aide de la moyenne de 5 premières images.

3.3 Détection des objets en mouvement

La détection des objets en mouvement (Figure III.2) se fait par soustraction de fond (soustraction d'arrière-plan) dont le principe consiste à faire la différence des intensités des pixels $I(x, y, t)$ d'une image I prise à l'instant t , par rapport aux intensités des pixels $B(x,y)$ d'une image de référence d'arrière-plan B , où il n'y a pas de mouvement.

Ceci permet de détecter les pixels (x, y) ayant subi un changement d'intensité significatif. Ces pixels forment les objets en mouvement.

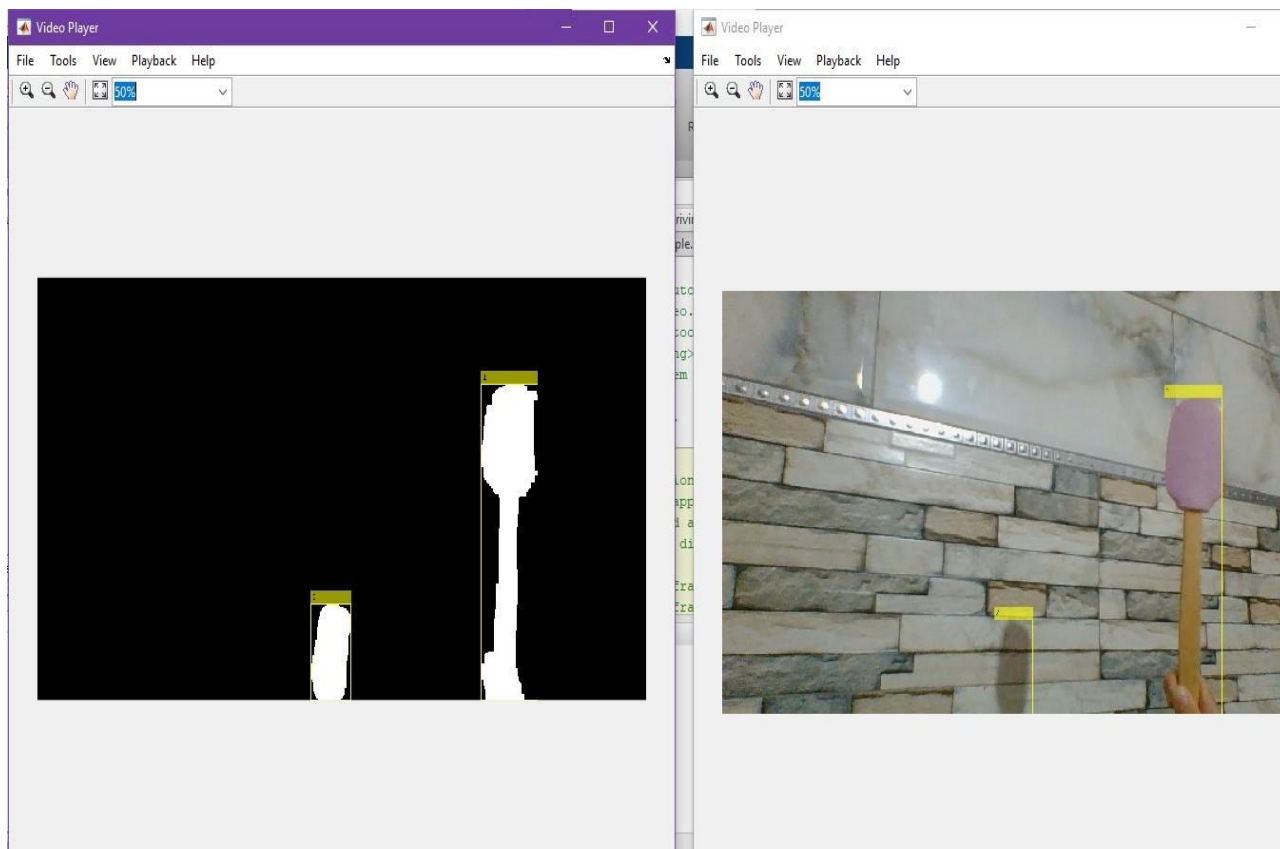


Figure III.2 Objet mobile détecté

Comme on peut le voir, l'objet ainsi que son ombre sont détectés vu que l'ombre aussi est mobile.

