



Université Sidi Mohammed Ben Abdellah
Faculté des Sciences Dhar El Mahraz- Fès
Centre d'Etudes Doctorales
"Sciences et Technologies"

Formation Doctorale : STIC

Discipline : Informatique

Spécialité : Informatique

Laboratoire : LESSI

THESE DE DOCTORAT

Présentée par
YOUNES DHASSI

**SUIVI VISUEL D'OBJETS PAR MODELE D'APPARENCE
MULTICRITERES ET ESTIMATION ADAPTATIVE**

Soutenue le 21 Décembre 2018 devant le jury composé de :

<i>Pr. Abdelmajid SAKA</i>	<i>Ecole Nationale des Sciences Appliquées FES</i>	<i>Président</i>
<i>Pr. MY Abdelouahed SABRI</i>	<i>Faculté des sciences dhar el mehraz Fes</i>	<i>Rapporteur</i>
<i>Pr. Mohammed OUANAN</i>	<i>Faculté des Sciences et Techniques Errachidia</i>	<i>Rapporteur</i>
<i>Pr. Mohamed SABANE</i>	<i>Faculté des Sciences Meknes</i>	<i>Rapporteur</i>
<i>Pr. Hassan QJIDAA</i>	<i>Faculté des sciences dhar el mehraz Fes</i>	<i>Examineur</i>
<i>Pr. El Hassan SBAI</i>	<i>Ecole supérieure de technologie Meknes</i>	<i>Examineur</i>
<i>Pr. Abderrahim SAAIDI</i>	<i>Faculté Polydisciplinaire, Taza</i>	<i>Examineur</i>
<i>Pr. Abdellah AARAB</i>	<i>Faculté des sciences dhar el mehraz Fes</i>	<i>Directeur de thèse</i>

REMERCIEMENTS

Cette thèse n'aurait aboutie sans le soutien et l'indulgence de mon directeur de thèse, le professeur Abdellah AARAB. Je tiens à le remercier pour m'avoir accepté dans son groupe de recherche, d'avoir cru en mes capacités, et aussi pour le temps et la patience qu'il m'a accordés durant toute la période de préparation de cette thèse. Ses exigences fonctionnelles pour un travail de qualité, ses conseils ont été utiles et fructueux.

J'adresse mes remerciements les plus sincères

Au Professeur Abdelmajid SAKA pour m'avoir fait l'honneur de présider cette thèse.

Aux Professeurs My Abdelouahed SABRI, Mohammed OUANAN, Mohamed SABBANE pour avoir accepté et m'accordé du temps pour rapporter sur le présent travail.

Aux Professeurs Hassan QJIDAA, Abdarrahim SAAIDI et El Hassan SBAI pour avoir accepté d'examiner cette thèse.

À tous les membres de laboratoire LESSI et plus particulièrement les Professeurs Hassan QJIDAA et Ismail BOUMHIDI pour leur encouragement sans oublier leur persévérance et leur souci pour maintenir un environnement de recherche sain.

Je tiens à présenter ma profonde gratitude à mes parents, les mots ne peuvent pas exprimer à quel point je leur suis reconnaissant pour m'avoir apporté un soutien sincère et des encouragements continus tout au long du processus de recherche et de rédaction de cette thèse. Cet accomplissement n'aurait pas été possible sans eux.

Mes pensées affectueuses vont à mon fils Yahya.

Enfin, j'aimerais remercier ma famille et mes amis qui m'ont soutenu et encouragé durant ce travail.

TABLE DES MATIÈRES

LISTE DES FIGURES	4
LISTE DES TABLEAUX	7
LISTE DES ABREVIATIONS	8
RESUME	10
ABSTRACT	12
INTRODUCTION GENERALE	16
1 PROBLEMATIQUES	16
2 CONTRIBUTIONS	18
3 STRUCTURE DE LA THESE	19
CHAPITRE I : ETAT DE L'ART	22
I.1. GENERALITE SUR LE SUIVI VISUEL	23
I.2. CARACTERISTIQUES POUR LE SUIVI VISUEL	26
I.2.1. CARACTERISTIQUES DE COULEUR	26
I.2.2. GRADIENT	27
I.2.3. TEXTURE	29
I.2.4. SPATIO-TEMPORELS	31
I.2.5. LA FUSION DE PLUSIEURS CARACTERISTIQUES.....	32
I.3. CLASSIFICATION DES METHODES DE SUIVI VISUEL	33
I.3.1. LES METHODES DISCRIMINATIVES.....	33
<i>1.3.1.1. Méthode basé sur la représentation de de l'arrière-plan</i>	34
<i>1.3.1.2. Méthodes discriminatives par apprentissage en ligne</i>	35
I.3.2. LES METHODES DE SUIVI GENERATIVES	40
<i>1.3.2.1. Méthodes basées sur des modèles</i>	40
<i>1.3.2.2. Méthodes basées sur l'analyse par sous-espace</i>	44
<i>1.3.2.3. Méthodes basées sur la représentation parcimonieuse</i>	47
I.3.3. METHODES COMBINEES.....	52
I.4. CONCLUSION	55
CHAPITRE II : METHODES BAYESIENNES POUR LE SUIVI VISUEL	57
II.1. INTRODUCTION	58
II.2. FILTRE BAYESIEN	60
II.2.1. PREDICTION BAYESIENNE.....	60
II.2.2. LE FILTRE DE KALMAN.....	62
II.2.3. LE FILTRE H^∞	64
II.2.4. LE FILTRE A PARTICULES.....	68
II.3. FILTRE H^∞-PF POUR LE SUIVI VISUEL	69
II.4. EVALUATION	72
II.5. CONCLUSION	80

CHAPITRE III : MODELISATION DE L'APPARENCE DE L'OBJET SUIVI	81
III.1. INTRODUCTION	82
III.2. REPRESENTATION DE L'OBJET SUIVI	83
III.3. MODELE D'APPARENCE BASE SUR LA COULEUR ET LA TEXTURE	87
III.4. MISE A JOUR DU MODELE D'APPARENCE	103
III.5. CONCLUSION	104
CHAPITRE IV : SUIVI VISUEL PAR LE FILTRE ADAPTATIF $H_{\infty}PF$	1045
IV.1. INTRODUCTION	105
IV.2. ESTIMATION GLOBALE.....	106
IV.2.1. MODELISATION	106
IV.2.2. ESTIMATION	109
IV.2.3. MISE A JOUR ADAPTATIVE.....	111
IV.2.4. ETUDES DES PERFORMANCES.....	112
IV.3. ESTIMATION LOCALE	115
IV.3.1. ECHANTILLONNAGE ET PONDERATION	115
IV.3.2. ESTIMATION	116
IV.3.3. MISE A JOUR DES COEFFICIENTS DE FUSION	117
IV.4. ALGORITHMME DE SUIVI	118
IV.5. EXPERIMENTATIONS ET DISCUSSIONS	122
IV.5.1. SEQUENCES VIDEO	122
IV.5.2. CRITERES DE PERFORMANCE.....	124
IV.5.3. RESULTATS QUALITATIFS.....	125
IV.5.4. RESULTATS QUANTITATIFS	138
IV.6. CONCLUSION	145
CONCLUSION ET PERSPECTIVES.....	146

LISTES DES FIGURES

Figure 1.	<i>Différents problèmes pouvant survenir lors du suivi. (a) Occlusion; (b) variation d'illumination; (c) arrière-plan encombré; (d) Variation d'échelle; (e) Déformation; (f) Mouvement rapide; (g) Rotation dans le plan; (h) Flou de mouvement</i>	17
Figure 2.	<i>Représentation schématique de la structure de la thèse</i>	21
Figure 3.	<i>Organigramme d'un algorithme de suivi visuel</i>	24
Figure 4.	<i>Quelques applications du suivi visuel</i>	24
Figure II-1.	<i>Un modèle d'état d'un objet en suivi visuel, avec x, y sont les positions cartésiennes, v est la vitesse, a est l'accélération</i>	59
Figure II-2.	<i>Un modèle d'état d'un objet en suivi visuel, avec x_c, y_c sont les positions cartésiennes, la taille de la zone suivi est définie par w, h et l'orientation</i>	59
Figure II-3.	<i>Diagramme du filtre bayésien</i>	62
Figure II-4.	<i>Diagramme du filtre de Kalman</i>	63
Figure II-5.	<i>Diagramme du filtre H^∞</i>	65
Figure II-6.	<i>Diagramme du filtre à particule</i>	68
Figure II-7.	<i>Diagramme du filtre H^∞ PF</i>	70
Figure II-8.	<i>Trajet réel</i>	73
Figure II-9.	<i>Trajetoires estimées par les méthodes de comparaison et vérité du terrain</i>	73
Figure II-10.	<i>Comparaison de l'erreur de localisation du centre (en pixels)</i>	74
Figure II-11.	<i>Comparaison de l'Erreur de Localisation Moyenne Normalisée pour différent nombre de particules</i>	75
Figure II-12.	<i>Comparaison de l'Erreur de Localisation du centre pour différentes valeurs de \square</i>	75
Figure II-13.	<i>Comparaison de l'Erreur de Localisation Moyenne Normalisée différentes valeurs de \square</i>	76
Figure II-14.	<i>Trajet réel</i>	77

<i>Figure II-15.</i>	<i>Trajectoires estimées par les méthodes de comparaison et vérité du terrain</i>	77
<i>Figure II-16.</i>	<i>Comparaison de l'erreur de localisation du centre (en pixels)</i>	78
<i>Figure II-17.</i>	<i>Comparaison de l'Erreur de Localisation Moyenne Normalisée pour différent nombre de particules</i>	79
<i>Figure II-18.</i>	<i>Comparaison de l'Erreur de Localisation du centre pour différentes valeurs de γ</i>	79
<i>Figure II-19.</i>	<i>Comparaison de l'Erreur de Localisation Moyenne Normalisée différentes valeurs de γ</i>	80
<i>Figure III-1.</i>	<i>Différentes représentations de la cible. (a) boîte, (b) ellipse, (c) contour, (d) point d'intérêt</i>	83
<i>Figure III-2.</i>	<i>Tracé de surface pour (a) GMM et (b) α-GMM avec $\alpha=-5$</i>	90
<i>Figure III-3.</i>	<i>La distribution α-Gaussien pour différentes valeurs de α</i>	90
<i>Figure III-4.</i>	<i>Répartition spatiale des données générées aléatoirement</i>	91
<i>Figure III-5.</i>	<i>Données originales et fonctions de distribution de probabilité estimées du GMM</i>	92
<i>Figure III-6.</i>	<i>Données originales et fonctions de distribution de probabilité estimées du α-GMM</i>	92
<i>Figure III-7.</i>	<i>Le profil du noyau Epanechnikov pour les positions de pixel normalisées</i>	93
<i>Figure III-8.</i>	<i>Le profil du noyau Gaussien pour les positions de pixel normalisées</i>	94
<i>Figure III-9.</i>	<i>GMM Histogramme et α-GMM Histogramme basés sur les caractéristiques H et S du mode HSV</i>	95
<i>Figure III-10.</i>	<i>LBP opérateur appliqué sur un voisinage 3x3</i>	96
<i>Figure III-11.</i>	<i>Exemples de l'opérateur LBP étendu: les voisinages circulaires (8,1), (16,2) et (24,3)</i>	96
<i>Figure III-12.</i>	<i>Exemples de motifs uniformes dans le cas d'un voisinage circulaires (8, 1)</i>	97
<i>Figure III-13.</i>	<i>Image LBP et histogramme LBP pour le même objet sous différents niveau de luminosité</i>	99
<i>Figure III-14.</i>	<i>Image ULBP et histogramme ULBP pour le même objet sous différents niveau de luminosité</i>	99
<i>Figure III-15.</i>	<i>Image LBP et histogramme LBP pour le même objet sous différentes orientations</i>	100

<i>Figure III-16</i> · Image ULBP et histogramme ULBP pour le même objet sous différentes orientations.	100
<i>Figure III-17</i> · Représentation de la composante V du mode HSV par l'histogramme ULBP.	101
<i>Figure III-18</i> · Exemples de motifs non uniformes dans le cas d'un voisinage circulaires (8,1).	101
<i>Figure III-19</i> · Division de l'objet en blocs (exemple de division en 2×2+1 blocs).	102
<i>Figure III-20</i> · Représentation des blocs par l'histogramme ULBP.	102
<i>Figure IV-1</i> · Structure de notre Algorithme de suivi visuel.....	105
<i>Figure IV-2</i> · Comparaison entre PF, KPF et AH ∞ PF en termes d'erreur de localisation du centre (en pixels)	114
<i>Figure IV-3</i> · Organigramme de la méthode proposée.....	120
<i>Figure IV-4</i> · Séquences vidéo utilisées dans les tests des expérimentations.....	123
<i>Figure IV-5</i> · Résultats de suivi pour la séquence Soccer-Ball	125
<i>Figure IV-6</i> · Trajectoires estimées et vérité du terrain pour la séquence Soccer-Ball.....	126
<i>Figure IV-7</i> · Résultats de suivi pour la séquence Ping-Pong.....	126
<i>Figure IV-8</i> · Trajectoires estimées et vérité du terrain pour la séquence Ping Pong	127
<i>Figure IV-9</i> · Résultats de suivi pour la séquence Skater	127
<i>Figure IV-10</i> · Trajectoires estimées et vérité du terrain pour la séquence Skater	128
<i>Figure IV-11</i> · Résultats de suivi pour la séquence Woman	128
<i>Figure IV-12</i> · Trajectoires estimées et vérité du terrain pour la séquence Woman	129
<i>Figure IV-13</i> · Résultats de suivi pour la séquence Car1	129
<i>Figure IV-14</i> · Trajectoires estimées et vérité du terrain pour la séquence Car1	130
<i>Figure IV-15</i> · Résultats de suivi pour la séquence Fish.....	130
<i>Figure IV-16</i> · Trajectoires estimées et vérité du terrain pour la séquence Fish	131
<i>Figure IV-17</i> · Résultats de suivi pour la séquence Jumping	131
<i>Figure IV-18</i> · Trajectoires estimées et vérité du terrain pour la séquence Jumping	132
<i>Figure IV-19</i> · Résultats de suivi pour la séquence David	133

<i>Figure IV-20</i> ·	Trajectoires estimées et vérité du terrain pour la séquence David	133
<i>Figure IV-21</i> ·	Résultats de suivi pour la séquence Deer	134
<i>Figure IV-22</i> ·	Trajectoires estimées et vérité du terrain pour la séquence Deer	134
<i>Figure IV-23</i> ·	Résultats de suivi pour la séquence Mhyang	135
<i>Figure IV-24</i> ·	Trajectoires estimées et vérité du terrain pour la séquence Mhyang	135
<i>Figure IV-25</i> ·	Résultats de suivi pour la séquence Car4	136
<i>Figure IV-26</i> ·	Trajectoires estimées et vérité du terrain pour la séquence Car4	136
<i>Figure IV-27</i> ·	Résultats de suivi pour la séquence Man	137
<i>Figure IV-28</i> ·	Trajectoires estimées et vérité du terrain pour la séquence Man	137
<i>Figure IV-29</i> ·	Performances de suivi par AH ∞ PF, de l'IMM-PF, du HKPF et du WLTMS en termes d'erreur de position	139
<i>Figure IV-30</i> ·	Performances de suivi par AH ∞ PF, de l'IMM-PF, du HKPF et du WLTMS en termes d'erreur de position	140
<i>Figure IV-31</i> ·	140

LISTE DES TABLEAUX

<i>Tableau IV-1</i> · Erreur de localisation moyenne normalisée (en pixels).	115
<i>Tableau IV-2</i> · Temps d'exécution moyen (en second).....	115
<i>Tableau IV-3</i> · Aspects de suivi des séquences vidéo utilisées dans les expérimentations.....	123
<i>Tableau IV-4</i> · Précision de suivi. Erreur de localisation moyenne normalisée (ELMN) entre le centre cible suivi et le centre d'objet de vérité terrain est présenté pour les méthodes de comparaison.....	143
<i>Tableau IV-5</i> · Précision du suivi. La précision de suivi moyenne (PSM) entre la fenêtre englobante estimée et celle de la vérité de terrain est présentée pour les méthodes de comparaison.....	143
<i>Tableau IV-6</i> · Précision de suivi. F1 score est présenté pour les méthodes de comparaison	144
<i>Tableau IV-7</i> · Précision du suivi. Taux de réussite (TR) pour les méthodes de comparaison	144
<i>Tableau IV-7</i> · Temps de calcul. Temps d'exécution moyens pour les méthodes de comparaison (s/image)	145

LISTE DES ABRÉVIATIONS

ACP	Analyse de Composante Principale
A-GMM	Alpha-Modèle de Mélange Gaussien
AH ∞ PF	Adaptative H-infini Filtre à Particules
AMBP	Adaptatif Motif Binaire Médian
CLBP	Coefficients du Motif Local Binaire
DBP	Motif Binaire Différentielle
EBGM	Graphique du Bouquet de Correspondance Elastique
ELC	Erreur de Localisation du Centre
ELMN	Erreur de Localisation Moyenne Normalisée
EM	Maximisation de l'Espérance
EMD	Maximisation Différentielle de l'Espérance
GMM	Modèle de Mélange Gaussien
H ∞	H-infini
H ∞ F	Filtre H-infini
H ∞ PF	H-infini Filtre à Particules
HOG	Histogramme des Gradients Orientés
HSV	Teinte-Saturation-Lumière
ICA	Analyse des Composantes principales
IMM	Interactive Multiple Modèles
IMM-PF	Filtre à Particules Interactif Multiple Modèles
KCF	Filtre de Corrélation à Noyau
KF	Filtre de Kalman
KLT	Kanade-Lucas-Tomaci
KPCA	Analyse de Composante Principale à Noyau
KPF	Kalman Filtre à Particules
LASSO	Moindre Rétrécissement Absolu et Opérateur de Sélection
LBP	Motif Local Binaire
LCG	Graphe Local Connecté
L-K	Lucas-Kanade
LPQ	Quantification de Phase Locale
LQE	Estimation Quadratique Linéaire
MACP	Modélisation de l'apparence par Codage Parcimonieux
MB-LBP	Blocs Multi échelle Motif Local Binaire
MBP	Motif Binaire Médian
MFID	Descripteur Conjoint Multifonctions
MIL	Apprentissage Multi-Instances
MRID	Descripteur de Rotation Mixte
MS	Mean Shift
NCA	Vitesse Quasi Constante
NCP	Prédiction de Chaîne de Voisinage
NCV	Accélération Quasi Constante
PF	Filtre à Particules
PLS	Moindre Carrés Partiels
PSM	Précision de suivi Moyenne

QBP	Motif Binaire Quadrilatère
RCRP	Recherche de Cible par Représentation Eparsée
RRSC	Codage Crypté Robuste Régularisé
RVB	Rouge-Vert-Bleu
SA	Accélération de Singer
SBP	Motif Structurel Binaire
SIFT	Transformation de Caractéristiques Visuelles Invariante
STC	Contexte Spatio Temporel
STI	Système de Trafic Intelligent
SURF	Caractéristiques Robustes Accélérées
SVI	Système de Surveillance Intelligent
SVM	Support Vecteur Machine
TR	Taux de Réussite
ULBP	Motif Binaire Locale Uniforme
WLTM	Probabilité Pondérée du Suivi des Mélanges Gaussien

RÉSUMÉ

Cette thèse propose une méthode de suivi visuel précise et stable, elle présente des solutions pour surmonter les majeurs problèmes rencontrés dans le suivi visuel, surtout dans les scénarios non contraints. Notre travail touche deux modules essentiels dans tout système de suivi visuel : Module d'estimation et module de modélisation.

Dans ce travail de thèse, de nouvelles approches pour le suivi visuel dans le cas des séquences vidéo présentant des situations contraignantes telles que la variabilité du comportement dynamique de la séquence, les changements de luminosité et l'encombrement de l'arrière-plan, sont proposées.

La première approche concerne l'estimation du mouvement et repose sur la combinaison séquentielle d'une prédiction globale et d'une prédiction locale. La prédiction globale utilise le filtre H_∞ qui ne nécessite aucune hypothèse sur les perturbations et minimise l'erreur d'estimation du pire cas. Elle contribue à une atténuation de la dégénérescence des particules notamment dans le cas où la cible effectue un mouvement complexe. Quant à la prédiction locale, le filtre à particules est utilisé pour gérer efficacement les situations non linéaires et non gaussiennes. Le filtre à particules exploite l'état dynamique global et permet ainsi de réduire le temps de calcul sans pour autant perdre dans la précision de suivi.

Dans le module de modélisation, le mouvement et l'apparence de la cible sont pris en compte. L'approche proposée est de nature adaptative, basée sur deux sous modèles paramétrés dont chacun peut couvrir différents modes de mouvements à savoir les mouvements uniformes et complexes. Les paramètres du système dynamique sont ajustés automatiquement suivant des règles logiques. Un nouveau modèle d'apparence basé sur les deux caractéristiques visuelles de couleur et de texture est élaboré. Dans ce modèle, l'histogramme de modèle de mélange α -gaussien des composantes H et S du système HSV et de l'histogramme distinctif de la composante V basé sur l'opérateur ULBP sont fusionnés de façon adaptative dans le cadre du filtre particulaire.

Plusieurs expérimentations qualitatives et quantitatives sur des séquences vidéo publiques, comprenant des scénarios non contraints, ont été menées. Les résultats issus de ces expérimentations ainsi que des comparaisons avec d'autres algorithmes de la littérature confirment les bonnes performances (précision et stabilité) des approches proposées.

Mots Clés : vision par ordinateur, suivi visuel, détection de mouvement, estimation de mouvement, prédiction Bayésienne, α -GMM.

ABSTRACT

In this work, new approaches are proposed for visual tracking in the case of video sequences presenting constraining situations such as, the dynamic behavior variability of the sequence, the changes in brightness and the background clutter. They deal with the two complementary modules of a tracking procedure: estimation module and modeling module.

The first approach concerns motion estimation and relies on the sequential combination of global prediction and local prediction. The global prediction uses the H^∞ filter which does not require any assumptions about the disturbances and minimizes the worst case estimation error. It contributes to a decrease of the degeneracy of the particles especially in the case where the target carries out a complex movement. As for local prediction, the particle filter is used to effectively achieve nonlinear and non-Gaussian systems. The particle filter exploits the global dynamic state and thus reduces the computation time without losing in tracking accuracy.

In the modeling module, the movement and appearance of the target object are taken into account. The proposed approach is of an adaptive nature, based on two parameterized sub-models, each of which can cover different modes of motion, namely uniform and complex motions. The parameters of the dynamic system are adjusted automatically according to logical rules. A new appearance model based on both visual characteristics of color and texture is developed. In this model, the α -Gaussian mixture model histogram of the H and S components of the HSV system and the distinctive histogram of the component V based on the ULBP operator are merged with adaptive weighting as part of the particulate filter.

Several qualitative and quantitative experiments were conducted on public video sequences, including unconstrained scenarios. The experiments results as well as comparisons with other algorithms of the literature confirm the suitable performances (precision and stability) of the proposed approaches.

Key words : vision par ordinateur, suivi visuel, détection de mouvement, estimation de mouvement, prédiction Bayésienne, α -GMM.

Introduction Générale

La vision par ordinateur est une discipline de l'intelligence artificielle, elle vise à traiter, analyser et comprendre une image ou une séquence d'images acquise par une ou plusieurs caméras ou autre dispositif d'acquisition.

Le suivi visuel constitue une composante principale de plusieurs applications de la vision par ordinateur. Il peut être défini comme l'estimation de la trajectoire d'un ou plusieurs objets, en localisant sa position dans une séquence d'images.

1 Problématiques

Le suivi visuel d'objets est un problème complexe dans le domaine de la vision par ordinateur. Plusieurs défis rendent le suivi visuel robuste très difficile, l'ensemble de ces problèmes est dû à des facteurs intrinsèques ou extrinsèques. La première catégorie comprend la déformation, le changement d'échelles, la rotation dans le plan, le mouvement rapide ainsi que les objets non rigides et articulés. Alors que la deuxième catégorie comprend la variation d'illumination, les objets similaires, l'occultation et le flou.

Développer une méthode de suivi dépend en grande partie de l'application et des informations préalablement disponibles. Le problème est en général traité sous différentes hypothèses et contraintes. Dans le cas où l'objectif visé consiste à suivre certains types ou certaines catégories d'objets (personnes, véhicules), l'apparence des objets peut être utilisée comme information préalable. D'autres cas posent des contraintes sur le pas de déplacement de l'objet en mouvement pour ne pas dépasser un certain nombre de pixels entre deux images consécutives. Alors que dans les méthodes génériques, seule la position cible initiale dans une séquence est supposée connue.

Pour développer un algorithme de suivi robuste, certaines exigences doivent être prises en considération, à savoir la robustesse et traitement en temps réel. Pour la robustesse, l'algorithme doit maintenir un suivi précis sans aucune dérive dans des conditions complexes. Pour les applications temps réel, l'algorithme doit traiter les flux vidéo en ligne dans un temps d'exécution très faible.

Notre objectif est de proposer des approches pour faire face à ces défis, en adoptant le suivi générique où seule la position initiale est prise en considération. Ainsi notre recherche est axée sur l'amélioration des approches proposées dans la littérature à travers l'élaboration d'une méthode de suivi qui fonctionne sous différentes hypothèses, d'un modèle dynamique qui peut couvrir différents type de mouvement et d'un modèle d'apparence robuste pour le suivi d'objet dans des scènes complexes.



Figure 1. Différents problèmes pouvant survenir lors du suivi. (a) Occlusion; (b) variation d'illumination; (c) arrière-plan encombré; (d) Variation d'échelle; (e) Déformation; (f) Mouvement rapide; (g) Rotation dans le plan; (h) Flou de mouvement

2 Contributions

L'objectif de ce travail est de proposer une approche qui permet un suivi visuel stable et robuste, en particulier, dans des situations où le comportement dynamique de l'objet suivi est variable allant d'un mouvement uniforme au mouvement rapide. Notre but est aussi de suivre la cible dans des séquences caractérisées par des changements de luminosité, encombrement de l'arrière-plan ou de flou, de telles conditions peuvent rendre le modèle d'apparence adopté faible. Dans ce travail nous proposons des solutions pour surmonter ces problèmes, en effet un algorithme de suivi est basé essentiellement sur deux modules complémentaires : Module de modélisation et module d'estimation. Les contributions de cette thèse se rapportent à chacun de ces deux modules, elles sont répertoriées dans trois articles et peuvent être récapitulées comme suit :

Estimation basé sur le filtre H_∞ infini et filtre à particules (H_∞ PF) : Sous l'hypothèse que le mouvement d'un objet dans sa forme globale est linéaire tandis que son comportement local est non linéaire. Par conséquent notre stratégie de prédiction repose sur la combinaison séquentielle de deux filtres de prédiction Bayésien :

Prédiction global utilisant le filtre H_∞ qui est une alternative intéressante du célèbre filtre de Kalman, en effet le filtre H_∞ ne fait aucunes hypothèses sur les perturbations et minimise l'erreur d'estimation du pire cas. Son utilisation permet d'atténuer le problème de dégénération des particules surtout dans le cas où la cible a un mouvement complexe ou de manœuvre.

Prédiction locale basé sur le filtre à particules, qui est plus précis et peut effectivement gérer les cas non linéaire et non gaussien, le filtre à particules exploite l'état dynamique globale pour générer un ensemble d'échantillons dont le nombre est considérablement petit, ce qui permet de réduire le temps de calcul sans aucun cas perdre dans la précision de suivi.

Modélisation adaptative du mouvement : Le choix du modèle dynamique pour décrire le mouvement de la cible est primordial pour un suivi stable et robuste, surtout dans le cas où l'objet a un mouvement complexe. L'utilisation d'un seul modèle peut aboutir à un suivi non précis, d'un autre côté intégrer plusieurs sous modèles augmente la complexité de l'algorithme. Nous proposons une approche

adaptative basée sur deux sous modèles paramétrés dont chacun peut couvrir différents mode de mouvements à savoir les mouvements uniformes et complexes. Par la suite les paramètres du système dynamique sont ajustés automatiquement suivant des règles logiques.

Modèle d'apparence basé sur la couleur et la texture : Un nouveau modèle d'apparence basé sur la couleur et la texture qui permet d'exploiter à la fois les caractéristiques statistiques globales et locales de l'apparence de la cible :

L'histogramme pondéré est construit sur les composantes H et S du système HSV, en utilisant un modèle de mélange α -gaussien (α -GMM), le paramètre α rend la modélisation plus flexible et permet à la fonction de probabilité de posséder une distribution plus plane. La zone objet est masquée avec un profil de noyau isotrope pour intégrer l'information spatiale.

L'histogramme distinctif basé sur le motif binaire local uniforme (ULBP) est construit sur la composante V du mode HSV, la zone objet est subdivisée en blocs et chaque bloc est décrit par un histogramme distinctif qui est défini comme la différence entre l'histogramme basé sur l'opérateur ULBP du bloc central et celui du bloc voisin. Par la suite tous les histogrammes distinctifs sont concaténés pour former un seul descripteur de texture.

Les deux descripteurs sont fusionnés dans le cadre du filtre particulaire. Les coefficients de fusion dépendent de la contribution de chaque caractéristique dans la qualité du suivi. Ainsi la mise à jour des poids des caractéristiques est basée sur l'analyse de la distribution spatiale des particules. En tant que paramètre de dispersion, le coefficient de variation est utilisé pour ajuster automatiquement les coefficients de fusion associés aux caractéristiques d'apparence.

3 Structure de la thèse

Ce mémoire de thèse est organisé en quatre chapitres. Le chapitre I présente dans la première partie un tour d'horizon critique des méthodes classiques et contemporaines sur le suivi visuel, trois classes d'approches sont décrites et analysées à savoir les méthodes discriminatives, les méthodes génératives ainsi que les méthodes combinées. Dans la deuxième partie, nous examinons les différentes

caractéristiques utilisées pour la représentation des objets suivis ainsi que les récentes avancées dans les descripteurs de caractéristiques.

Le chapitre II présente le principe et l'essentiel de la théorie de la prédiction Bayésien. Trois filtres de prédiction sont décrits dont chacun est spécifique pour un type de représentation de l'incertitude et de système. Par la suite le principe de notre approche de prédiction basée sur le filtre H_∞ et le filtre particulière est présenté. Dans la deuxième partie les simulations et les résultats expérimentaux sont exposés et discutés.

Le chapitre III introduit les détails de notre méthode de modélisation de l'apparence de la cible basée sur la fusion des caractéristiques de couleur et de texture permettant d'exploiter à la fois les caractéristiques statistiques globales et locales de l'apparence de la cible. Nous présentons la modélisation de la distribution des caractéristiques de couleur effectuée à travers un modèle de mélange de α -gaussiennes, ainsi de la modélisation des caractéristiques texture utilisant l'opérateur motif binaire locale uniforme. Finalement, nous décrivons la dernière étape de modélisation qui explique une fusion des descripteurs et une mise à jour du modèle de base.

Le chapitre IV présente l'approche proposée pour le suivi visuel qui comprend principalement deux étapes. La première étape concerne l'estimation de l'état globale de la cible par le filtre de prédiction H_∞ adaptatif appliqué à un système dynamique paramétré, dont les paramètres sont ajustés de manière automatique. La deuxième étape, concerne l'estimation de l'état local de la cible et s'appuie sur le modèle d'apparence décrit dans le troisième chapitre. La dernière partie est réservée à l'évaluation des performances de l'approche proposée ou une étude comparative expérimentale est conduite sur une série de scènes vidéo.

Finalement le rapport est terminé par une conclusion résumant l'ensemble des contributions, des perspectives pour des travaux futurs sont présentées.

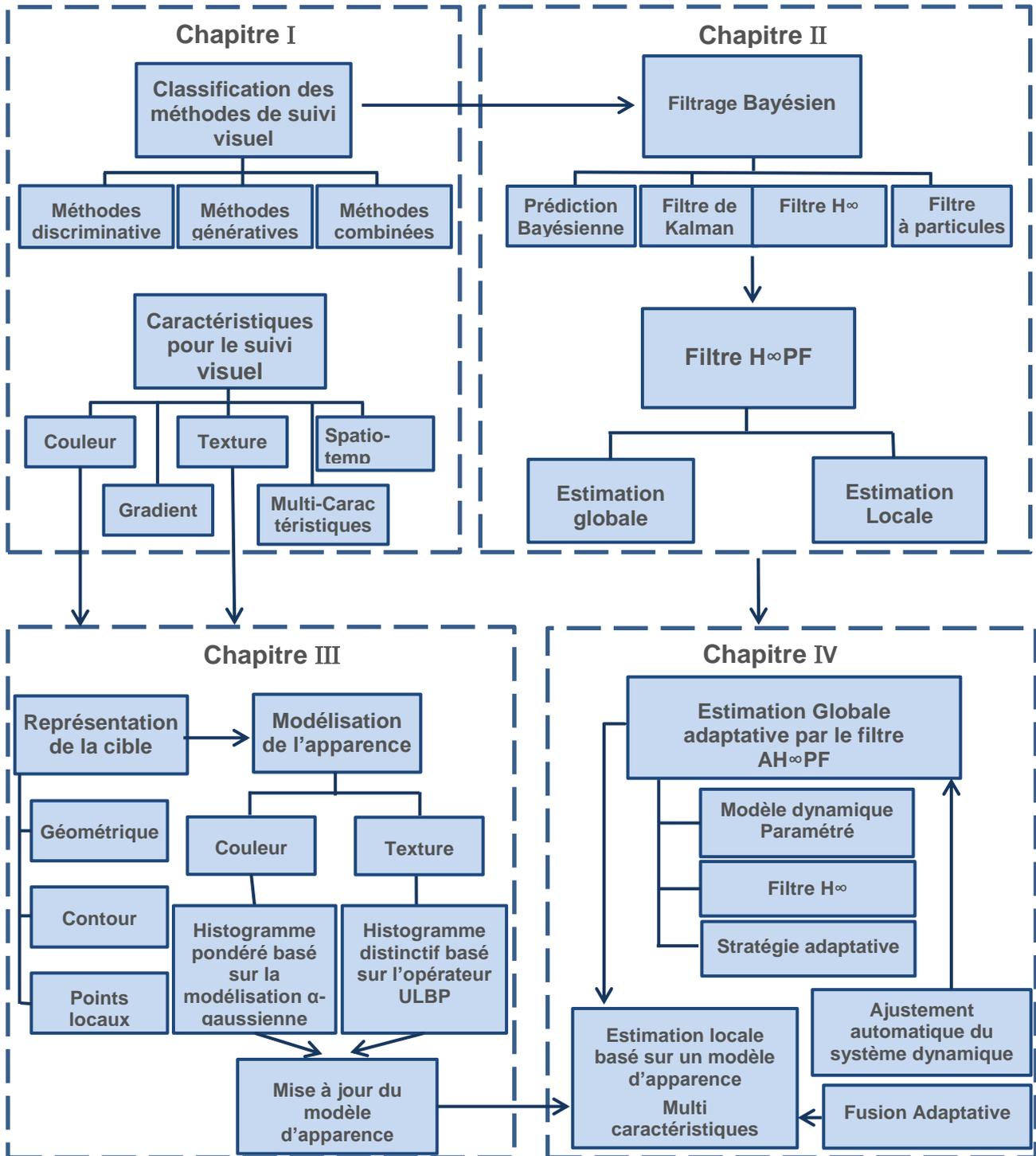


Figure 2. Représentation schématique de la structure de la thèse

Chapitre I : Etat de l'art

L'objectif de ce chapitre est de présenter les principales caractéristiques ainsi que les tendances des descripteurs utilisés pour la représentation des objets en mouvement. Les approches classiques ainsi que les progrès les plus récents en matière de méthodes de suivi visuel sont passées en revue, un classement en deux catégories principales est présenté : discriminative et génératives ainsi la théorie de base et les principes des différentes approches sont examinés à savoir les méthodes basées sur l'apprentissage en ligne, les méthodes basées sur l'analyse sous espace et les méthodes basées sur la modélisation de l'apparence et la modélisation parcimonieuse.