3.4 Suivi avec filtre kalman

Les principales phases du filtre de Kalman sont illustrées sur la figure III.3

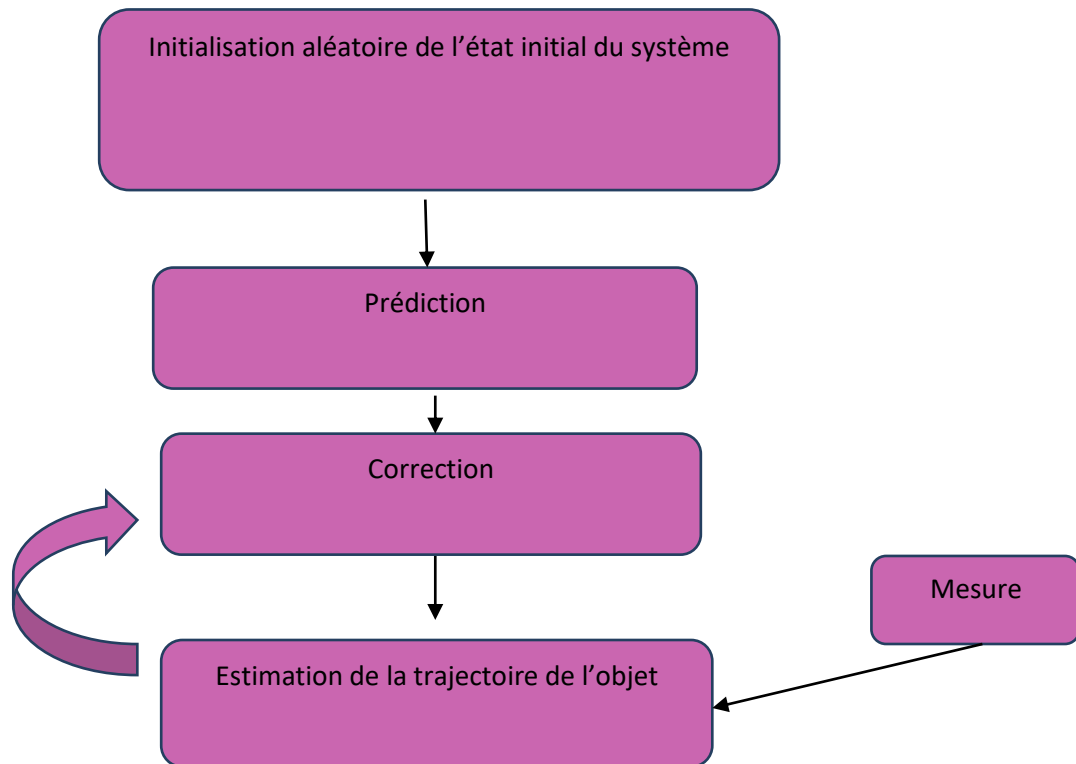


Figure III.3 Architecture du filtre de Kalman

Définir un filtre Kalman revient à passer par les étapes suivantes :

Étape 1 : Définir le modèle de mouvement et l'état

Étape 2 : Définir le bruit du processus

Étape 3 : Définir le modèle de mesure

Étape 4 : Initialiser le vecteur d'état en fonction de la mesure du capteur*

Étape 5 : Définir une covariance d'état initiale

Étape 6 : Créer le bon filtre

Dans notre cas, tous les modèles sont linéaires, alors on a utilisé la fonction '**trackingKF**' .

La fonction '**detectObjects**' renvoie les centroïdes et les boîtes limitantes.

4. Tests et Résultats

4.1 A partir de la webcam

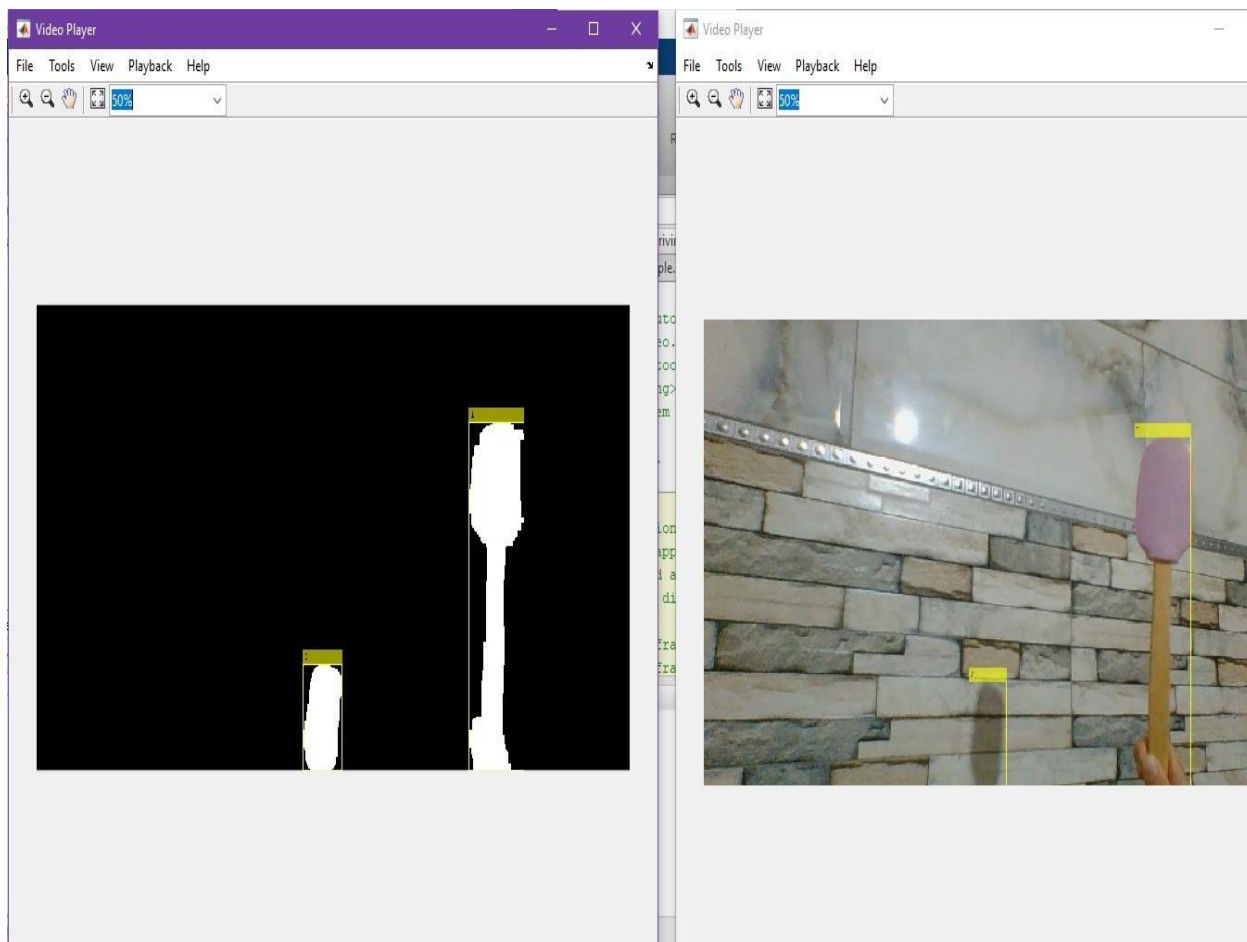


Figure III.4 Resultat 1

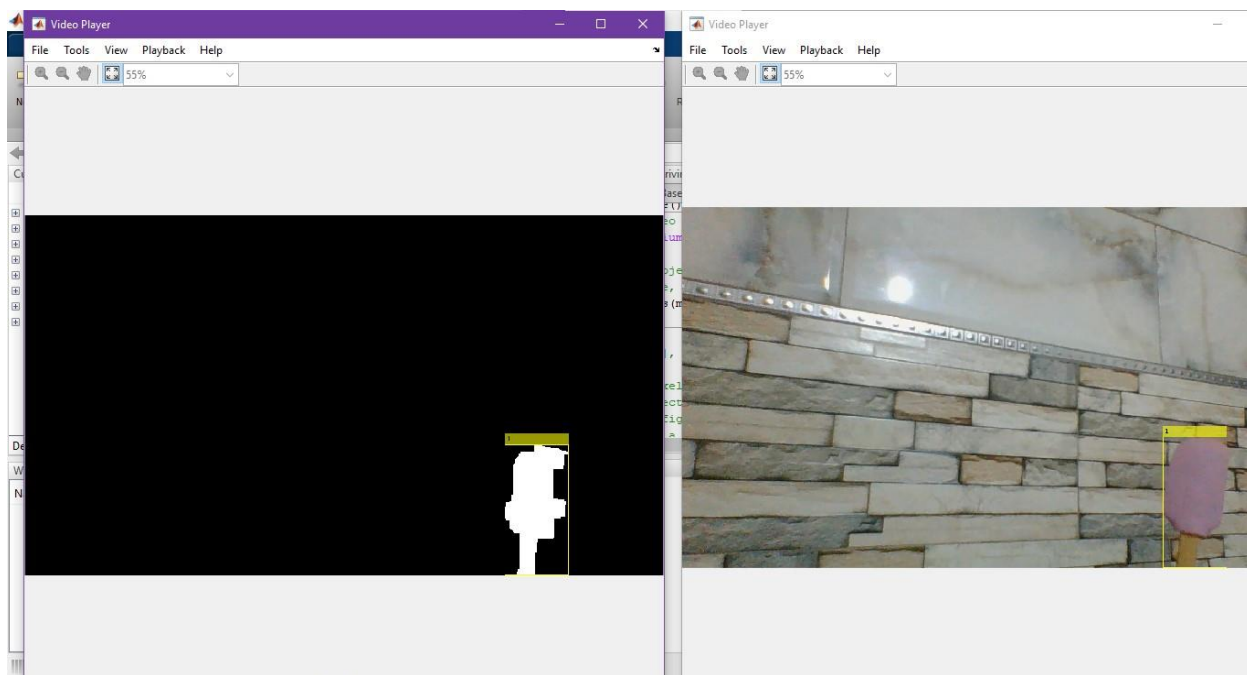


Figure III.5 Resultat 2

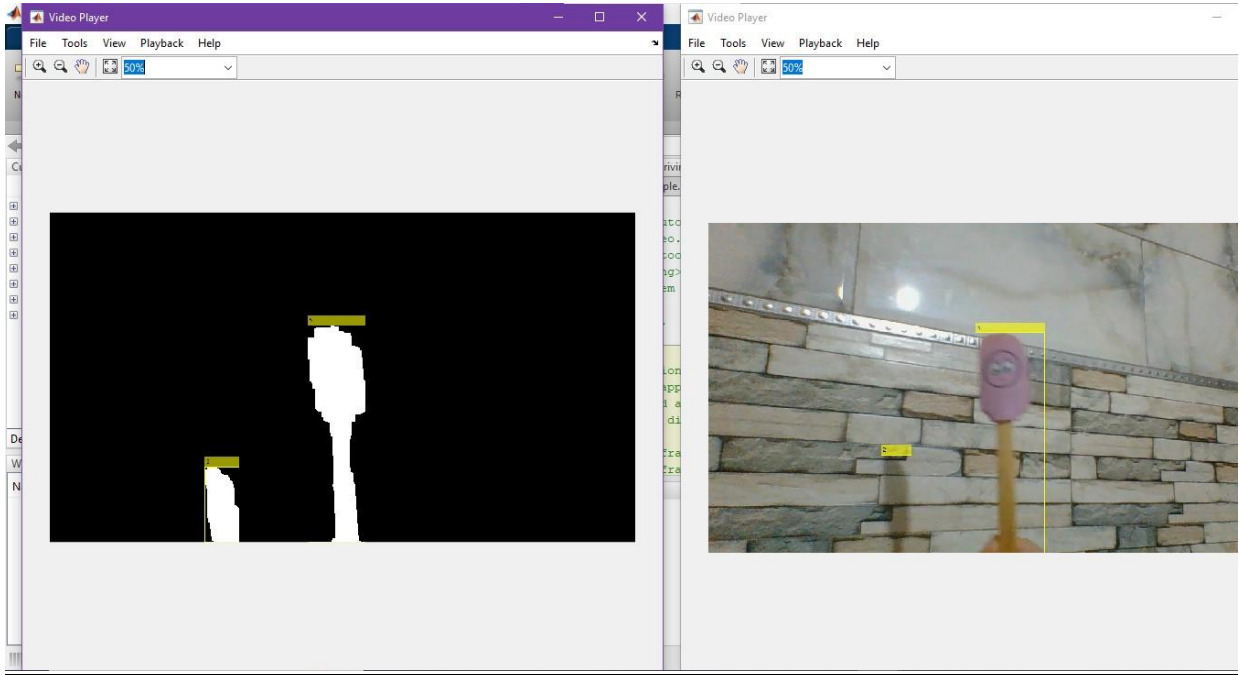


Figure III.6 Resultat 3