CHAPITRE I : ETAT DE L'ART	22
I.1. GENERALITE SUR LE SUIVI VISUEL	23
I.2. CARACTERISTIQUES POUR LE SUIVI VISUEL	26
I.2.1. CARACTERISTIQUES DE COULEUR	26
I.2.2. GRADIENT	27
I.2.3. TEXTURE	29
I.2.4. SPATIO-TEMPORELS	31
I.2.5. LA FUSION DE PLUSIEURS CARACTERISTIQUES.....	32
I.3. CLASSIFICATION DES METHODES DE SUIVI VISUEL	33
I.3.1. LES METHODES DISCRIMINATIVES.....	33
1.3.1.1. Méthode basé sur la représentation de de l'arrière-plan.....	34
1.3.1.2. Méthodes discriminatives par apprentissage en ligne	35
I.3.2. LES METHODES DE SUIVI GENERATIVES	40
1.3.2.1. Méthodes basées sur des modèles	40
1.3.2.2. Méthodes basées sur l'analyse par sous-espace	44
1.3.2.3. Méthodes basées sur la représentation parcimonieuse	47
I.3.3. METHODES COMBINEES	52
I.4. CONCLUSION	55

Le développement du suivi visuel est étroitement lié aux avancées technologiques dans le domaine de la vision par ordinateur. Avant les années 1980, la recherche et l'analyse des séquences d'images numériques étaient similaires à la technique appliquée dans la recherche d'images statiques. La recherche dans les séquences d'images dynamiques est devenue un point fort après la méthode du flux optique. Les années 1980 ont connu l'élaboration de méthodes de suivi visuel, dont certaines sont devenues actuellement des méthodes classiques.

Ce chapitre est organisé en deux parties. Dans la première partie, la structure et les applications du suivi visuel sont présentées, ainsi que les récents progrès les plus représentatifs, en rapport avec les descripteurs de caractéristiques qui s'appuient sur les dernières études dans le domaine de la détection d'objets visuels. Dans la deuxième partie, nous décrirons les approches classiques et contemporaines pour le suivi visuel des objets selon une classification trois catégories : méthodes discriminatives, méthode génératives et méthodes combinées.

1.1. Généralité sur le suivi visuel

Le suivi visuel peut être défini comme un problème de localisation de la cible tout au long d'une séquence vidéo. Différentes représentations pour les objets en mouvement sont considérées, en particulier les formes géométriques, contour ou points locaux. Différents paramètres sont associés à la zone de représentation de l'état de l'objet en mouvement, à savoir les paramètres en rapport avec la position, la vitesse et l'accélération, les paramètres d'échelle et d'orientation, ainsi que les paramètres d'apparence. Ces paramètres sont traités, et analysés en vue de localiser l'objet suivi.

En général, la procédure de suivi visuel est composée de cinq étapes : Initialisation, modélisation, Estimation, Mise à jour et localisation (Figure 1). Dans l'étape d'initialisation, l'objet est détecté soit manuellement ou automatiquement, ensuite l'état et le modèle de base sont construits. L'étape de modélisation consiste d'une part à modéliser l'apparence de l'objet suivi à base de différents types de caractéristiques visuelles, et d'autre part à formuler l'état dynamique généralement sous forme d'un système d'état. Dans l'étape d'estimation, les paramètres du système sont estimés par des filtres de prédiction ou des méthodes de régression linéaire. L'étape de mise à jour permet d'adapter le modèle d'apparence aux

différentes conditions de suivi. La dernière étape concerne la localisation de l'objet dans l'image suivante.

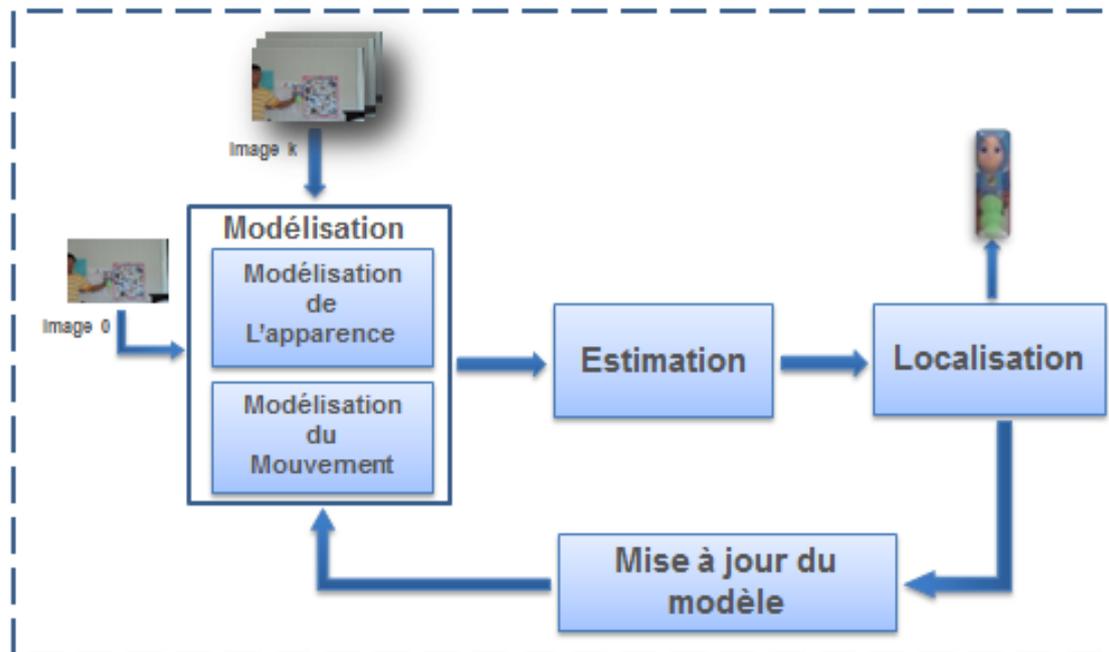


Figure 3. Organigramme d'un algorithme de suivi visuel

Actuellement, l'application du suivi visuel recouvre plusieurs domaines (Figure 4), tels que la vidéo surveillance, la compression vidéo, la robotique industrielle, système de trafic intelligent, reconnaissance d'activité, etc.

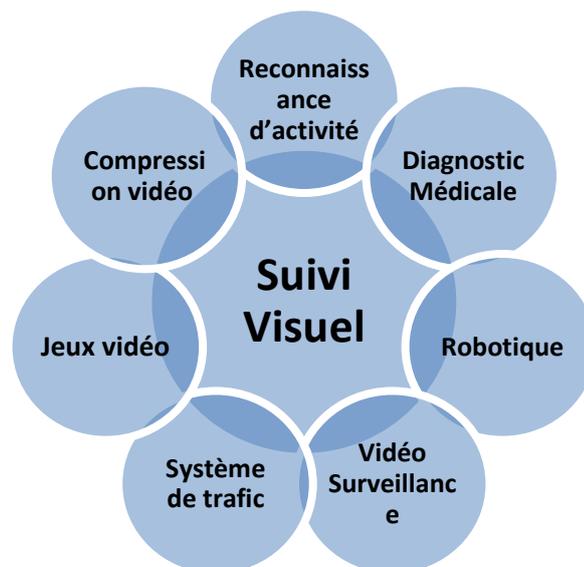


Figure 4. Quelques applications du suivi visuel

Vidéo Surveillance intelligente (SVI) : le système de surveillance automatique supervise et contrôle le comportement de la cible en temps réel dans une zone spécifique, tels que les services de sécurité, les résidences privées et les zones publiques pour des fins de d'enquête ou de prévention à travers la reconnaissance des comportements suspects. Les principales fonctions sont la détection, la reconnaissance et le suivi des visages dans la vidéo.

Compression vidéo : l'objectif est la réduction du coût de stockage et de transmission. L'utilisation du suivi permet de détecter la redondance temporelle dans la séquence d'images et ainsi de décrire une image par des images temporellement voisines.

Système de trafic intelligent (STI) : le suivi visuel permet de calculer le débit, l'état et le comportement des véhicules en circulation. Il est aussi utile pour repérer et compter les piétons en temps réel et pour calculer la densité des piétons et peut ainsi aider à la gestion du trafic en toute sécurité sur le carrefour.

Robotique industrielle: Le suivi visuel est utilisé dans le système de contrôle de la robotique industrielle et humanoïde avec algorithme de suivi dans la boucle de production.

Jeux vidéo : Le suivi visuel est utilisé pour offrir un meilleur contrôle aux utilisateurs, un suivi des mouvements d'utilisateurs et des visages pour améliorer l'interactivité.

Reconnaissance d'activité : Concerne la reconnaissance d'activité dans la vidéo surveillance (aéroport, patients et personnes âgées) elle vise à étiqueter des actions comme marcher, sauter, s'asseoir, courir, tomber etc.

Le suivi visuel peut être appliqué dans de nombreux autres domaines, en tant que technique principale dans l'aéronautique pour le contrôle de vol des avions sans pilote dans la médecine moderne où l'imagerie médicale constitue une aide au diagnostic et au suivi thérapeutique. En outre, il peut être appliqué pour suivre la trajectoire des protéines dans les cellules et analyser les caractéristiques dynamiques de la structure cellulaire.

1.2. Caractéristiques pour le suivi visuel

La majorité des méthodes de suivi [Arn, 2014] se basent sur des caractéristiques d'apparence pour représenter la cible. Le changement de l'apparence de l'objet tout au long des séquences vidéo rend la tâche du suivi difficile car l'apparence de l'objet peut varier en raison de nombreux facteurs, notamment le point de vue, les occlusions, la luminosité, les objets articulés. Le choix des caractéristiques discriminantes joue un rôle crucial pour le suivi visuel. La propriété la plus souhaitable d'une caractéristique visuelle est son caractère unique, de sorte que les objets peuvent être facilement distingués.

Au cours de la dernière décennie, la détection d'objets d'une classe particulière, tels que les humains ou les voitures, a suscité un intérêt croissant dans le domaine de la vision par ordinateur. Cela a motivé l'élaboration de différentes caractéristiques d'image qui capturent différentes propriétés. Certaines méthodes existantes se basent sur un seul type de caractéristiques. Cela permet une comparaison directe des performances de détection. D'autres essaient d'intégrer plusieurs types de fonctionnalités pour améliorer les performances. En effet, tout descripteur de caractéristique utilisé pour la détection visuelle peut être adapté pour un suivi visuel. Nous présentons, dans la suite, des récentes avancées représentatives dans des descripteurs de caractéristiques, à partir d'innovations dans le domaine du suivi visuel d'objets.

1.2.1. Caractéristiques de couleur

La couleur apparente d'un objet est principalement influencée par deux facteurs physiques, la distribution de puissance spectrale de l'illuminant et les propriétés de réflectivité de la surface de l'objet. La caractéristique couleur a été largement utilisée pour la représentation des caractéristiques des objets dont les descripteurs sont généralement basés sur l'histogramme. Dans [Ire, 2017] les auteurs ont proposé un histogramme pondéré au lieu d'utiliser un histogramme de couleur usuel afin de rendre la distribution plus fiable et de réduire l'effet de l'arrière-plan. De plus, afin de réduire l'effet de variation de l'éclairage, l'espace HSV est utilisé au lieu de l'espace RGB. Afin de surmonter le changement de l'apparence des objets au fil du temps en raison du changement de l'éclairage, de la déformation des objets et de l'angle de vue, le modèle cible est adapté au fil du temps.

Dans [Jun, 2018] les auteurs ont proposé une méthode de super pixel multi-échelle pour segmenter un objet en utilisant un masque de confiance global qui détermine automatiquement le centre et la taille des structures de l'objet. Le filtre de corrélation à noyau (KCF) a été intégré dans le cadre du filtrage bayésien, dans lequel la carte de confiance guidée par la couleur proposée est considérée comme un modèle d'observation.

Dans [Min, 2017] un espace colorimétrique basé sur la préservation des teintes est conçu pour remplacer l'espace RGB et pour stabiliser la précision du suivi en cas de variation importante de la luminosité. L'espace RGB est converti en l'espace HSV et la composante de teinte reste constante par le biais d'un principe de préservation de la teinte.

Dans [Yaw, 2017] les auteurs ont présenté des champs de distribution de couleurs multicanaux pour la modélisation de l'apparence, qui permet la distinction des couleurs entre la cible et l'arrière-plan. Une structure hiérarchique adaptative de ces champs de distribution de couleurs a été adoptée. Cette structure hiérarchique est générée via un l'algorithme de k-means, ainsi que le poids de chaque canal est ajusté en ligne en fonction de sa discrimination.

1.2.2. Gradient

La caractéristique du Gradient est largement appliquée dans les applications de détection et de suivi d'objets. D'une manière générale, les opérateurs de gradient de bord permettent la détection des contours. L'opérateur Prewitt est plus sensible aux bords horizontaux et verticaux, tandis que l'opérateur Sobel est plus sensible aux bords diagonaux. Essentiellement, Prewitt et Sobel sont complémentaires. Les entités de contour sont extraites de la cible pour construire le modèle d'apparence initiale, les parties claires fournissent souvent des informations plus utiles pour décrire les caractéristiques de la cible.

Les auteurs de [Kan, 2018] ont adopté des caractéristiques de contour pour représenter l'apparence de la cible. En effet l'apparence initiale d'une cible a toujours un contour clair, qui est invariant par la lumière et robuste au changement d'illumination. Les auteurs ont proposé une stratégie en ligne pour

mettre à jour de manière incrémentale le modèle afin de s'adapter au changement d'apparence.

Dans [Yua, 2012], les auteurs ont proposé une sélection de caractéristiques à base de gradient associé à l'apprentissage multi-instance (MIL) en ligne. Ils ont introduit un nouveau modèle d'apparence d'objet composé de caractéristiques histogramme des gradients orientés (HOG) et l'ensemble de caractéristiques correspondantes à partir d'échantillons positifs et négatifs, et appliquer un cadre de simulation pour combiner les composants HOG dans un classifieur fort discriminant. Un schéma d'optimisation est utilisé pour mettre à jour le modèle d'apparence. Il met à jour de manière itérative chaque caractéristique en utilisant la descente du gradient et les approches MIL en maximisant la probabilité de formation des caractéristiques HOG.

Les auteurs dans [Mee, 2017] ont présenté une nouvelle méthode de fusion de plusieurs algorithmes de suivi pour améliorer les performances du suivi humain dans une application robotique. La synthèse statistique des gradients pour la caractéristique accélérée (SURF) est utilisée comme unique caractéristique visuelle pour représenter et suivre l'objet. Afin de supprimer les descripteurs de SURF de l'arrière-plan du modèle, une ellipse dans la zone rectangulaire est considérée. Les descripteurs SURF présents dans cette région elliptique sont calculés. Pour supprimer le reste des descripteurs appartenant à l'arrière-plan, un algorithme de croissance de région (flood-fill) est appliqué aux points-d'intérêt situés en dehors de la région elliptique. De cette manière, la région d'arrière-plan est segmentée à partir du premier plan. Tous les descripteurs, qui appartiennent à cette région d'arrière-plan segmentée, sont maintenant supprimés de la description du modèle. Cette étape est répétée chaque fois qu'un nouveau modèle est sélectionné. C'est une étape cruciale qui permet d'éviter la déviation du suiveur pendant une durée plus longue.

Dans [Din, 2018] les auteurs ont une approche pour un système de suivi qui utilise les caractéristiques de Haar pour la structure de l'objet et les fonctions HOG pour le contour. Un ensemble de caractéristiques mixtes ont été développé avec ces deux caractéristiques. Le Boosting de la sélection en ligne de fonction est utilisée pour sélectionner des fonctions importantes et mettre à jour ces fonctions en ligne pour comprendre le choix optimal en utilisant un classificateur support vecteur machine (SVM) en cascade. La nécessité d'une extraction en cascade en utilisant

des caractéristiques de type Haar consiste essentiellement à diminuer l'utilisation de valeurs de pixels brutes, et à faciliter la classification

1.2.3. Texture

La caractéristique texture est une mesure de la variation d'intensité d'une surface qui quantifie des propriétés telles que la douceur et la régularité. L'ondelette de Gabor [Sho, 2008] est probablement la caractéristique de texture la plus utilisée. Les filtres de Gabor peuvent être considérés comme des détecteurs de bord et de ligne accordables à l'orientation et à l'échelle, et les statistiques de ces micro-caractéristiques dans une région donnée sont souvent utilisées pour caractériser les informations de texture sous-jacentes.

Au cours des dernières années, un intérêt croissant a été porté à l'étude des schémas locaux de l'image pour une meilleure détection et reconnaissance. En particulier, les modèles locaux binarisés avec un seuil adaptatif fournissent des bons résultats sur divers sujets, tels que la détection des visages et la classification des images. Dans [Oja, 2002], les auteurs ont développé un descripteur de texture, appelé motifs binaires locaux (LBP). L'opérateur d'analyse de texture LBP est défini comme une mesure de texture d'échelle de gris dérivée d'une définition générale de texture dans un voisinage local. La principale caractéristique de l'opérateur LBP est sa tolérance aux changements d'éclairage. Une autre caractéristique tout aussi importante est sa simplicité de mise en oeuvre.

De nombreuses variantes de LBP ont été récemment proposées. Dans [Hwa, 2014] les auteurs ont proposé un nouveau modèle, appelé modèle structural binaire (SBP), pour le suivi des objets où un schéma de seuillage alternatif est introduit pour générer un ensemble de SBP multiples. Le SBP code non seulement le modèle binaire constitué de différences binarisées entre les intensités moyennes des sous-régions à l'intérieur de la région cible, mais aussi la configuration spatiale des sous-régions. Les auteurs ont défini une métrique pour la similarité entre les modèles SBP de la cible et du candidat pour la localisation de la cible, qui est basée sur une distance de Hamming pondérée par le noyau isotrope. Dans [Yon, 2017], un nouvel algorithme de suivi basé sur un descripteur de modèle binaire local de niveau de gris (GLLBP) est proposé. Le descripteur décrit non seulement la caractéristique de texture locale mais également la distribution de la luminance. L'algorithme

augmente la robustesse du système de suivi dans une scène à basse résolution et dans un arrière-plan complexe. Dans des circonstances extrêmes, il n'y a pas d'informations de texture, LBP devient sans signification, mais GLLBP peut toujours fonctionner comme un histogramme de la valeur de l'échelle de gris.

Afin d'améliorer la précision, gérer la disparition de la cible et surtout diminuer le coût de calcul pour des applications en temps réel, de nombreux algorithmes de suivi visuel ont été proposés. Les auteurs dans [Zeb, 2016] ont proposé une nouvelle méthode de suivi qui utilise des caractéristiques binaires locales à blocs multi-échelle modifiés (MB-LBP). Une telle caractéristique a une faible charge de calcul et une invariance vis-à-vis du changement de l'éclairage. En effet l'opérateur LBP de base ne code que les microstructures et peut facilement être affecté par le bruit tandis que MB-LBP code à la fois les microstructures et les macrostructures de l'image. Le modèle d'apparence est mis à jour dans chaque image en remplaçant le modèle d'apparence par les entités extraites du correctif d'image le plus récent de la cible. La présence d'arrière-plan dynamique et les effets d'ombrage diminuent la précision du suivi.

Dans [Moh, 2017] les auteurs ont proposé une nouvelle technique de normalisation d'arrière-plan avec analyse de modèle de texture pour supprimer la région d'ombre. L'algorithme de prédiction de chaîne de voisinage (NCP) est utilisé pour regrouper l'arrière-plan inégal et le modèle de limite différentielle (DBP) extrait la texture de l'image vidéo pour supprimer les pixels d'ombre présents dans l'image. L'estimation d'intensité plus faible et la prédiction de la zone autour de l'intensité plus faible permettent une élimination de l'ombre. Le cadre sans ombre se divise en plusieurs grilles et les histogrammes des entités sont extraits du cadre formaté en grille.

Les auteurs dans [Moh, 2018] introduisent une nouvelle technique de suivi en utilisant le motif binaire adaptatif médian (AMBP) comme caractéristique de texture pour les algorithmes de suivi. Le descripteur AMBP combine à la fois le modèle local binaire (LBP) et le modèle binaire médian (MBP) en fonction du bruit et des informations de microstructure. L'AMBP utilise une procédure de filtrage adaptatif médian qui le rend non sensible au bruit.

Dans [Hua, 2016], les auteurs ont proposé un nouveau descripteur de texture local, appelé motif binaire quadrilatère (QBP). Comparé au modèle binaire local (LBP), le

QBP est plus adapté pour l'extraction de caractéristiques sous une scène complexe (changement de luminance, de la couleur de cible et arrière-plan similaire). Le QBP est calculé sur la base d'un bloc 2×2 , en utilisant la moyenne de ces quatre pixels comme seuil, y compris le pixel courant, son pixel gauche, son pixel supérieur et son pixel supérieur gauche. Le QBP contient 16 motifs, qui peuvent être synthétisés en 6 classes selon le critère d'invariance de rotation. Comparé au LBP, le QBP a une meilleure capacité discriminante comparée à LBP, puisque deux pixels avec des valeurs de niveaux de gris assez différentes ont des codes QBP différents tout en ayant les mêmes codes LBP

1.2.4. Spatio-temporels

Les caractéristiques spatio-temporelles locales sont récemment devenues une représentation populaire pour la reconnaissance d'action et la détection visuelle. Les fonctions spatio-temporelles locales capturent des motifs et les caractéristiques de mouvement et fournissent une représentation relativement indépendante des événements par rapport à leurs décalages spatio-temporels, ainsi que l'encombrement de fond et les mouvements multiples dans la scène.

Plusieurs méthodes de localisation et de description des caractéristiques ont été proposées dans la littérature. Les auteurs dans [She, 2017] considèrent une image dans sa forme courante comme une matrice, et construisent la représentation d'objet basée sur le tenseur du 3^{ème} ordre pour préserver la corrélation spatiale dans l'image 2D et exploiter pleinement l'information temporelle utile. Une mise à jour du modèle d'objet est effectuée en utilisant la décomposition de la valeur singulière N-mode pour modéliser les variations d'apparence, ce qui réduit l'influence de la dérive du modèle et des occlusions d'objets. Le schéma proposé apprend efficacement une représentation tensorielle de faible dimension grâce à la mise à jour adaptative de la base propre du tenseur.

Dans [Yan, 2015] Les auteurs ont développé un moyen de mettre à jour un modèle d'apparence adaptatif. Le modèle se compose de plusieurs modules appelés indices temporels, qui sont mis à jour d'une manière alternative, ce qui permet de conserver les informations antérieures et courantes de l'objet suivi, par conséquent de gérer les changements d'apparence brusque. Chaque module est représenté par plusieurs fragments appelés indices spatiaux. Afin d'incorporer tous les indices

spatiaux et temporels, les auteurs ont développé un critère d'évaluation efficace de la qualité des repères, qui combine les informations d'apparence et de mouvement.

Dans [Yun, 2016] les auteurs ont présenté une méthode adaptative de modélisation de fond. La méthode de détection d'espace mesure d'abord la distance entre un objet en mouvement et un objet immobile. La première méthode de détection temporelle conduit une analyse de changement temporel puis une analyse de changement spatial.

Les auteurs dans [Yong, 2017] ont combiné les avantages des algorithmes de flot optique L-K et du contexte spatio-temporel (STC). Dans la première étape, ils ont modélisé la relation spatiale entre l'objet et sa région dans le contexte local, à travers le cadre bayésien. La corrélation statistique entre l'objet et sa région environnante est ainsi obtenue. Ensuite, cette relation spatiale est fusionnée avec le concours focalisé sur le système biométrique afin d'évaluer la carte de confiance de l'emplacement de l'objet dans la nouvelle image. Par conséquent, l'emplacement de l'objet dans la nouvelle image est obtenu. Dans un second temps, le suivi par flot optique de Lucas-Kanade a été implémenté en adoptant les informations de contexte temporel.

1.2.5. La fusion de plusieurs caractéristiques

Depuis l'émergence de divers descripteurs de caractéristiques, la fusion de caractéristiques est devenue de plus en plus importante pour la restauration d'images et de vidéos, le suivi visuel et la détection. Le schéma de fusion de caractéristiques conduit généralement à des performances ou une robustesse améliorée, ce qui attire largement l'attention des chercheurs de la vision par ordinateur. Dans ce contexte, Jin et al. [Jin, 2018], ont proposé une approche intégrant deux caractéristiques la couleur dans l'espace HSV et la texture. L'information d'espace colorimétrique est prise en compte par la méthode des blocs, mais vu la nécessité d'un processus en temps réel, la zone cible est divisée en neuf parties. Afin de mettre en évidence la contribution de chaque bloc, des poids leurs sont affectés. Pour chaque bloc, un histogramme conjoint de la couleur et de la texture est établi. Le nombre de bin de l'histogramme est de 26×5 . Dans [Saa, 2018], les auteurs ont proposé un algorithme de suivi utilisant l'histogramme de couleur et de texture. Il combine l'opérateur de quantification de phase locale

(LPQ) de la représentation de caractéristiques de texture et l'histogramme de couleur. Les deux descripteurs sont fusionnés dans l'algorithme de Mean Shift.

Dans [Pra, 2017], les auteurs ont proposé un algorithme de suivi d'objet utilisant un codage crypté régularisé (RRSC) pour représenter les modèles multi-caractéristiques des objets candidats. La fonction LBP et la fonction de couleur Ohta, réunies dans la modélisation de l'apparence, améliorent la représentation discriminante.

Les auteurs dans [Yon, 2016] ont proposé un algorithme de suivi basé sur la description d'invariant de rotation mixte (MRID). Le MRID adopte un histogramme LBP annulaire mixte et un histogramme de gradient annulaire avec une direction dominante pour décrire la texture et les informations de bord d'un objet. Dans l'histogramme de gradient annulaire, les gradients horizontaux et verticaux de chaque pixel sont d'abord calculés. Ensuite, l'amplitude et l'orientation du gradient sont calculées. L'histogramme d'orientation est obtenu avec une plage de 45° dans chaque bin d'orientation. Après, le pic le plus élevé est détecté en tant que direction dominante de l'objet. Le MRID surmonte les grands problèmes de calcul relatif aux histogrammes binaires locaux conventionnels et le gradient orienté (LBP-HOG).

Dans [Zhi, 2015] les auteurs ont utilisé un descripteur conjoint multifonctions (MFJD), et la distance usuelle entre les histogrammes en tant que mesure de la similarité entre une cible et ses candidats. Les caractéristiques couleurs et HOG sont fusionnés. Une stratégie de fusion multi-caractéristique auto-adaptative ajuste le poids des éléments conjoints fusionnés en fonction de leurs scores de stabilité.

1.3. Classification des méthodes de suivi visuel

On peut distinguer principalement deux catégories des approches pour le suivi visuel : les méthodes discriminatives et les méthodes génératives. Dans la première catégorie, le problème de suivi est formulé comme un problème de classification binaire. Quant à la seconde catégorie, l'objet est représenté par un modèle d'apparence qui peut être évidemment masqué par un noyau. Le processus de suivi recherche le candidat dont le modèle d'apparence observé est le plus proche de celui du modèle de base.

1.3.1. Les méthodes discriminatives

Les méthodes discriminatives, considèrent le suivi des objets comme un problème de classification. Elles tiennent compte simultanément des informations de la cible et de l'arrière-plan. Un classifieur binaire permet d'extraire la cible de l'arrière-plan, une mise à jour en ligne est effectuée pour gérer les changements d'apparence et d'environnement. Dans ce contexte plusieurs algorithmes discriminatives ont été élaborés.

1.3.1.1. Méthode basé sur la représentation de de l'arrière-plan

Les auteurs dans [Elg 2002] utilisent les techniques d'estimation à noyau pour créer une représentation à la fois pour l'arrière-plan et le premier plan. Ils proposent une technique de modélisation de l'arrière-plan adaptative et une technique de soustraction de fond capable de détecter des cibles mobiles dans des environnements extérieurs difficiles comme le changement de luminosité, les objets non statiques. Dans [Kum, 2016], les auteurs proposent un algorithme efficace de soustraction de fond pour le suivi des objets dans des conditions de fond statiques et dynamiques. Le travail est axé sur la réalisation des blocs mobiles pertinents dans le premier plan, aidant à l'initialisation et à la mise à jour appropriée du module d'arrière-plan afin d'améliorer la précision du suivi. L'entropie des blocs est évaluée au-dessus d'une certaine distance des pixels de l'image de soustraction, afin d'extraire les pixels en mouvement pertinents du champ de mouvement initial.

Les auteurs dans [Hua, 2017], proposent une nouvelle approche qui permet de détecter les objets du premier plan avec des mouvements fréquents et / ou peu fréquents. Plus précisément, ils utilisent un mécanisme d'assignement visuel pour inférer l'arrière-plan complet à partir d'un sous-ensemble d'images, et ensuite le propager vers les autres images pour une soustraction de fond précise. En outre, un algorithme local de stabilisation de mouvement est proposé. Il est basé sur les caractéristiques pour identifier les mouvements locaux fréquents en arrière-plan pour réduire les faux positifs dans le premier plan détecté.

Le travail réalisé dans [Yad, 2016], présente une méthode de suivi des objets dans des vidéos acquises par des caméras thermiques basé sur la distance de Kullback-Leibler, et la technique de soustraction de l'arrière-plan, pour minimiser le bruit et extraire les objets dynamiques. Les auteurs proposent un modèle basé sur la

moyenne ajustée. La distance de Kullback-Leibler est calculée entre le pixel d'arrière-plan, et un pixel courant à l'aide du modèle de mélange gaussien.

Pour résoudre le problème des occlusions, [Bin, 2014] présente un algorithme de suivi visuel via un modèle d'apparence adaptative basé sur une estimation d'arrière-plan local. Un objet est représenté par un modèle d'apparence basé sur les parcelles, dans lequel chaque parcelle produit une carte de confiance pendant le processus de suivi, lesquelles sont combinées via un estimateur. De plus, une approche de modélisation d'arrière-plan basée sur la cooccurrence spatiale locale est proposée pour estimer automatiquement le modèle local de fond d'un objet dans une image acquise à partir d'une seule caméra, qui peut être fixe ou mobile.

Pour faire face au problème de la présence des ombres qui entraînent des confusions, les auteurs dans [Hak, 2012] proposent une approche de la détection et de l'extraction automatiques des ombres, qui exploite plusieurs propriétés dérivées de l'analyse spectrale, géométrique et temporelle. Un modèle générique sélectionne des zone d'ombre candidates en se basant sur les caractéristiques temporelles d'orientation de l'ombre suivi d'une vérification de la détection par l'analyse photométrique de chaque pixels. Dans un autre travail [Sab, 2018], les auteurs proposent une approche de détection automatique des ombres basée sur la sur la décomposition empirique bidimensionnelle. L'image ombrée est décomposée en composantes intrinsèques contenant uniquement la texture et un résidu de la forme de l'objet. Ensuite, la détection des ombres est effectuée sur les composantes des textures par correspondance des régions, tandis que l'élimination de l'ombre est effectuée via une approximation gaussienne appliquée uniquement sur le résidu. Enfin, l'image sans ombres est obtenue en ajoutant toutes les composantes de texture et le résidu sans ombres.

1.3.1.2. Méthodes discriminatives par apprentissage en ligne

Pour gérer efficacement les variations de structure non rigide, la déformation de forme, les changements de forme et de mouvement brusque et de l'environnement, de nouvelles méthodes basées sur l'apprentissage en ligne ont été développées. Elles sont basées sur la sélection de caractéristiques, sur des classifieurs ou sur l'apprentissage métrique.

Apprentissage par la sélection de caractéristiques

Dans ce type de méthodes, un mécanisme de classification de caractéristiques permet de sélectionner de façon adaptative les caractéristiques discriminantes les mieux classées. La capacité discriminante d'un système de suivi est directement liée à l'espace caractéristique utilisé. En effet une sélection en ligne adaptative d'un espace de caractéristiques approprié est plus adéquate que d'utiliser un ensemble fixe de caractéristiques déterminé a priori. Dans [Han, 2016], les auteurs proposent une méthode de suivi basée sur une représentation des caractéristiques à apprentissage en ligne, via un seul réseau de neurones convolutif. La méthode génère des scores pour tous les candidats possibles dans une image donnée, puis le candidat ayant le score le plus élevé est sélectionnée comme prédiction de l'état de mouvement dans l'image courante. En outre, l'apprentissage est effectué sans avoir besoin d'informations supplémentaires ou d'information hors ligne.

Pour faire face au changement d'apparence causé par un mouvement brusque, la variation de l'éclairage, la déformation et l'occlusion, les auteurs dans [Qin, 2017] proposent une approche de suivi, qui sélectionne les caractéristiques les plus discriminantes pour concevoir un modèle d'apparence efficace. Les caractéristiques les plus discriminantes sont sélectionnées via une méthode de stimulation vectorielle en ligne, et la représentation de l'objet est mise à jour en ligne. Cela qui préserve les fonctionnalités stables tout en filtrant les bruits parasites. Par la suite une approche de rectification de trajectoire est adoptée pour une estimation plus précise. Enfin, pour réduire les interférences provenant des informations de l'arrière-plan, un mécanisme d'adaptation à plusieurs échelles est adopté. La méthode réalise un suivi stable et précis pour la majorité des séquences choisies pour les expérimentations. Toutefois le choix des modèles et des collections influence sensiblement les performances de l'algorithme. La séparation des objets cibles de leur arrière-plan est susceptible d'inclure des échantillons bruités, provoquant ainsi une perte de suivi.

Les auteurs dans [Kai, 2013] proposent un algorithme de sélection de caractéristiques discriminantes en ligne qui optimise une fonction objective dans la direction ascendante par rapport aux échantillons positifs et dans la direction de descente par rapport à échantillons négatifs. L'algorithme proposé réalise un suivi précis, qui mène à un taux de réussite favorable et une erreur de suivi minimale

pour la plupart des séquences vidéo utilisées dans les expérimentations. Mais dans le cas des séquences dont l'objet effectue un mouvement rapide et brusque la méthode proposée conduit à des performances plus au moins faible.

Apprentissage par mise à jour de classifieurs

Dans le cas où aucune information préalable sur l'apparence de l'objet n'est disponible, un apprentissage en ligne à l'aide d'un classifieur binaire est réalisé en raison du manque d'échantillons d'information. A partir d'une position initiale de la cible dans l'image actuelle, le classifieur évalue plusieurs positions possibles dans une région de voisinage dans l'image suivante.

Les auteurs dans [Jun, 2016], proposent une représentation visuelle dans laquelle les principales informations sur la structure spatiale locale et globale de la cible sont incorporées de manière flexible dans l'ensemble de classificateur candidat avec des membres provenant de plusieurs familles de complexité. Pour éviter d'ajuster souvent l'algorithme de suivi une méthode d'apprentissage profond en ligne est adoptée. En outre, les auteurs proposent un système de suivi multi-période pour améliorer la capacité en cas des échecs. Les résultats des expérimentations et des comparaisons réalisés sur une variétés de séquences présentant des conditions difficiles de suivi, montrent la forte capacité de cette méthode à éviter les dérives même lorsque l'apparence de l'objet change de manière significative, mais elle a du mal à se remettre des échecs de suivi causés par des situations telles que les occlusions ou la disparition de cible.

Les auteurs dans [Fen, 2018] s'intéressent aux problèmes en rapport avec les occlusions et l'irrégularité de la forme de l'objet, alors dans de telles situations, certaines parties de la cible ne sont pas contenues dans l'objet, par conséquent ne doivent pas être considérées dans le processus d'apprentissage. Pour surmonter ce problème, les auteurs proposent une approche de suivi en ligne par une mise à jour qui considère des composantes fiables de l'objet. Les modèles d'arrière-plan et de l'objet sont à la fois conservés pendant le suivi. En effet le résultat du suivi courant est comparé avec ces deux modèles, de manière à déterminer les pixels appartenant à l'objet. De plus, afin de mieux représenter les images, les caractéristiques de l'histogramme du gradient orienté sont utilisées pour une méthode d'apprentissage faible du classifieur en ligne. Les résultats des expérimentations menées sur des

séquences complexes ont démontré que la méthode proposée est performante, cependant lorsque tous les pixels sont similaires au modèle de l'arrière-plan, ou dans le cas où l'objet est complètement occlus le suivi échoue.