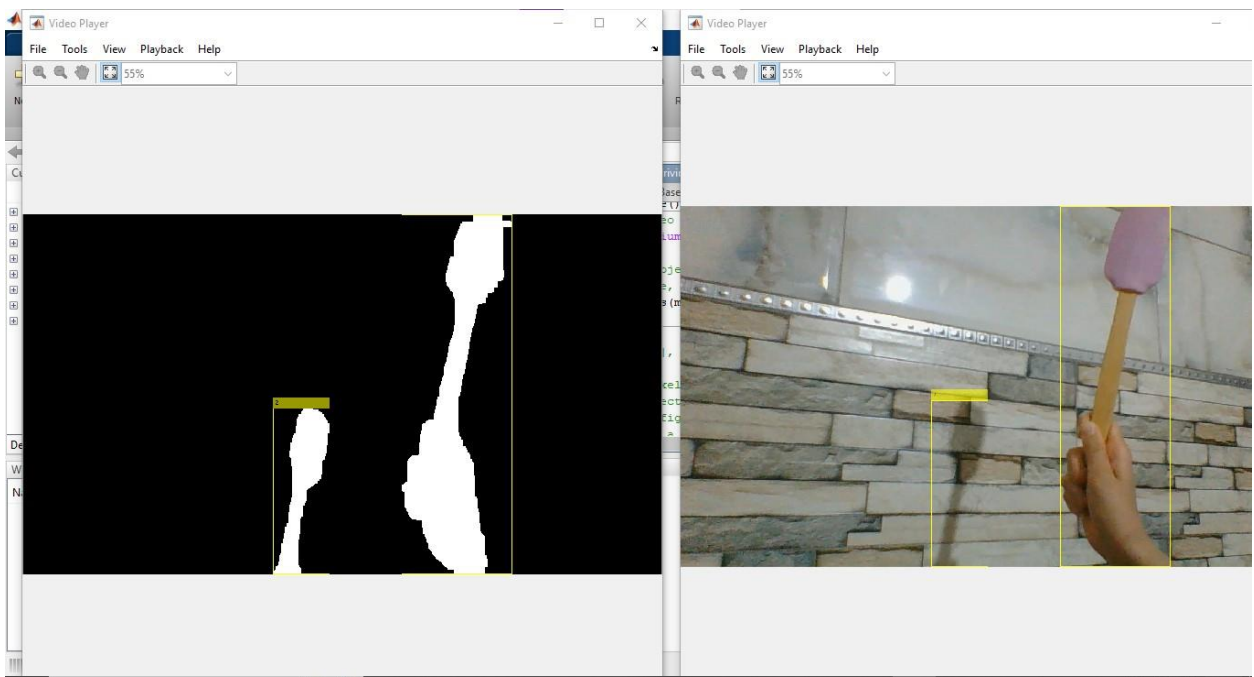


Figure III.7 Resultat 4

4.2 Vidéo téléchargée (Youtube)

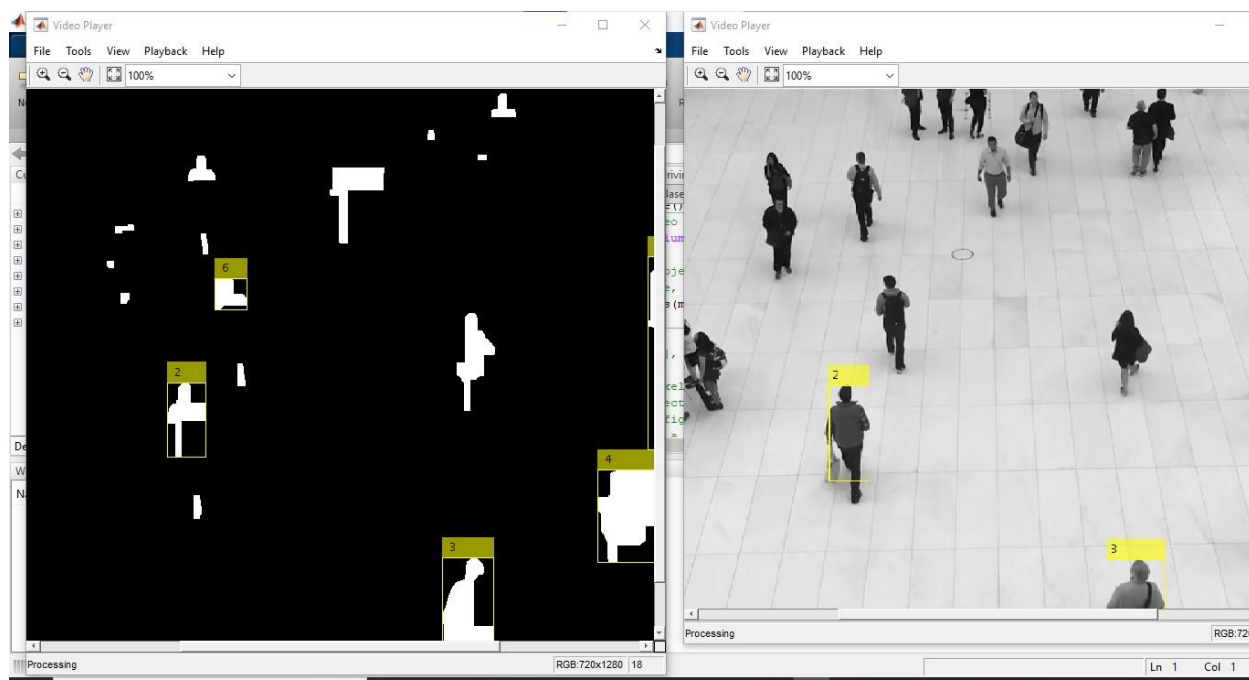


Figure III.8 Résultat 1

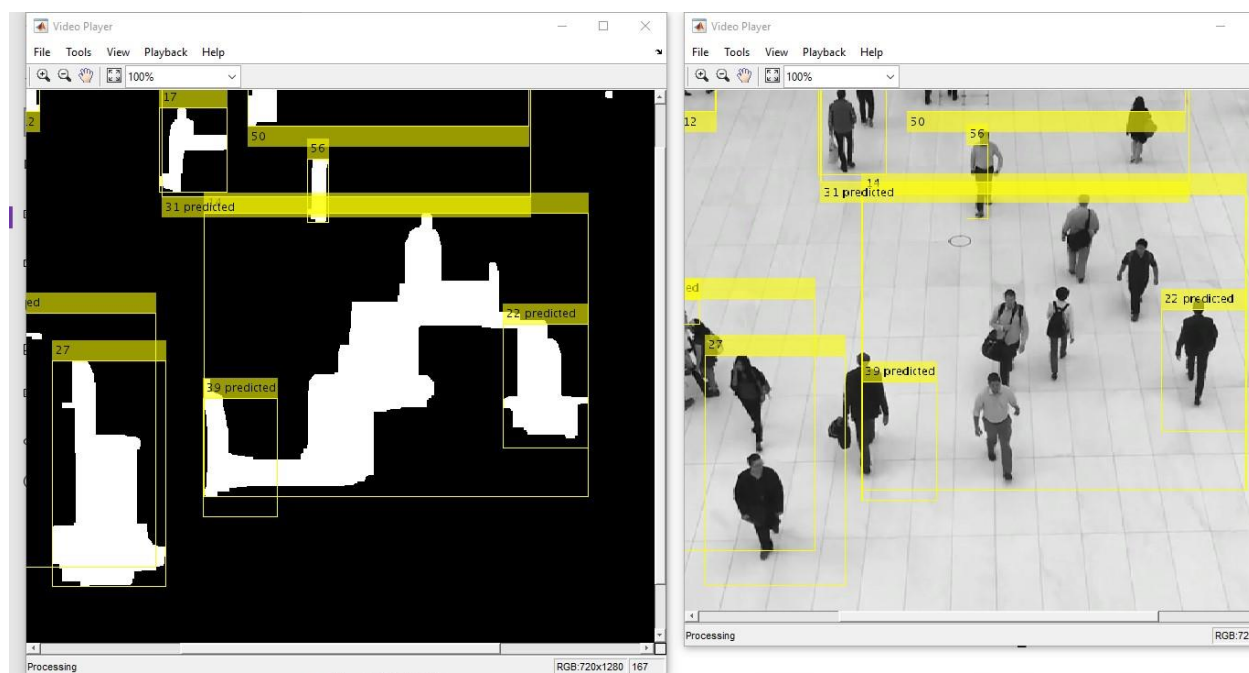


Figure III.9 Résultat 2



Figure III.10 Résultat 3