Le suivi d'objets arbitraires dans des séquences publiques est difficile en raison des différentes difficultés qui peuvent provenir de variances extrinsèques, telles que l'illumination, encombrement de fond, l'occlusion et etc., ainsi que les déformations intrinsèques. Dans ce cadre les auteurs dans [Pen, 2016], proposent un nouvel algorithme bayésien basé sur un ensemble de classifieurs en ligne. Le suivi est basé sur une stratégie d'apprentissage en ligne de multiples algorithmes de suivi, à travers une modélisation d'un processus de minimisation des pertes cumulatives avec un schéma d'échantillonnage par paire pour le classifieur SVM. Les expériences sur une base de données standard, montrent que la méthode proposée réalise un suivi précis. En outre, elle est capable de gérer les défis dans de nombreux environnements difficiles. Toutefois dans le cas des changements imprévisibles de l'apparence de la cible, il est difficile d'obtenir de bonnes performances.

Apprentissage par sélection de métrique en ligne

Pour la plupart des modèles discriminants, des métriques à distance fixe spécifiées à l'avance sont employées, tels que la métrique euclidienne, le coefficient de Bhattacharyya, la divergence de Kullback-Leibler, la métrique de Matusita et leur combinaison. Puisque la métrique prédéfinie ne garantit pas nécessairement que la correspondance la plus proche soit la cible d'intérêt, des méthodes utilisant une stratégie d'apprentissage ou un ajustement de la métrique de distance de manière adaptative, ont été récemment proposées. Elles opèrent par la recherche de la meilleure projection dans l'espace des caractéristiques.

Pour faire face au problème de dérive du suivi causé par l'accumulation des erreurs dues à la mise à jour du modèle d'apparence utilisant des échantillons positifs non fiables, et qui peuvent contenir des informations provenant de l'arrière-plan, les auteurs dans [Hon, 2018] proposent un algorithme de suivi basé sur l'apprentissage multi-métriques. Dans une première étape, une méthode de recherche est conçue pour réduire la charge de calcul de recadrage d'échantillons candidats pour l'image suivante. Ensuite, une collection de métrique de similarité est proposée pour

sélectionner les instances positives les plus correctes pour former la collection positive.. De plus, une méthode de sélection de classificateurs discriminatifs en ligne, qui exploite la stratégie de gradient moyen et de classifieurs faibles moyens, pour optimiser la fonction de marge entre les collections positives et négatives est présentée pour résoudre le problème du processus de sélection des faibles classifieurs. La méthode proposée réalise un suivi précis dans la majorité de séquences utilisées dans les expérimentations, cependant dans le cas des séquences qui présentent de fortes occlusions et dont l'objet se déplace rapidement, les résultats sont moins satisfaisants.

Pour gérer les changements de mouvement rapide, d'occlusion partielle et d'illumination, les auteurs dans [Mar, 2017] proposent un algorithme chaotique d'apprentissage multi-instance basé sur la théorie du chaos pour un suivi en ligne. Un modèle chaotique peut capturer la dynamique complexe de la région cible, pour former les classifieurs faibles, et la dimension fractale du modèle dynamique peut être ajustée, en tant que poids d'instance pour un apprentissage en ligne efficace. En plus, une approximation chaotique pour mettre à jour le modèle d'apparence est introduite, et pour sélectionner les caractéristiques discriminantes. L'apprentissage en ligne chaotique explore l'espace caractéristique pour mettre à jour le modèle d'apparence de la cible au moyen d'une carte chaotique. L'algorithme proposé réalise une bonne performance dans la plupart des séquences, alors que dans certaines situations difficiles où la cible est fortement occluse, ou dans le cas où l'objet a une apparence similaire de l'arrière-plan, la performance du suivi diminue.

La modélisation de l'apparence de l'objet joue un rôle crucial dans le processus de suivi, laquelle doit être capable de bien séparer l'objet de son arrière-plan, dans ce cadre les auteurs dans [Che, 2017] proposent un schéma de régression par paire pour modéliser l'apparence de la cible en utilisant la machine d'apprentissage extrême, deux contraintes sont appliquées, à savoir les observations cibles doivent avoir des résultats de régression différents de ceux de l'arrière-plan; tandis que les diverses observations cibles pendant le suivi doivent avoir des résultats de régression approximatifs. Une machine d'apprentissage extrême séquentielle en ligne est utilisée pour mettre à jour le modèle d'apparence résultant. La méthode proposée fonctionne bien et peut gérer de nombreux problèmes difficiles, mais dans

le cas d'une forte occlusion ou lorsque la cible disparaît pendant une longue période, elle fonctionne mal.

1.3.2. Les méthodes de suivi génératives

Les méthodes de suivi génératives s'appuient sur le modèle d'apparence de l'objet cible et le suivi s'effectue à travers la recherche de la position de la cible correspondante au meilleur score à l'aide d'une certaine mesure. Dans la suite nous allons présenter quelques avancées récentes dans les méthodes de suivi génératives.

1.3.2.1. Méthodes basées sur des modèles

La faiblesse des modèles d'apparence est l'un des principaux facteurs qui limitent les performances des algorithmes de suivi visuel. Les méthodes traditionnelles de suivi basé sur la correspondance de modèles ne s'adaptent pas aux changements d'apparence en raison des modèles fixes, par conséquent, les modèles dynamiques basés sur l'apprentissage en ligne sont utilisés pour représenter les changements d'apparence d'une cible causés par des changements de forme et d'illumination.

Dans [Jas, 2015] les auteurs ont proposé un algorithme pour le suivi des objets visuels dans lequel la validation et la mise à jour sont effectuées en fonction de seuil des séquences vidéo d'entrée. Le suivi fonctionne en mettant à jour le modèle d'apparence d'abord, puis en utilisant les informations de l'image entrante pour vérifier la validité du modèle. L'algorithme proposé peut conduire à des meilleures performances, en particulier dans les scènes où l'occlusion et la variation de l'apparence se produisent. Cependant, dans le cas des déformations, la méthode présente une grande sensibilité.

Dans [Jun, 2015] les auteurs proposent un algorithme de suivi visuel qui intègre l'apprentissage en ligne et le filtre à particules. Les variations d'apparence d'un objet sont effectivement apprises via un mécanisme d'apprentissage en ligne de la métrique. De plus, des représentations de caractéristiques spatialement pondérées sont utilisées en tenant compte à la fois des informations couleur et spatiales d'objets.

Dans [Luk, 2013] un modèle visuel à couche couplée est proposé pour un suivi efficace des cibles qui subissent d'importants changements d'apparence. Le modèle est une combinaison d'une couche locale et globale. La couche locale est un ensemble de correctifs locaux qui contraint géométriquement les changements dans l'apparence de la cible. L'ensemble s'adapte de manière probabiliste à l'apparence de la cible, en maximisant la répartition de l'articulation sur les contraintes géométriques, et les observations visuelles du modèle. Comme l'apparence de la cible change de manière significative, certaines des patches du modèle visuel cessent de correspondre aux parties visibles de la cible. Ces correctifs sont identifiés par la couche locale et sont progressivement retirés du modèle. La contribution des nouveaux correctifs dans la couche locale est limitée par la couche globale, qui encode les fonctions visuelles globales de la cible. La couche globale maintient un modèle probabiliste des caractéristiques visuelles globales de la cible telles que la couleur, la forme et le mouvement apparent. L'algorithme de suivi, fonctionne bien dans les cas où la couleur de l'arrière-plan est similaire à celle de la cible, mais il est difficile de faire face à des situations où l'apparence des objets est très similaire au fond et aux occlusions complètes de l'objet.

Pour surmonter les problèmes liés aux changements du mouvement de la cible, les auteurs dans [Yin, 2011] proposent une méthode d'estimation hiérarchique. La stratégie est basée sur une recherche approximative et stimulée par le phénomène selon lequel les mouvements les plus naturels sont cohérents avec le modèle gaussien linéaire dans un aspect global, tandis que le mouvement non linéaire est généralement limité dans l'aspect local. Les caractéristiques de mouvement globales telles que la position, la vitesse et l'accélération du mouvement sont estimées au niveau bas par le filtre de Kalman (KF) et l'estimation locale est réalisée par le filtre à particules pour traiter les non-linéarités et les statistiques non gaussiennes.

Le suivi visuel des objets en mouvement avec une apparence variable dans des environnements complexes est très difficile, les difficultés proviennent principalement de différents aspects, comme la complexité de l'arrière-plan, les changements de l'apparence de la cible causés par les changements environnementaux, et les changements de forme des cibles non rigides. Pour développer un algorithme de suivi assez robuste contre toutes sortes de conditions de suivi, l'une des solutions consiste à développer un modèle d'apparence de la

cible, qui permet d'extraire les caractéristiques robustes dans des conditions de suivi complexes. La modélisation de l'apparence d'une cible dans des vidéos est un problème d'extraction de caractéristiques. Les caractéristiques comprennent la région, la forme, la texture, les points etc.... Ces caractéristiques montrent différents avantages et inconvénients.

Dans le but de surmonter les inconvénients de chaque caractéristique dans la modélisation des cibles, l'intégration de multiples indices a reçu beaucoup d'attention de la part des chercheurs. Dans [Yur, 2012] les auteurs présentent un modèle d'apparence dynamique basé sur plusieurs indices pour le filtre à particules (PF). Pour adapter le modèle à l'apparence changeante de la cible, le poids de chaque indice est mis à jour dynamiquement à chaque image de la séquence vidéo. Dans cet objectif, les auteurs ont proposé deux façons pour évaluer le poids de chaque caractéristique, l'une consiste à comparer la précision de suivi et l'autre à évaluer la distribution des particules. En plus un système de mise à jour de poids en ligne est conçu. A chaque image vidéo, le modèle d'intégration multi indices dynamique proposé est mis à jour en ligne en fonction du changement des conditions de suivi, de cette façon le filtre à particules devient adapté aux environnements complexes et aux conditions modifiables.

Dans l'objectif de combiner plusieurs caractéristiques, les auteurs dans [Dou, 2014] utilise le filtre à particules et le modèle multiple interactif (IMM) pour intégrer trois sous modèles d'apparences, à savoir l'histogramme pondéré en arrière-plan corrigé, les motifs ternaires locaux parfaits et l'histogramme des gradients orientés. Les poids de chaque entité sont ajustés dynamiquement, ainsi l'algorithme proposé permet de suivre la cible dans des situations d'occlusions partielles.

Dans [Hui, 2016] les auteurs présentent un nouvel algorithme de suivi d'objets en ligne, en utilisant des caractéristiques multifonctions avec des poids adaptatifs ; en effet, deux caractéristiques, à savoir l'histogramme de gradient et de couleur, sont utilisées pour générer un ensemble de cartes de confiance, dans lesquelles la valeur de confiance de chaque pixel indique la probabilité que ce pixel appartienne à l'objet suivi. La caractéristique d'intensité couvre les informations d'énergie et la fonction histogramme du gradient couvre les informations de texture de l'objet suivi et de son arrière-plan environnant respectivement. Les caractéristiques de couleur mentionnées ci-dessus fournissent non seulement des caractéristiques de haut

niveau pour construire un modèle d'apparence plus stable, mais traitent également le suivi avec un arrière-plan de couleur encombré. La carte de confiance finale est obtenue en fusionnant les cartes de confiance de différents indices de manière adaptative. La position optimale de l'objet suivi peut être déterminée sur la base de la valeur maximale dans la carte de confiance finale fusionnée.

D'autres méthodes utilisent le mélange de modèles Gaussien (GMM) pour représenter l'apparence de l'objet, dans [Jun, 2014] les auteurs présentent une méthode adaptative avec des modèles de mélange gaussien, pour faire face à des variations significatives de l'apparence de la cible pendant le suivi des objets. Le modèle d'apparence est progressivement mis à jour à l'aide d'un algorithme en ligne de maximisation d'espérance (EM). Pour atteindre la robustesse vis-à-vis des valeurs aberrantes provoquées par une erreur de suivi ou une occlusion partielle lors de la mise à jour des modèles d'apparence. La zone cible est divisée en sous-régions et les modèles d'apparence sont estimés indépendamment pour chacune de ces sous-régions.

Pour faire face à la variabilité de l'apparence de l'objet et des perturbations extrinsèques, [Xue, 2012] présentent une nouvelle approche pour le suivi basée sur le tenseur des caractéristiques. En effet l'image est convertie en champ tensoriel pour produire des caractéristiques plus discriminantes et coder l'apparence cible de manière probabiliste avec le modèle de mélange gaussien. Les paramètres du modèle sont obtenus par un algorithme EM modifié en utilisant tous les échantillons de tenseurs extraits de la zone cible. Une procédure d'apprentissage différentiel est utilisée pour mettre à jour les paramètres du modèle pour s'adapter aux changements d'apparence dans le temps.

Pour gérer les changements d'échelle et de rotation de la cible, ainsi que les changements d'apparence, dans [Vas, 2015] les auteurs ont proposé un algorithme de suivi en temps réel probabiliste, où l'apparence de la cible est représentée par un modèle de mélange gaussien, la fonction de vraisemblance des mélanges gaussiens est pondérée spatialement pour intégrer les informations de localisation de la cible, par conséquent la détermination de la position de la cible est obtenue en maximisant sa vraisemblance pondérée dans la séquence d'images en utilisant l'optimisation basée sur le gradient. Pour surmonter la difficulté de changement d'apparence, les auteurs proposent de mettre à jour dynamiquement le modèle de la

cible en insérant de nouveaux composants au GMM en utilisant des pixels près de la cible qui ont peu de vraisemblance. Ainsi, si l'importance d'une composante devient suffisamment petite, la composante est éliminée du GMM.

Dans [Vas, 2011] les auteurs proposent une méthode de suivi d'objets visuels reposant sur la modélisation de l'apparence de l'objet à l'aide d'un mélange gaussien. L'algorithme EM est appliqué pour calculer le GMM initial, et l'emplacement de l'objet est estimé dans un cadre différentiel par la direction du gradient de maximisation différentielle de l'espérance (EMD) par rapport à l'espace d'image bidimensionnel. L'algorithme EM n'est pas appliqué, car les moyennes et les variances ne changent pas. En outre, l'algorithme proposé est combiné avec un filtre de Kalman pour gérer les occlusions. La prédiction réalisée par EMD est considérée comme l'observation d'un filtre de Kalman dont les paramètres d'état sont automatiquement déterminés sur la base de l'historique des positions récentes.

Dans [Vas, 2017] les auteurs utilisent un mélange de Von Mises pondéré comme modèle d'apparence de la cible. Les poids de mélange, fournis par un noyau spatial, ainsi que les valeurs de teinte sont utilisés pour estimer les paramètres du modèle. La distribution de Von Mises est adaptée aux données circulaires et elle est utilisée afin d'éliminer les inconvénients du suivi basé sur le noyau causé par des décalages éventuels des classes d'histogrammes de la cible. Vu que seule la composante teinte de la cible est considérée, de nombreux paramètres de l'algorithme peuvent être pré-calculés pour des paramètres donnés.

1.3.2.2. Méthodes basées sur l'analyse par sous-espace

Au lieu d'utiliser un modèle simple pour représenter le modèle d'apparence pour le suivi, une représentation sous-spatiale par apprentissage apprise en ligne peut être utilisée pour fournir une représentation compacte d'une cible, et refléter les changements d'apparence pendant le suivi. Dans les modèles d'apparence basés sur les sous-espaces, les matrices des valeurs de pixels dans les régions d'image sont converties en vecteurs, et des informations statistiques globales sur les valeurs de pixel sont obtenues par analyse en composantes principales (ACP) pour les vecteurs. Dans [Ste 2012] les auteurs proposent un algorithme pour l'apprentissage en ligne avec une famille de noyaux indéfinis (non positifs). Dans cet article

l'analyse des composantes principales à noyau (KPCA) a été étendue à un espace de Hilbert à noyau de reproduction à l'espace de Kerin. Ensuite une KPCA incrémentale a été formulée dans l'espace Kerin, qui ne nécessite pas le calcul de pré-images.

Dans [Wei, 2012], les auteurs ont proposé un algorithme d'apprentissage incrémentiel par sous-espace, basé sur la métrique Riemannienne log-euclidienne et des matrices positives symétriques, dans lequel les matrices de covariance des caractéristiques d'image sont mappées dans un espace vectoriel. A la Base de l'algorithme d'apprentissage de sous-espace, les auteurs ont développé un modèle d'apparence de division par bloc log-euclidien, qui capture à la fois les informations de la disposition spatiale, globale et locale sur les apparences d'objets. Le suivi d'objet unique et le suivi multi-objet avec gestion des occlusions sont ensuite réalisés par inférence bayésienne d'état basée sur le filtrage de particules. Durant le suivi, la mise à jour incrémentale du modèle d'apparence de division de bloc log-euclidienne, capture les changements de l'apparence de l'objet.

Dans [Li, 2008], les auteurs présentent un algorithme de suivi basé sur l'apprentissage par sous-espace. Une représentation de la base propre basée sur la vue est utilisée pour modéliser les variations d'apparence, en supposant qu'elles sont contenues dans un sous-espace fixe. Cependant, l'algorithme ne fonctionne pas bien dans les scènes encombrées avec de grands changements d'éclairage.

Les auteurs dans [Ho, 2004], présentent un algorithme de suivi visuel basé sur l'apprentissage par sous-espace linéaire. Dans chaque mise à jour, le sous-espace est recalculé en utilisant uniquement les lots récents des résultats de suivi. Cependant, l'utilisation des moyens du suivi, entraîne un nombre d'images consécutives car les échantillons d'apprentissage peuvent perdre de leur précision, et le calcul du sous-espace en utilisant uniquement les récents résultats peut rendre le suivi instable. Dans [Sko, 2003] les auteurs ont présentés un algorithme PCA incrémental pondéré pour l'apprentissage sub-spatial. Sa limite est que chaque mise à jour inclut seulement un nouvel échantillon, plutôt que des multi-échantillons, et par conséquent il est nécessaire de mettre à jour le sous-espace à chaque image.

Dans [Li, 2004] l'auteur a proposé un algorithme PCA incrémental pour l'apprentissage des sous-espaces. Il peut gérer plusieurs multi-échantillons à chaque fois. La mise à jour du modèle PCA est effectuée directement à partir des vecteurs propres précédents et d'un nouveau vecteur d'observation. Néanmoins, la méthode proposée rencontre des problèmes lorsque le modèle PCA initial contient des valeurs aberrantes significatives. Dans ces circonstances, le modèle prendrait beaucoup de temps à se rétablir. Dans [Ros, 2008] les auteurs ont proposés un algorithme de suivi généralisé basé sur la méthode incrémentale d'apprentissage de sous-espace image-vecteur. La méthode proposée supprime l'hypothèse que la moyenne des données précédentes est égale à la moyenne des nouvelles données dans [Wu, 2009]. Cependant, elle ne capture et ne modélise pas directement les corrélations spatiales entre les valeurs de pixels dans la région de suivi.

Pour réduire le coût de calcul, les auteurs dans [Wu, 2013] ont présenté une approche de suivi visuel en temps réel avec des mises à jour du modèle de covariance incrémentale. Chaque échantillon est associé à un poids pour donner aux nouveaux échantillons une plus grande influence sur l'estimation de la représentation du tenseur de covariance actuelle.

Les auteurs dans [Ke, 2013] ont présenté un algorithme d'apprentissage de sous-espace basé sur l'incorporation de graphe localement connecté (LCG). Le modèle de sous-espace sémantique est appris en construisant un graphe supervisé avec plusieurs types d'échantillons cibles étiquetés. Le LCG combine les cibles avec des erreurs mineures pour construire un sous-espace à travers la projection qui est préalablement établie. De plus, selon les informations sémantiques, les échantillons de la cible sont séparés en catégories d'états. LCG applique des contraintes d'étiquette supplémentaires pour lier le sous-graphe de chaque catégorie d'état afin de rendre le graphique plus compact et plus raisonnable pour gérer le problème de dérivations.

Les limitations générales des modèles d'apparence basés sur les sous-espaces courants, est qu'ils n'utilisent pas directement les relations locales des valeurs de pixels de l'objet, qui peuvent être quantitativement représentées par des dérivées d'intensité de pixel, etc. Ces relations locales sont, dans une large mesure, invariantes à des changements environnementaux complexes. Par exemple, les variations dans l'éclairage peuvent provoquer de grands changements dans les

valeurs des pixels, tandis que les changements dans les dérivées spatiales des intensités des pixels peuvent être beaucoup moins importants. En plus dans le cas des occlusions, il est difficile de mettre à jour les modèles d'apparence d'objet pendant les occlusions. C'est un défi pour les modèles d'apparence basés sur des sous-espaces d'utiliser les relations locales entre les pixels des objets pour augmenter la robustesse du suivi des objets.

1.3.2.3. Méthodes basées sur la représentation parcimonieuse

La modélisation d'apparence est l'une des conditions préalables les plus critiques pour un suivi visuel robuste. Cependant, la conception d'un modèle d'apparence efficace est une tâche difficile en raison des variations d'apparence causées par les encombrements de fond, la déformation de l'objet, les occlusions partielles et les changements d'illumination. Pour gérer ces variations, un modèle de bonne apparence est souhaité pour répondre à deux exigences: l'adaptabilité aux variations d'apparence intrinsèques et extrinsèques. En plus la complexité de calcul est un problème très important, puisque la vitesse de suivi en temps réel est une exigence pratique de la plupart des applications de haut niveau, telles que la reconnaissance et l'analyse. La modélisation de l'apparence basée sur la représentation parcimonieuse a reçu une attention considérable dans la communauté du suivi visuel. Les méthodes de suivi visuel basé sur le codage parcimonieux peuvent être grossièrement classées en trois classes: modélisation de l'apparence par codage parcimonieux (MACS), recherche de cibles basée sur représentation parcimonieuse (RCRP) et la combinaison de MACS et de RCRP.

Modélisation de l'apparence par codage parcimonieux (MACP)

Les auteurs dans [Zha, 2010], ont proposé un modèle d'apparence basé sur la distribution de la réponse des fonctions de base, apprises en utilisant l'analyse des composantes indépendantes (ICA). Il est à noter que ICA est un cas particulier de codage parcimonieux et a également des propriétés similaires avec des champs réceptifs de cellules simples dans le cortex visuel. Les fonctions de base apprises sont appelées «fonctions de base générales», car les images d'apprentissage ne proviennent pas de la séquence de suivi mais de toute image naturelle. Pour adapter les fonctions de base générales à la tâche de suivi, une stratégie de sélection de base basée sur le gain d'entropie a été utilisée pour sélectionner les

fonctions de base qui codent les caractéristiques dominantes dans la cible. Enfin, le modèle d'apparence de la cible suivie est représenté par la distribution de réponse des fonctions de base sélectionnées. Le modèle d'apparence proposé est robuste aux variations d'apparence, en particulier à l'occlusion partielle due à la stratégie de sélection des caractéristiques. Toutefois, ce modèle n'est pas suffisamment discriminant pour distinguer la cible de l'arrière-plan, car les fonctions de base apprises sont trop générales pour coder la différence entre les correctifs cible et d'arrière-plan.

Pour augmenter encore la discrimination, un modèle d'apparence plus discriminatif basé sur un codage parcimonieux a été proposé dans [Liu, 2011] où le dictionnaire a été directement appris, en utilisant des parties échantillonnées à partir de l'image cible. Les fonctions de base apprises sont plus spécifiques pour représenter les correctifs échantillonnés à partir de la cible. D'un autre côté, le dictionnaire appris a pour résultat une représentation non- parcimonieuse pour les parcelles d'arrière-plan. Par conséquent, les coefficients de représentation sont plus faciles à distinguer la cible de l'arrière.

Dans [Wan, 2012] a également utilisé les parties échantillonnées depuis l'image cible pour l'apprentissage du dictionnaire. En effet pour atténuer le problème de dérive visuelle souvent rencontré, un algorithme en deux étapes est proposé pour exploiter à la fois les informations de vérité de terrain de la première image et les observations obtenues en ligne. Les auteurs utilisent un ensemble de fonctionnalités ou un ensemble de classifieurs boostés, l'algorithme proposé apprend des codes parcimonieux et un classifieur linéaire directement à partir de correctifs d'image bruts. Contrairement aux méthodes de suivi basées sur la représentation parcimonieuse, qui codent l'apparence holistique des objets dans un cadre génératif, l'algorithme proposé utilise une formulation de discrimination qui facilite la tâche de suivi dans des environnements complexes. Cependant, il y a deux inconvénients, le premier est qu'il est très sensible au bruit d'image car chaque code local est un élément du vecteur de caractéristiques concaténé. Une fois que le bruit apparaît dans l'image, le vecteur de caractéristiques résultant sera directement affecté. Le deuxième, est que la dimension du vecteur final est extrêmement élevée.

Dans [Jia, 2012], le modèle parcimonieux de codage local a été proposé, qui utilise un ensemble de modèles cibles. Chaque image du modèle est divisée en un ensemble de parcelles locales superposées. Les parcelles échantillonnées à partir de toutes les images cibles sont utilisées comme dictionnaire pour représenter de manière parcimonieuse, les parcelles échantillonnées à partir du candidat, en utilisant le même chemin. Tous les parcelles locales de la cible avec une relation spatiale fixe peuvent refléter la structure cible. Pour calculer le vecteur final qui capture la structure possible, tous les codes locaux calculés à partir de l'image candidate ont été ensuite regroupés par un opérateur d'alignement, qui calcule l'élément caractéristique pour chaque parcelle, comme la somme des codes correspondant aux fonctions de base à la même position avec la parcelle. Toutefois pour améliorer ce modèle de codage structurel parcimonieux, le dictionnaire devrait également inclure des parcelles de l'image de fond pour décrire la structure de l'arrière-plan, ce qui devrait être utile pour distinguer la cible de l'arrière-plan.

Recherche de cibles basée sur représentation parcimonieuse (RCRP)

Dans [Wri, 2009], une méthode de reconnaissance faciale basée sur la représentation parcimonieuse a été proposée qui exploite la capacité de recherche de sous-espace minimal, motivés par ce travail, les auteurs dans [Mei, 2009] ont considéré le suivi visuel comme un problème de recherche à deux sous-espaces, où le premier sous-espace était couvert par un ensemble de modèles cibles et le second sous-espace par un ensemble de modèles triviaux. Les modèles triviaux sont des vecteurs de colonne d'une matrice d'identité, également appelée base de pixels d'identité. Dans le cadre du filtre à particules, le suivi visuel a pour but de rechercher à quel sous-espace appartient chaque candidat cible. Si le résultat de la recherche est le premier sous-espace, le candidat est considéré comme le résultat du suivi.

Au lieu de faire une recherche sur chaque candidat, les auteurs dans [Zha, 2009] ont proposé de rechercher le modèle cible à partir de deux sous-espaces. Le premier sous-espace est couvert par un ensemble de candidats similaires au modèle cible. Le second sous-espace est couvert par un ensemble de candidats correspondant à l'arrière-plan. Lors de la représentation du modèle en utilisant tous les candidats des deux sous-espaces, le candidat correspondant au plus grand coefficient est considéré comme le résultat du suivi.

Dans les méthodes de suivi mentionnées ci-dessus, le dictionnaire est la collection d'un certain nombre d'images vectorisées. La plupart des travaux utilisent les images en niveaux de gris échantillonnées pour obtenir une représentation de faible dimension. Cependant, les images en niveaux de gris sont sensibles aux bruits, en particulier dans les scènes extérieures. D'autre part, une cible est plus facile à distinguer des autres cibles lorsque certaines caractéristiques sont utilisées.

Dans [Tzi 2011], des gradients le long des directions horizontales et verticales ont été utilisés pour remplacer les caractéristiques d'intensité sous-échantillonnées. Cependant, il est difficile de représenter un échantillon comme une combinaison linéaire d'un sous-espace dans le domaine de l'angle. Pour utiliser des données angulaires, chaque vecteur d'angle est mappé sur un nouveau vecteur en concaténant ses valeurs de cosinus et de sinus. Des descripteurs de plusieurs caractéristiques sont complémentaires pour la recherche de la cible.

Au lieu d'utiliser un seul type de descripteur d'entités pour construire le dictionnaire. Les auteurs dans [Wu, 2011], ont proposé d'utiliser plusieurs descripteurs d'entités dans le dictionnaire. Chaque atome de dictionnaire est obtenu en concaténant plusieurs descripteurs d'entités extraits de l'image cible. Lors du calcul des coefficients de représentation, la contrainte parcimonieuse joue un rôle important pour faire en sorte que les différents descripteurs soient en concurrence les uns avec les autres pour représenter le candidat cible, ce qui permet de fusionner plusieurs types de descripteurs de caractéristiques. Un dictionnaire fixe ne peut pas capturer la variation d'apparence. Par conséquent, le sous-espace doit être mis à jour au fil du temps lorsque le résultat du suivi est disponible.

Dans [Lin, 2010], les auteurs proposent un schéma de mise à jour, pour faire face aux problèmes de dérive, lors de la poursuite d'un véhicule dans des vidéos infrarouges, où la qualité d'image et le contraste entre cible et arrière-plan sont faibles. Pour surmonter ce problème, ils ont proposé un schéma de mise à jour de modèle probabiliste, dans lequel la mise à jour de pondération se produit uniquement, lorsque la probabilité d'observation du résultat de suivi actuel est supérieure à un seuil. Pour réduire la complexité du calcul, deux types de méthodes peuvent être adoptés, l'un d'entre eux consiste à réduire le coût de calcul de chaque minimisation de la norme L1. L'autre consiste à réduire le nombre de minimisations

de la norme L1. Au lieu de réduire la dimension de la minimisation de la norme L1, en échantillonnant les images recadrées à la fois pour la cible et les candidats.

Pour réduire la dimension de la représentation linéaire, dans [Liu, 2010] les auteurs ont proposé d'utiliser la matrice de covariance pour représenter la cible ou le candidat. La matrice de covariance présente plusieurs avantages significatifs, elle permet une fusion de différentes caractéristiques, où les propriétés spatiales et statistiques ainsi que leur corrélation sont caractérisées, ainsi que sa petite dimension. Dans [Li, 2011], un algorithme de suivi en temps réel est proposé, il exploite la propriété d'isométrie restreinte pour réduire la dimension du dictionnaire en multipliant une matrice de hachage.

D'autres méthodes visent à réduire la taille du dictionnaire, en particulier la taille du dictionnaire d'occlusion, pour réduire la complexité du calcul. Dans [H. Liu, 2011], les auteurs ont intégré un modèle de mouvement dans la représentation parcimonieuse. A partir d'un état initial de la cible, une procédure d'optimisation basée sur la descente de gradient est utilisée itérativement pour trouver les coefficients de représentation parcimonieuse et le vecteur de gradient correspondant. A chaque itération, un nouveau candidat est obtenu sur la base du vecteur gradient calculé et du dernier candidat. Après un petit nombre d'itérations, la valeur minimale locale de la minimisation de la norme L1 est obtenue, par conséquent, le nombre de minimisations de la norme L1 est réduit

Combinaison des méthodes MACP et RCRP

Certains chercheurs ont essayé de combiner les méthodes MACP et RCRP pour créer un modèle collaboratif basé sur la densité spatiale pour le suivi visuel. Le RCRP utilise le dictionnaire cible et le dictionnaire d'arrière-plan pour représenter chaque candidat. La méthode MACP utilise la mise en pool de concaténation pour obtenir un vecteur de caractéristique de type histogramme pour le modèle cible et pour le candidat. Le poids du candidat basé sur MACP est calculé par interaction histogramme. Le poids final du candidat est calculé par une simple fusion. Leur travail combine les avantages de MACP et RCRP et a rapporté une meilleure performance de suivi, par rapport à d'autres méthodes de pointe.

Dans [Zho, 2012], les auteurs proposent un algorithme de suivi d'objets utilisant un modèle collaboratif. Comme le principal défi pour le suivi des objets est de prendre

en compte les changements d'apparence, dans ce travail un modèle d'apparence qui exploite à la fois les modèles holistiques et les représentations locales est présenté. Les auteurs ont développé un classifieur discriminatif basé sur la densité et un modèle génératif basé sur la parcimonie. Dans le module de classifieur discriminatif basé sur la densité, une méthode est introduite pour calculer la valeur de confiance qui assigne plus de poids au premier plan qu'à l'arrière-plan. Dans le module génératif basé sur la parcimonie, une nouvelle méthode est proposée basée sur l'histogramme qui prend en compte les informations spatiales de chaque parcelle avec un schéma de traitement d'occlusion. En outre, le schéma de mise à jour prend en compte à la fois les dernières observations et le modèle d'origine, permettant ainsi au suivi de gérer le changement d'apparence et d'atténuer le risque de perdre le suivi.

Dans [Min, 2016], les auteurs ont proposé une méthode de suivi, utilisant Haar collaboratif compressif de la collection de modèles, comprenant des modèles positifs (premier plan), négatifs (arrière-plan) et modèles carrés. En tant que méthodes traditionnelles, la cible est peu représentée dans les modèles positifs uniquement. Mais, comme le calcul de la complexité de la représentation parcimonieuse, les auteurs ont utilisé la caractéristique de type Haar compressive des modèles positifs et carrés pour faire face à l'occlusion. Afin d'améliorer la fiabilité au cours du processus de suivi, une expression discriminative basée sur la parcimonie est développée en utilisant des informations de premier plan et d'arrière-plan. L'expression avec erreur de reconstruction de premier plan et erreur de reconstruction de fond est formulée, pour trouver l'échantillon le plus correct.

1.3.3. Méthodes combinées

Les méthodes de suivi génératives, modélisent uniquement l'apparence de la cible et échouent facilement dans le cas d'un arrière-plan encombré. En revanche, les méthodes discriminatives sont capables de gérer des changements d'apparence significatifs, des arrière-plans encombrés et des occlusions à court terme. Cependant, elles sont sensibles au bruit, et souffrent généralement du problème de la dérive et sont facilement perturbées par d'autres objets qui sont similaires en apparence à la cible. Par conséquent, certains chercheurs ont récemment tenté de tirer parti des deux types de méthodes.

Pour aborder le problème des changements dynamiques de l'apparence et de l'environnement, les auteurs dans [Jia, 2017] ont proposé un algorithme de suivi, en intégrant le modèle génératif et discriminatif. Le modèle d'apparence d'objet est constitué d'un modèle cible génératif et d'un classifieur discriminatif. Pour le modèle cible génératif, le modèle d'apparence parcimonieux structurel pondéré est considéré, combinant les parcelles caractérisées par la valeur du niveau de gris et l'histogramme des dégradés orientés. En échantillonnant les positifs et les négatifs, les fonctionnalités de regroupement d'alignement sont obtenues sur la base du dictionnaire de correctifs, par le biais du codage parcimonieux local, puis SVM est utilisé pour former le classifieur discriminatif. La méthode proposée est intégrée dans un cadre d'inférence bayésienne pour le suivi visuel. Une méthode d'appariement combinée est adoptée pour améliorer la distribution des propositions du filtre à particules. De plus, afin de rendre l'algorithme adaptatif, le dictionnaire de correctifs et le classificateur discriminatif sont mis à jour par apprentissage incrémental toutes les cinq images.