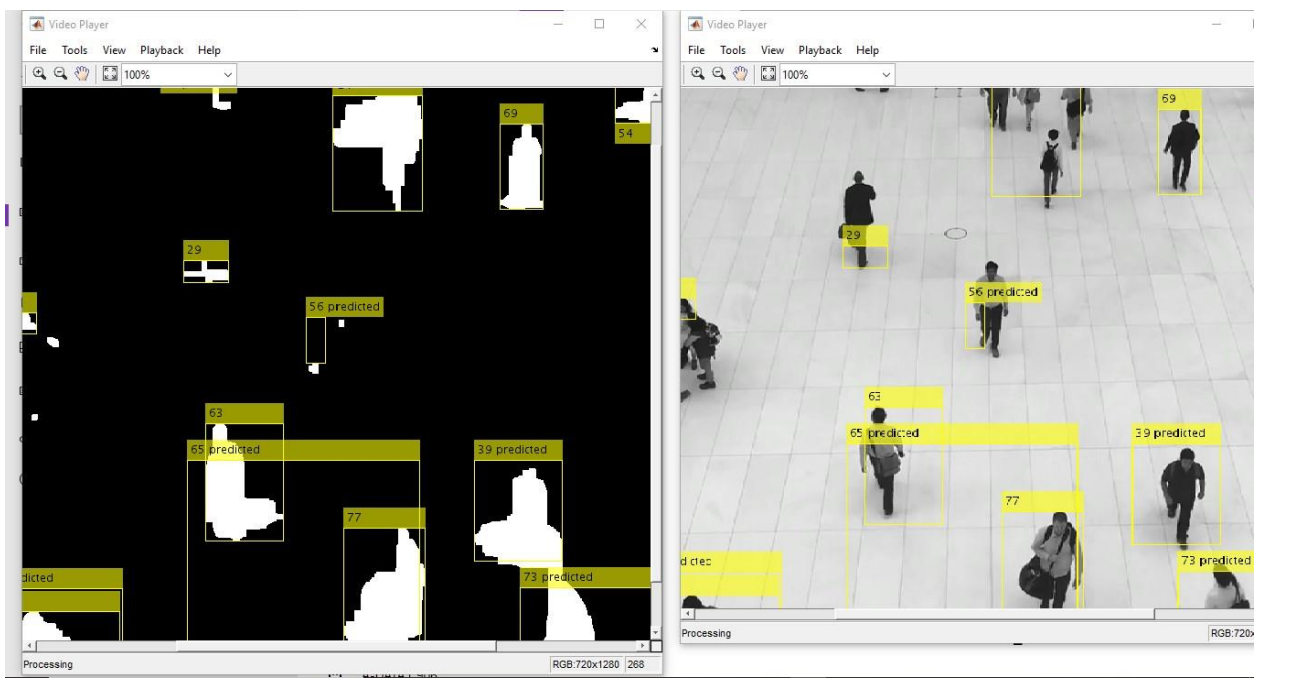


Figure III.11 Résultat 4

4.3 Vidéo capturée sur téléphone

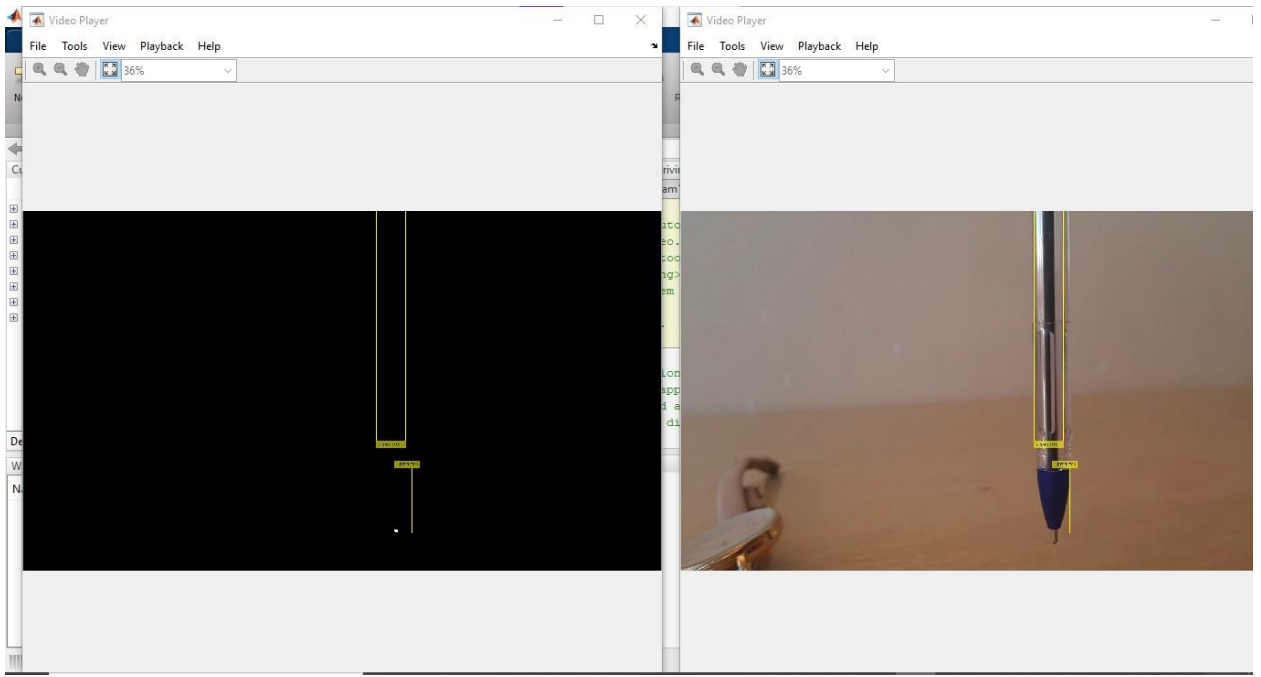


Figure III.12 Résultat 1

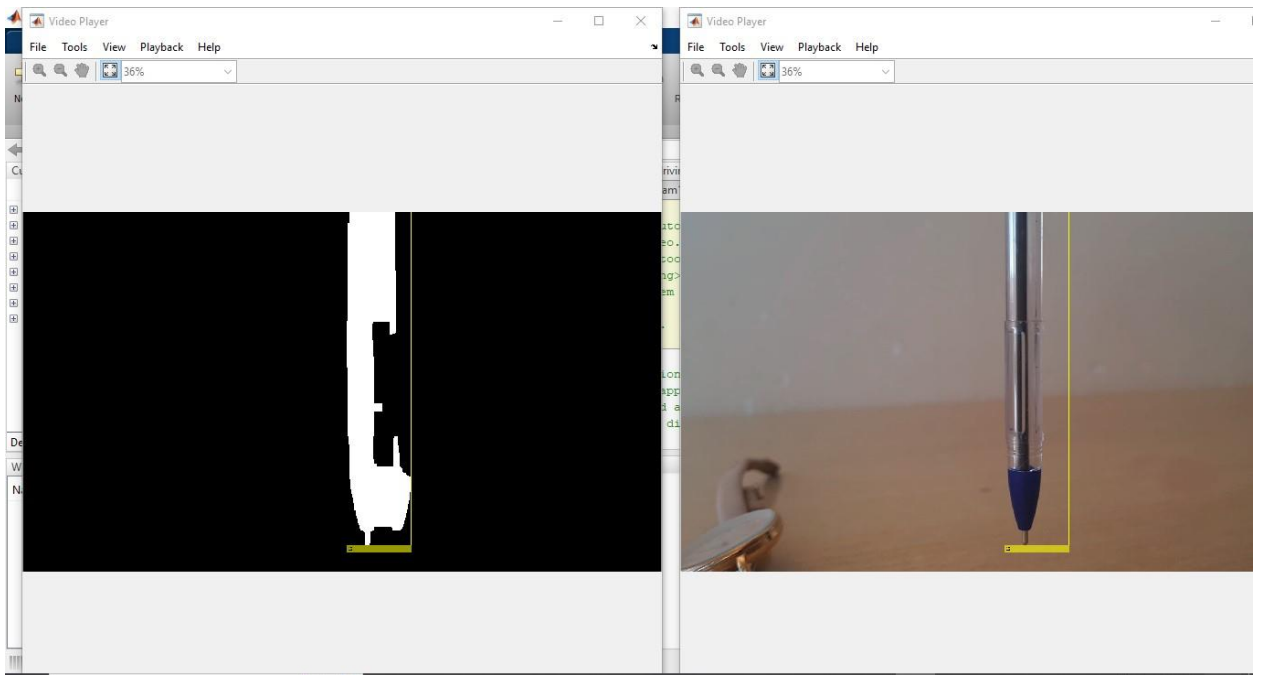


Figure III.13 Résultat 2

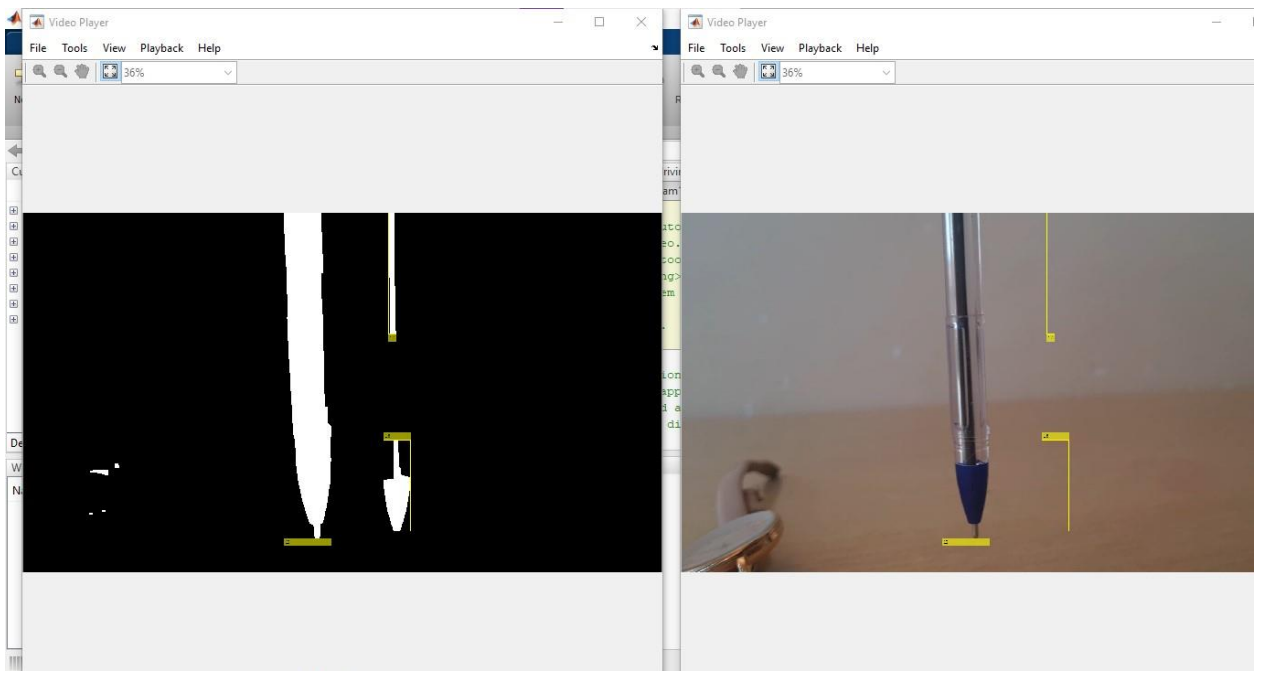