Dans [Man, 2018], un algorithme de suivi temps réel est conçu sur la base d'un algorithme de hachage avec des indices spatio-temporels, la relation spatiale est combinée avec la relation temporelle entre les entités sur différentes images vidéo. Un cadre hiérarchique est proposé et effectue des tâches d'appariement et de classification pour générer un système de suivi approximatif. Un modèle génératif est développé sous un filtre à particules modifié avec des empreintes digitales de hachage pour l'appariement approximatif, par le maximum a posteriori et un modèle discriminatif pour la classification fine en maximisant une carte de confiance basée sur un modèle de contexte. La carte de confiance révèle la dynamique spatio-temporelle de la cible vue que, l'empreinte digitale est simplement un vecteur binaire et le filtre à particules modifié utilise seulement un petit nombre de particules.

Les auteurs dans [Ash, 2017] ont proposé une méthode pour suivre un objet dans l'imagerie infrarouge basée sur une combinaison d'approches discriminatives et génératives. La technique discriminatif utilise deux méthodes complémentaires telles que le filtre de corrélation à noyau avec la caractéristique spatiale, et le classificateur d'AdaBoost avec des caractéristiques d'intensité de pixel pour fonctionner en parallèle. Après avoir obtenu des positions optimisées grâce à des

approches discriminatives, la technique générative est appliquée pour déterminer la meilleure localisation de la cible en utilisant une méthode de recherche linéaire. Contrairement aux algorithmes de base, la méthode proposée estime l'échelle de la cible par l'estimation de l'homographie de Lucas-Kanade (L-K).

Afin de suivre et réacquérir un objet inconnu avec des données d'étiquetage limitées, les auteurs dans [Tha, 2014] ont proposés d'apprendre en ligne les changements de l'apparence, en raison des variations du point de vue, des changements d'éclairage et de l'occlusion et incrémenter un modèle qui encode toutes les variations d'apparence tout en effectuant un suivi. Pour résoudre ce problème d'apprentissage semi-supervisé, les auteurs ont proposés une approche de d'apprentissage avec un filtre à particules en cascade pour étiqueter les données entrantes en continu, et actualiser en ligne les modèles hybrides génératifs et discriminatifs. La manière d'organisation en cascade du filtre à particules, permet l'évaluation efficace des modèles d'apparence multiples avec des coûts de calcul différents. L'algorithme en ligne proposé, fournit un suivi temporel local qui s'adapte aux changements d'apparence. De plus, il fournit une capacité de détection spécifique qui permet de réacquérir un objet après une occlusion totale.

Afin de suivre les objets qui subissent des changements de rotation et de pose, les auteurs dans [Liu, 2015] ont proposés un algorithme combinant le modèle discriminatif global et le modèle génératif local. Initialement, les coefficients d'approximation des ondelettes et le motif binaire local (CLBP) sont combinés pour représenter les caractéristiques globales de l'objet. Avec le descripteur d'apparence globale obtenu, les auteurs ont utilisés l'apprentissage métrique discriminatif en ligne pour différencier l'objet cible de l'arrière-plan. Pour éviter le problème de dérive résultant du modèle discriminatif global, un nouveau modèle local géométrique spatial génératif est introduit. Basé sur les fonctions SURF, le modèle local génératif quantifie l'information de structure géométrique en échelle et en angle. Ensuite, ces modèles globaux et locaux sont combinés pour qu'ils puissent être bénéfiques les uns aux autres.

Dans [Lei, 2013] les auteurs ont incorporés un algorithme génératif et discriminatif en cascade. Le suiveur génératif filtre les candidats les plus faciles au début et conserve quelques échantillons les plus confus. Le suiveur discriminatif réévalue ensuite ces échantillons en utilisant l'analyse discriminante Moindres carrés

partiels (PLS). Les deux suiveurs sont mis à jour en collaboration pour s'adapter aux changements d'apparence pendant le suivi. L'approche proposée apprend explicitement la différence d'apparence entre la cible et les distractions les plus confus.

Dans [Yun, 2008], les auteurs présentent une solution pour suivre un objet visuel sous des changements de l'illumination, le point de vue, la pose, l'échelle et l'occlusion. Dans le cadre de l'apprentissage bayésien séquentiel, un suiveur basé sur un modèle discriminatif avec un algorithme de machine vectorielle de pertinence rapide est introduit en premier, puis un suiveur basé sur un modèle génératif avec un modèle de mélange séquentiel gaussien est présenté après. Enfin, une hiérarchie à trois niveaux est présentée pour étudier différents schémas pour combiner les modèles discriminatifs et génératifs pour le suivi. La combinaison de modèles hiérarchiques présentées, contient la combinaison d'apprenants (au niveau un), la combinaison de classificateurs (au niveau deux) et la combinaison de décisions (au niveau trois).

1.4. Conclusion

Ce chapitre à été consacré à un état d'art sur la représentation des objets et les méthodes récentes de suivi de la littérature. La représentation des objets constitue une étape majeure dans le suivi visuel, le choix des bonnes caractéristiques pour représenter l'objet suivi est important. Ce choix doit être fait de façon à rendre le suivi robuste vis-à-vis des problèmes en rapports avec le changement de l'apparence tout au long de la séquence ainsi que dans le cas des scènes encombrées où plusieurs objets ont une apparence assez similaire. Dans ce contexte, plusieurs caractéristiques ainsi que les avancées représentatives récentes dans des descripteurs ont été présentées.

Les méthodes classiques et contemporaines de suivi visuel mono-objet ont été présenté en termes de trois classes : discriminative, générative et combinées. Les méthodes discriminatives considèrent le problème de suivi comme un problème de classification dont l'objectif est de séparer l'objet et l'arrière-plan en utilisant à la fois les informations sur l'objet et son voisinage. Alors que les méthodes génératives, représentent l'objet par un modèle d'apparence qui peut être évidemment convolué par un noyau, ainsi le processus de suivi recherche le

candidat dont le modèle d'apparence observé est le plus proche de celui du modèle de base. Différentes méthodes représentant les deux classes sont discutées à savoir les méthodes basées sur un apprentissage en ligne, les méthodes basées sur une analyse sous espace et les méthodes basées sur la modélisation d'apparence et modélisation parcimonieuse. En plus les principales caractéristiques et la théorie de base des approches combinées sont examinées, en effet ces méthodes modélisent l'apparence de l'objet suivi tout en tenant compte de son entourage.

Il ressort de la classification précédente, plusieurs remarques et réflexions importantes. Premièrement, l'utilisation d'une seule caractéristique pour modéliser l'apparence de l'objet, ne favorise pas l'aspect discriminatif vis-à-vis des conditions complexe. Deuxièmement, la fusion de plusieurs caractéristiques améliore le pouvoir discriminatif du modèle, par contre le temps de calcul devient élevé. Finalement, la majorité des méthodes de suivi s'intéressent au modèle d'apparence vu son importance, alors que la modélisation du mouvement de la cible est négligée, malgré son influence majeure pour le succès du suivi surtout dans le cas du mouvement rapide ou complexe.

Dans la suite de ce rapport, nous proposons des approches pour surmonter les problèmes discutés avant. Nous proposons une méthode de suivi qui fonctionne sous différentes hypothèses et contraintes, en plus les caractéristiques d'apparence et de mouvement sont à la fois prises en compte. Ainsi, un modèle dynamique adaptative est élaboré pour couvrir différents styles de mouvement, alors qu'un modèle visuel basé sur la fusion de plusieurs indices complémentaires est proposé pour améliorer le pouvoir discriminatif avec un coût de calcul raisonnable.

Chapitre II : Méthodes bayésiennes pour le suivi visuel

Dans ce chapitre nous présentons une description générale des filtres de prédiction pour l'estimation d'états dans le cadre bayésien en particulier pour le problème de suivi visuel. Premièrement la modélisation mathématique d'un système sous forme de représentation de l'état dynamique est introduite. Puis un aperçu de la théorie du filtrage bayésien est exposé, suivi des différents filtres prédictifs, à savoir le filtre de Kalman, le filtre H^∞ et le filtre à particules, dont chacun est approprié pour un type de représentation de l'incertitude et de système. Nous terminons ce chapitre par présenter le principe de base de notre approche pour la prédiction de l'état pour le suivi visuel. La dernière partie porte sur des expérimentations réalisées pour l'évaluation de méthode de prédiction, les simulations, ainsi les résultats expérimentaux sont exposés et discutés..

CHAPITRE II : METHODES BAYESIENNES POUR LE SUIVI VISUEL.....	57
II.1. INTRODUCTION.....	58
II.2. FILTRE BAYESIEN	60
<i>II.2.1. Prédiction bayésienne</i>	<i>60</i>
<i>II.2.2. Le filtre de Kalman</i>	<i>62</i>
<i>II.2.3. Le filtre H^∞.....</i>	<i>64</i>
<i>II.2.4. Le Filtre à Particules</i>	<i>68</i>
II.3. FILTRE H^∞ PF POUR LE SUIVI VISUEL.....	69
II.4. EVALUATION	72
II.5. CONCLUSION.....	80

II.1. Introduction

De nombreux problèmes en sciences de l'ingénieur exigent l'estimation de l'état d'un système qui évolue avec le temps, par exemple la robotique, la vision par ordinateur, etc. La prédiction de l'état du système et l'estimation des paramètres sont réalisées essentiellement par des filtres prédictifs également appelés filtres bayésien. Ils appliquent les règles bayésienne de probabilité conditionnelle pour combiner le comportement prédit et les observations les plus significatives. Dans le processus de prédiction, une formulation mathématique du problème étudié est nécessaire. La configuration du système est décrite par un ensemble de paramètres représentés par le vecteur d'état. Une bonne représentation mathématique permet de mieux déduire les propriétés du système réel et de mieux comprendre à l'avance son comportement et ainsi faire face aux facteurs dus à une modélisation qui peut être plus au moins loin de la réalité ainsi qu'aux manques de précision des observations ou même leur absence, ce qui peut conduire à des résultats légèrement incorrects, plusieurs filtres prédictifs doivent intégrer la notion d'incertitudes.

En général, une large gamme de systèmes peut être représentée sous forme d'espace d'état, plusieurs variables à valeurs réelles sont utilisés pour décrire l'état du système dont elles sont regroupées dans un vecteur d'état. Une formulation naturelle de la dynamique du système est représentée par une équation de différence.

Il y a deux espaces d'état : l'espace d'état à temps continu et l'espace d'état à temps discret. En effet la plupart des grandeurs physiques sont considérées comme étant à temps continu et à valeurs réelles:(températures, vitesses, pressions, etc.). Elles sont modélisées par des systèmes dynamiques continus. Cependant, dans les traitements informatiques, ces grandeurs physiques sont représentées par des quantités discrètes. Les systèmes à temps discret sont intrinsèquement obtenus par discrétisation de systèmes à temps continu. Dans de tels types de systèmes, les entrées, les variables d'espace d'état et les sorties ont une forme discrète et les modèles de système sont généralement représentés sous la forme de matrice de transition. Dans notre travail, le problème est formulé de façon mathématique par un système dynamique discret dont le vecteur d'état $X_k \in \mathbb{R}^n$ regroupe plusieurs

paramètres réels qui décrivent le comportement dynamique de la cible Figure II-1 et Figure II-2.

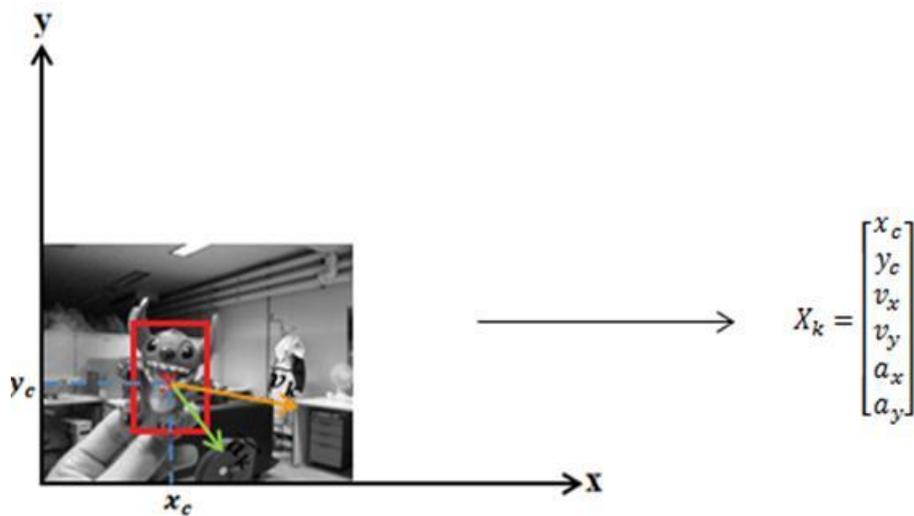


Figure II-1. Un modèle d'état d'un objet en suivi visuel, avec x , y sont les positions cartésiennes, v est la vitesse, a est l'accélération.



Figure II-2. Un modèle d'état d'un objet en suivi visuel, avec x_c , y_c sont les positions cartésiennes, la taille de la zone suivi est définie par w et h , θ est l'orientation.

Le système dynamique discret est construit de façon à considérer que l'état suivant ne dépend que de l'état actuel, par conséquent la dynamique est représentée comme une simple multiplication matricielle Equation II-1.

$$X_{k+1} = A X_k + G w_k \quad (\text{II.1})$$

Où w_k est l'incertitude associée à l'état, A est la matrice de transition à temps discret.

Pour compléter la formulation mathématique du système dynamique, nous modélisons l'observation par Equation II-2.

$$Y_k = H X_k + v_k \quad (\text{II.2})$$

Où H est la matrice de mesure, alors que v_k est l'incertitude associée à la mesure.

Le signal à estimer est donnée par :

$$Z_k = L X_k \quad (\text{II.3})$$

Dans le processus de prédiction, d'estimation et le contrôle, pour faire face à des quantités inexacts ou des mesures incertaines. On a besoin de modéliser les incertitudes, différentes options pour modéliser et représenter les incertitudes existes, telles que les intervalles et les régions de confiance, ou les variables aléatoires et les probabilités. Dans notre approche de suivi visuel, nous utilisons uniquement des variables aléatoires et spécialement les variables aléatoires gaussiennes pour représenter les incertitudes associées à l'état et aux observations.

II.2. Filtre Bayésien

II.2.1. Prédiction bayésienne

Dans la théorie de l'estimation et la théorie de la décision, l'estimation Bayésienne permet de déterminer la meilleure prédiction d'un vecteur de paramètre inconnu, elle est basée sur un modèle d'évolution du système combiné aux mesures ou observations de l'état du système.

Considérons l'estimation de la valeur d'un vecteur de paramètres aléatoires X , et un vecteur d'observation connexe Y . De la règle de Bayes, la prédiction Bayésienne consiste à construire la fonction de densité de probabilité postérieure $p(X_k|Y_{1:k})$ de l'état basé sur toutes les informations disponibles, en effet les filtres prédictifs

trouvent récursivement $p(X_k|Y_{1:k})$ à partir de $p(X_{k-1}|Y_{1:k-1})$ et Y_k . Ceci est fait en deux étapes. L'étape de prédiction et l'étape de correction.

Assumons que $p(X_{k-1}|Y_{1:k-1})$ est connue à l'itération $k - 1$.

Prédiction :

$$p(X_k|Y_{1:k}) = \int p(X_k|X_{k-1}) p(X_{k-1}|Y_{1:k-1}) dX_{k-1} \quad (\text{II.4})$$

Correction :

Par application directe de la règle de Bayes, on obtient :

$$p(X_k|Y_{1:k}) = \frac{p(Y_k|X_k)p(X_k|Y_{1:k-1})}{p(Y_k|Y_{1:k-1})} \quad (\text{II.5})$$

Avec $p(Y_k|Y_{1:k-1})$ est une constante de normalisation donnée par :

$$p(Y_k|Y_{1:k-1}) = \int p(Y_k|X_k) p(X_k|Y_{1:k-1}) dX_k \quad (\text{II.6})$$

Les équations récursives Bayésiennes Equation II-4 et Equation II-5 sont résolues différemment par les filtres prédictifs, sous différentes hypothèses. Le filtre de Kalman et le filtre H^∞ introduit respectivement dans la section II.2.2 et II.2.3 ont une solution algébrique fermée, ils assument que le système d'état est d'observation sont linéaire, en plus le filtre de Kalman nécessite une connaissance préalable de la covariance des incertitudes supposées être gaussiennes alors que le filtre H^∞ ne fait aucune hypothèse sur les perturbations et minimise l'erreur d'estimation du pire cas. Dans un autre côté le filtre particulière introduit dans la section II.2.4, qui est une technique d'échantillonnage Bayésienne permet d'estimer la distribution a posteriori dans des environnements non linéaires, non gaussiens. Le diagramme de base des étapes du filtre Bayésien est représenté par la Figure II-3.

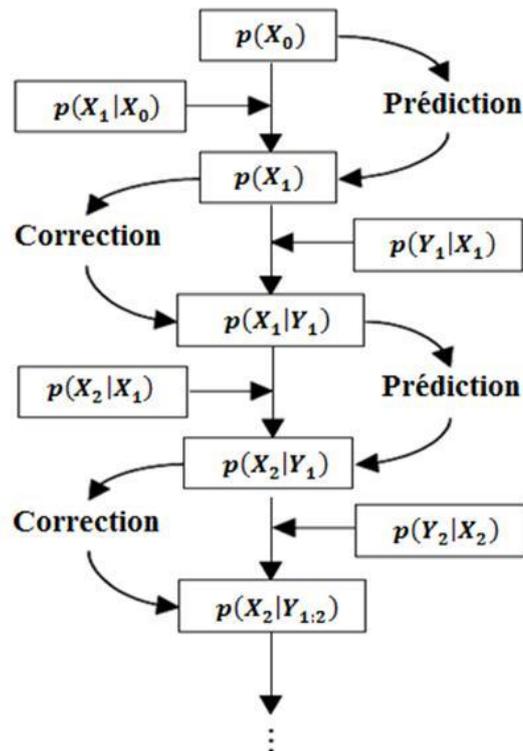


Figure II-3. Diagramme du filtre bayésien

II.2.2. Le filtre de Kalman

Le filtre de Kalman a eu un impact considérable sur l'estimation linéaire en raison de son adaptabilité à la mise en œuvre sur un ordinateur numérique pour l'estimation en ligne et l'utilité de l'approche espace d'état. Aujourd'hui, le filtre de Kalman est largement utilisé dans les domaines suivants: navigation, guidage, contrôle d'attitude, détermination de l'orbite des satellites, radar, traitement des signaux sonar et biomédicaux, etc. Le filtre de Kalman est également connu sous le nom d'estimation quadratique linéaire (LQE). C'est une méthode de filtrage récursif basée sur la théorie de l'estimation statistique, utilisant l'erreur quadratique moyenne minimale. Le filtre de Kalman est implémenté sous l'hypothèse que le modèle du système et les statistiques de bruit sont connus et que les perturbations liées à l'état et à la mesure sont considérées comme du bruit gaussien stationnaire.