Figure III.14 Résultat 3

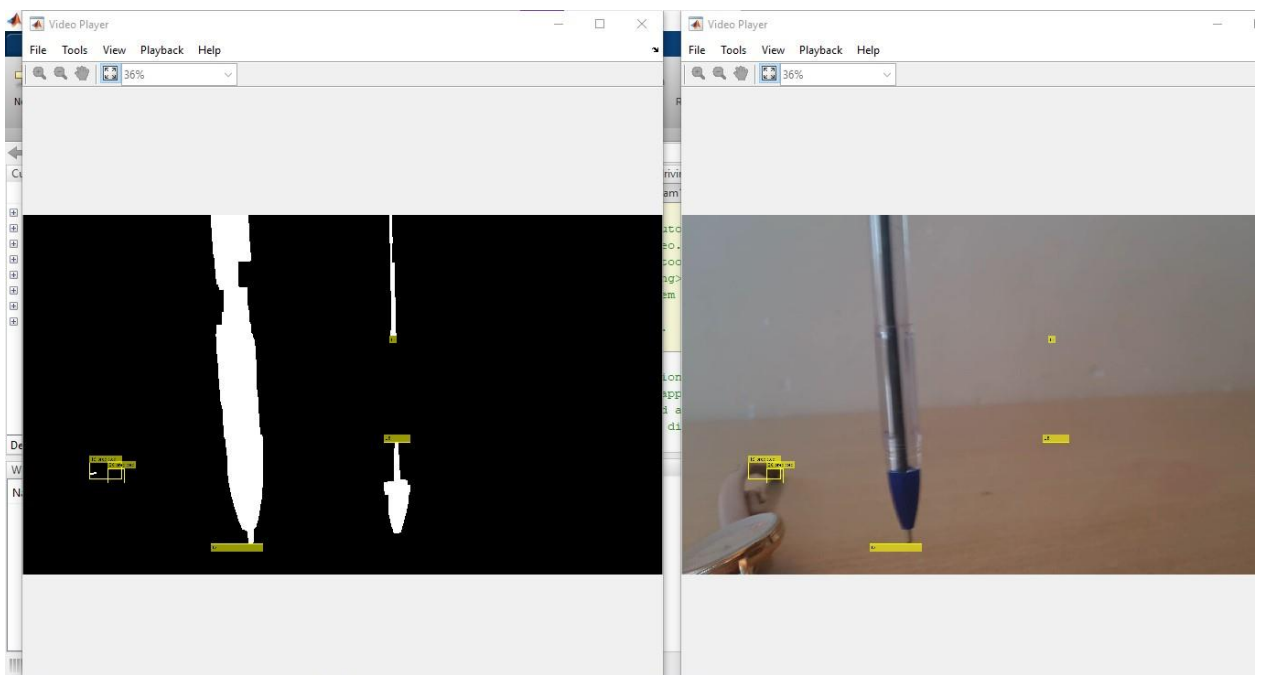


Figure III.15 Résultat 4

5 Discussion des résultats :

- L'objet est détecté mais même l'ombre est détectée, ce qui soulève le problème d'éclairage et de l'angle de la prise de vue pour y remédier.
- La stabilité (l'immobilité) du système d'acquisition de la vidéo est très importante pour mieux détecter l'objet. Ainsi qu'on peut le voir, l'utilisation du téléphone introduit beaucoup de bruit alors que les résultats avec la vidéo téléchargée du net sont assez bons.

6 Conclusion :

Dans ce chapitre, nous avons présenté les différents résultats que nous avons obtenus en intégrant les algorithmes de détection et poursuite développés dans le chapitre 2 de ceux de poursuite. La combinaison de ces deux types d'algorithmes permet de détecter et suivre les objets en mouvement dans une séquence vidéo.

La première étape a porté donc sur la détection de l'objet en mouvement par la méthode de soustraction d'arrière-plan qui donne des bons résultats dans le cas d'un mouvement rapide.

La deuxième étape a traité le problème de suivi d'objets multiples en mouvement basé sur l'estimation de la trajectoire d'objet mobile utilisant le filtre de Kalman linéaire. Les résultats de cette partie sont acceptables point de vue précision de l'estimation de la trajectoire du mouvement et temps de calcul requis.

Conclusion générale :

Dans ce travail, nous nous sommes intéressés au problème de détection et suivi d'objets dans une séquence de vidéo. Nous avons présenté des notions de base sur les images et les vidéos. Puis, nous avons étudié les différentes méthodes de détections et suivi.

Le problème de détection et suivi d'objet dans les séquences de vidéo est, ces dernières décennies, un thème de recherche très actif dans le monde de la vision par ordinateur. L'objectif est de définir un ensemble d'opérations génériques qui peuvent être utilisées quel que soit le domaine d'application du système de suivi.

L'étude de plusieurs exemples nous a permis d'identifier deux étapes fondamentales : la détection de mouvement et le choix du modèle du filtre de kalman.

Le choix du filtre de Kalman est une bonne approche pour estimer l'état d'un processus. Selon les paramètres, en particulier le processus de bruit d'erreur, le filtre peut rapprocher les valeurs de l'état réelles des deux systèmes statiques et dynamiques.

Dans notre cas, vous avons implémenté un algorithme basé sur le filtre de Kalman linéaire via la fonction **trackingKF** de Matlab pour détecter et suivre plusieurs objets en mouvement avec la supposition que tous les objets se déplacent en ligne droite avec une vitesse constante. Lorsque le mouvement d'un objet s'écarte de manière significative de ce modèle, des erreurs de suivi peuvent être produite.

Une extension de ce travail pourrait être l'emploi d'un filtre non linéaire pour prendre en compte des mouvements plus complexes avec une accélération non nulle et des trajectoires non rectilignes.

Néanmoins, ce que l'on peut affirmer d'après le tour d'horizon sur la littérature qu'à ce jour, il n'y a aucun tracker capable de maîtriser toutes les situations difficiles pouvant apparaître lors du suivi d'un objet : changements d'illumination, d'échelle, occultations, mouvement de la caméra, déformation d'objet, la similarité entre l'objet et le fond, etc.

Bibliographie

- [1] Rafael C. Gonzalez, Richard E. Woods et Steven L. Eddins, "Digital Image Processing Using MATLAB", Pearson Prentice-Hall, 2004
- [2] A. Gagalowic, de cours dispensés à l'ESIEA, 2002.
- [3] Benchrif Ali, " Traitement d'image ", Université Abou Bakar Belkaid Tlemcen, 2008.
- [4] Chérif TAOUCHE, " Implémentation d'un Environnement Parallèle pour la Compression d'Images à l'aide des Fractales", Université Mentouri Constantine, 2005
- [5] Zeroual Djazia, "Implémentation D'un Environnement Parallèle Pour La Compression D'Image A L'Aide Des Fractales", Université de Batna, 2006.
- [6] Thi-Lan le, "Indexation et recherche de vidéo pour la vidéo surveillance", université de NICE-SOPHIA ANTIPOLIS, 3 février 2009
- [7] Djamila Mokhtari, " Détection des chutes par calcul homographique ", Université de Montréal Département d'Informatique et de Recherche Opérationnelle Faculté des Arts et des Sciences, Août 2012.
- [8] MEDJAHED Fatiha, " Détection et Suivi d'Objets en Mouvement Dans Une Séquence d'Images ", Université Mohamed Boudiaf Oran, 2011
- [9] David Salomon, "Data-compression-the-complete-reference ", 2007.
- [10] Ahmed Ben Atitallah, "Etude et Implantation d'Algorithmes de Compression d'Images dans un Environnement Mixte Matériel et Logiciel", Université Bordeaux 1, Juillet 2007.
- [11] Ahmed Ben Atitallah, " Etude et Implantation d'Algorithmes de Compression d'Images dans un Environnement Mixte Matériel et Logiciel", thèse de doctorat en électronique, Université Bordeaux 1, Juillet 2007.
- [12] Nicolas DUMOULIN, "Compression de séquences vidéo choix des images de références ", Université de Rennes1-IFSIC, Juin 2003.
- [13] MEDOUAKH Saadia; ' Détection et suivi d'objets', Université Mohamed Khider – Biskra, 2019
- [14] Sadoun abdelbaki & Ouellabi yacine , 'Détection et suivi d'objets mobiles. Application dans un environnement de foule', UNIVERSITÉ ECHAHID HAMMA LAKHDAR D'EL OUED, 2016
- [15] _HACHEMI Fouzia; ' La détection et suivi des objets en mouvement dans une scène vidéo en utilisant la bibliothèque OpenCV', Université Abou Bakr Belkaid– Tlemcen, 2016