Considérons le système dynamique représenté par Equation II-1 et Equation II-2. Pour chaque itération, le filtre de Kalman effectue trois tâches: la prédiction, la pondération et la correction. Comme il s'agit d'un processus récursif, le filtre de Kalman ne garde que la trace de la distribution de l'état actuel X_k . Le diagramme de base des étapes du filtre de Kalman est représenté par la Figure II-4.

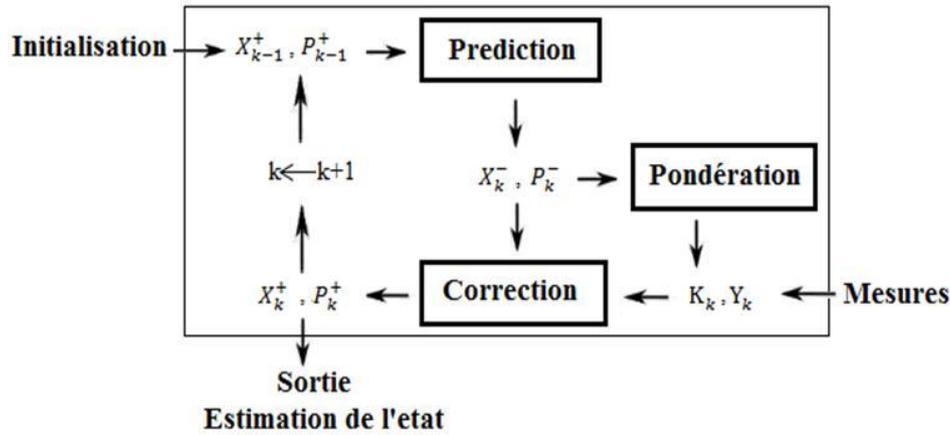


Figure II-4. Diagramme du filtre de Kalman

L'étape de prédiction Equation II-6 et Equation II-7, le filtre de Kalman utilise la dynamique du système, dans Equation II-1, pour calculer l'estimation a priori de l'état X_k^- et l'estimation a priori de la covariance de l'erreur P_k^- en utilisant respectivement l'estimation a posteriori X_{k-1}^+ et la covariance d'estimation a posteriori P_{k-1}^+ .

$$X_k^- = AX_{k-1}^+, \quad (\text{II.7})$$

$$X_k^- = AX_{k-1}^+, P_k^- = AP_{k-1}^+A^t + Q, \quad (\text{II.8})$$

Avec Q est la matrice de covariance du bruit du système.

L'étape de pondération calcule le gain de Kalman K à chaque itération Equation II-8. Pour ce faire, cette étape n'utilise que la matrice de covariance P_k^- de la prédiction, la matrice covariance de bruit de mesure R , et le modèle mathématique H de l'observation. En effet à chaque itération, le filtre de Kalman reçoit en entrée un vecteur d'observation Y_k . Ce vecteur est une mesure indirecte de l'état inconnu du système, modélisé par Equation II-2. L'observation a une incertitude, représentée par le bruit additif v_k , qui est une variable aléatoire gaussienne.

$$K_k = P_k^- H^t [HP_k^- H^t + R]^{-1} \quad (\text{II.9})$$

Le gain de Kalman K a deux rôles: premièrement, il projette des valeurs à partir de leur espace d'état dans l'espace des paramètres de mesures. Deuxièmement, il

décide combien doit être utilisée la différence entre l'observation réelle et l'observation prédite.

L'étape de correction du filtre de Kalman, la matrice de pondération K est utilisée pour combiner de manière optimale la prédiction et l'observation. Dans cette étape l'estimation de l'état a posteriori X_k^+ Equation II-9, et l'estimation a posteriori de la covariance de l'erreur P_k^+ Equation II-10, sont calculés.

$$X_k^+ = X_k^- K [Y_k - H X_k^-], \quad (\text{II.10})$$

$$P_k^+ = [I - KH] P_k^- \quad (\text{II.11})$$

L'ensemble des étapes et les calculs du filtre de Kalman sont donnés par l'Algorithme II-1

Algorithme II.1 : Filtre de Kalman

Initialisation :

L'état a posteriori initiale X_0^+ et la covariance a posteriori associée P_0^+ sont initialisés.

Itérations :

-Prédiction : Utilise la dynamique du système pour calculer l'estimation a priori de l'état X_k^- et l'estimation a priori de la covariance de l'erreur P_k^- .

$$X_k^- = A X_{k-1}^+,$$

$$P_k^- = A P_{k-1}^+ A^t + Q.$$

-Pondération : Calcule le gain de Kalman K_k

$$K_k = P_k^- H^t [H P_k^- H^t + R]^{-1}.$$

-Correction : Calculer l'estimation de l'état a posteriori X_k^+ et l'estimation a posteriori de la covariance de l'erreur P_k^+ .

$$X_k^+ = X_k^- K [Y_k - H X_k^-],$$

$$P_k^+ = [I - KH] P_k^-.$$

11.2.3. Le filtre H_∞

Le filtre de Kalman est largement utilisé pour estimer l'état des systèmes, il est considéré comme étant un outil efficace, en effet la robustesse du filtre de Kalman est basée sur la connaissance préalable des covariances associées à l'état et aux mesures, cependant ces informations a priori sont indisponibles pour plusieurs cas réels. D'autre part, le filtre H^∞ donne des limites supérieures strictes aux erreurs d'estimation, quelles que soient les perturbations, tant qu'elles sont d'énergie finie. Le filtrage H^∞ est appliqué dans le contrôle linéaire des systèmes automatiques, et également dans l'estimation des états des systèmes, en effet aucune supposition concernant les perturbations du système, il a la caractéristique de considérer l'erreur d'estimation du cas le plus défavorable, seul l'état antérieure et la mesure de l'état actuel sont requis mais aucun historique des observations n'est nécessaire. Par conséquent, aucune exigence pour une capacité élevée de stockage informatique. Dans le réglage du filtre H^∞ , le signal d'entrée exogène est supposé être à énergie limité plutôt que gaussien, le filtre H^∞ est conçu de telle sorte que la norme H^∞ du système, qui reflète le gain le plus défavorable du système, est minimisée. Le diagramme de base des étapes du filtre H^∞ est représenté par la Figure II-5.

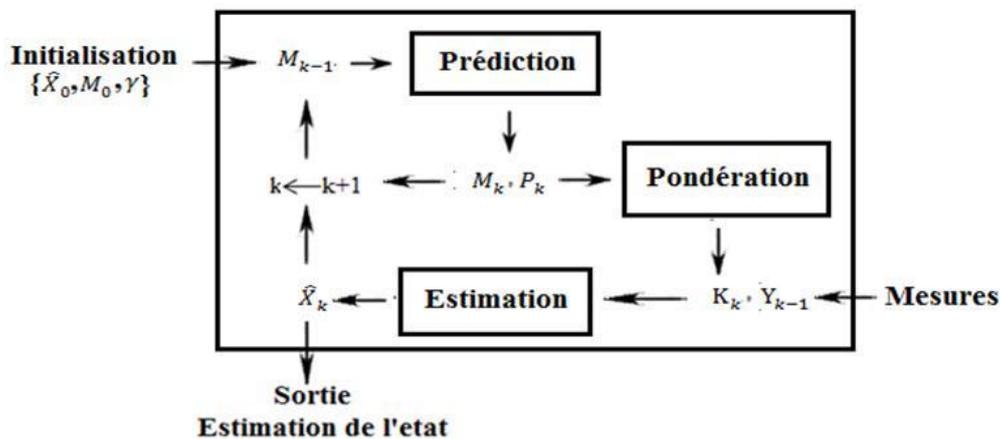


Figure II-5. Diagramme du filtre H^∞

Considérons le système dynamique représenté par Equation II-1 à Equation II-3.

Considérons le signal en temps discret $\{Y_0, Y_1 \dots Y_N\}$ assumé être à énergie limitée tel que :

$$\|Y\|_{2,[0,N]} < \infty \quad (\text{II.12})$$

Nous considérons un filtre basé sur un observateur de la forme :

$$\hat{X}_{k+1} = A\hat{X}_k + K_k(Y_k - H\hat{X}_k) \quad (\text{II.13})$$

Où \hat{X}_{k+1} est l'estimation d'état du système, K_k est le gain du filtre H^∞ .

Le problème de filtrage H^∞ peut être interprété comme un problème minimax où l'estimation est déterminée. Étant donné un scalaire $\gamma > 0$, le filtrage H^∞ a priori consiste à trouver une estimation \hat{Z} de Z basée sur les mesures $\{Y_0, Y_1 \dots Y_{k-1}\}$ tel que pour tout $w \in l_2[0, N-1]$, $v \in l_2[0, N-1]$, et $X_0 \in R^n$, la fonction J coût vérifie :

$$J = \|\hat{Z} - Z\|_{2,[1,N]}^2 - \gamma^2 \left\{ (\hat{X}_0 - X_0)^t P_0^{-1} (\hat{X}_0 - X_0) + \|w\|_{2,[0,N-1]}^2 + \|v\|_{2,[0,N-1]}^2 \right\} < 0 \quad (\text{II.14})$$

Où P_{X_0} est une matrice de covariance définie positive reflétant l'incertitude de l'estimation de l'état initial X_0 et son estimation \hat{X}_0 .

Ce problème peut être résolu via l'approche de l'équation de Riccati [Nag, 1991], le théorème suivant donne une solution au filtrage H^∞ a priori [Fra, 2007]:

Théorème II.1

Considérons le système de Equation II-1 à Equation II-3. Pour un $\gamma > 0$ donné, il existe un filtre H^∞ a priori $[0, N]$ si et seulement s'il existe une solution $P_k = P_k^t > 0$, $k=0,1,\dots,N$ pour l'équation de Riccati suivante :

$$P_{k+1} = AP_k A^t + \gamma^{-2} P_{k+1} H^t (I + \gamma^{-2} H P_k H^t)^{-1} H P_{k+1} - AP_k H^t (I + H P_k H^t)^{-1} H P_k A^t + G G^t \quad (\text{II.15})$$

Un filtre H^∞ a priori adapté peut être donné par Equation II-4 avec le gain K_k :

$$K_k = AP_k H (I + H P_k H^t)^{-1} \quad (\text{II.16})$$

Equation II-14 est équivalente à l'équation de Riccati suivante :

$$M_{k+1} = AM_k A^t - AM_k \bar{H}^t (\bar{R} + \bar{H} M_k \bar{H}^t)^{-1} \bar{H} M_k A^t + G G^t \quad (\text{II.17})$$

Avec

$$P_k = (M_k^{-1} - \gamma^{-2} L^t L)^{-1} > 0 \quad (\text{II.18})$$

$$M_0 = (P_{X_0}^{-1} + \gamma^{-2}L^tL)^{-1} \quad (\text{II.19})$$

Les matrices \bar{H} et \bar{R} sont définies par : $\bar{H} = \begin{bmatrix} H \\ L \end{bmatrix}$; $\bar{R} = \begin{bmatrix} I & 0 \\ 0 & \gamma^{-2}I \end{bmatrix}$

Ainsi, la matrice M_k est calculée via l'équation récursive Equation II-17, puis la matrice P_k est calculée par Equation II-18, par la suite le gain du filtre H^∞ a priori est calculé par Equation II-16.

L'essentiel des étapes du filtre H^∞ sont donnés par Algorithme II-2.

Algorithme II.2 : Filtre H^∞

Initialisation :

Initialiser le vecteur d'état \hat{X}_0 .

Initialiser la matrice M_0 .

Initialiser γ .

Itérations :

-Prédiction : Utilise l'équation de Riccati équivalente pour calculer la prédiction la covariance de l'erreur P_k .

$$M_k = AM_{k-1}A^t - AM_{k-1}\bar{H}^t(\bar{R} + \bar{H}M_{k-1}\bar{H}^t)^{-1}\bar{H}M_{k-1}A^t + GG^t$$

$$P_k = (M_k^{-1} - \gamma^{-2}L^tL)^{-1}$$

-Pondération : Calcule le gain H^∞ K_k

$$K_k = AP_k H^t [HP_k H^t + I]^{-1}.$$

-Estimation : Calculer l'estimation du vecteur d'état \hat{X}_k .

$$\hat{X}_k = A\hat{X}_{k-1} + K_k(Y_{k-1} - H\hat{X}_{k-1}).$$

II.2.4. Le Filtre à Particules

Le filtre à particules a été inventé pour implémenter numériquement l'estimation bayésienne, il approxime récursivement la distribution postérieure en utilisant un ensemble fini d'échantillons ou de particules pondérés. Il est connu sous le nom de Monte Carlo séquentiel, les caractéristiques du filtre à particules proviennent de ses propriétés de maintien de plusieurs hypothèses en présence de bruits non gaussiens et de modèle dynamiques ou d'observation non linéaires. Dans ce contexte, la densité postérieure est approchée par un ensemble de particules pondérées. Le fonctionnement du filtre à particules est basé sur un partitionnement de l'espace d'état en autant de parties, lesquelles des particules sont associées selon une certaine mesure de probabilité. Plus la probabilité est élevée, plus les particules sont denses. Comme le nombre de particules devient très grand, cette caractérisation d'échantillonnage séquentiel devient une représentation équivalente à la description fonctionnelle habituelle de la fonction de densité de probabilité postérieure, et le filtre se rapproche de l'estimation bayésienne optimale.

Par la suite nous présentons ci-dessous l'algorithme dit d'échantillonnage séquentiel pour le filtre à particules. Comme décrit en détail dans les références [San, 2002]. L'algorithme utilise une densité d'importance proposée pour représenter une densité qui ne peut pas être exactement calculée, à savoir la densité postérieure recherchée dans le cas présent. Ensuite, des échantillons sont tirés de la densité d'importance au lieu de la densité réelle. Le diagramme de base des étapes du filtrage à particules est présenté par Figure II-6.

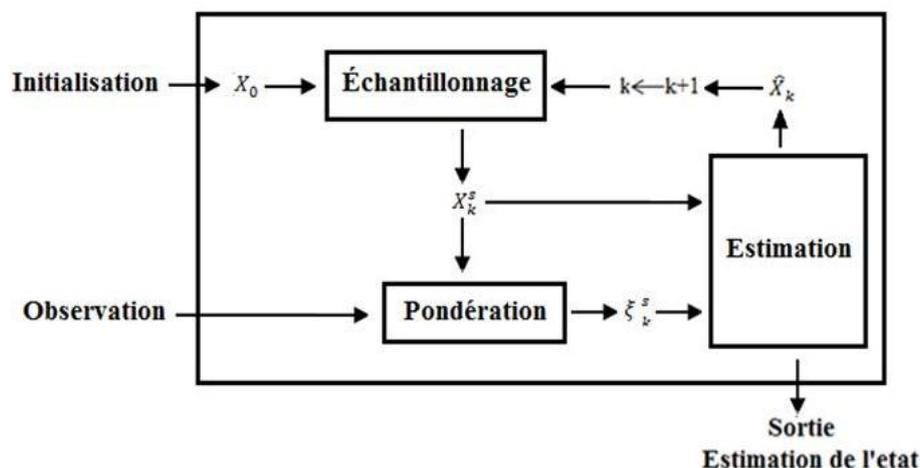


Figure II-6. Diagramme du filtre à particule

Où $\delta(\cdot)$ est la fonction delta de Dirac. Cette représentation de la densité postérieure facilite la phase de mise à jour en utilisant la règle de Bayes Equation II-5.

$$p(X_{0:k}|Y_{1:k}) = \frac{p(Y_k|Y_{1:k})p(X_{0:k}|Y_{1:k-1})}{p(Y_k|Y_{1:k-1})} \quad (\text{II.20})$$

Et la densité filtrée postérieure peut être approximée par :

$$p(X_k|Y_{1:k}) \approx \sum_{s=1}^{NP} \xi_k^s \delta(X_k - X_k^s) \quad (\text{II.21})$$

Lorsque $NP \rightarrow \infty$, l'approximation Equation II-21 approche la vraie densité postérieure $p(X_k|Y_{1:k})$.

L'essentiel des étapes du filtre à particules sont donnés par Algorithme II-3.

Algorithme II.3 : Filtre à particules

Initialisation :

Générer aléatoirement NP particules $\{X_0^s, s = 1, \dots, NP\}$

Initialiser les poids $\{\xi_0^s, s = 1, \dots, NP\}$ associés aux particules.

Échantillonnage séquentiel :

Propager chaque échantillon selon le modèle dynamique.

Calculer la distribution de densité de l'observation pour chaque particule.

Affecter les poids normalisé à chaque particule

Estimation

L'état est mis à jour par la moyenne pondérée des particules.

II.3. Filtre H_∞ PF pour le suivi visuel

Dans cette section nous allons proposer une méthode de prédiction Figure II-7 basée sur le filtre H_∞ et le filtre à particules pour estimer l'état dynamique dans le cadre du suivi visuel, a ce stade de notre thèse nous allons s'intéresser essentiellement à prouver l'efficacité de la méthode de prédiction proposée en

masquant l'étape de la détection locale de la cible basé sur le modèle d'apparence qu'on va étudier amplement dans le chapitre suivant.

Sous l'hypothèse que le comportement dynamique globale de l'objet en mouvement est linéaire tandis que son comportement locale est non linéaire, par conséquent notre approche de prédiction est divisée en deux modules, module globale basé sur le filtre H^∞ permet d'estimer l'état dynamique globale de la cible tandis que le module local utilise le filtre particulaire pour l'état local, vu qu'il est plus précis et peut effectivement gérer les cas non linéaire et non-gaussien.

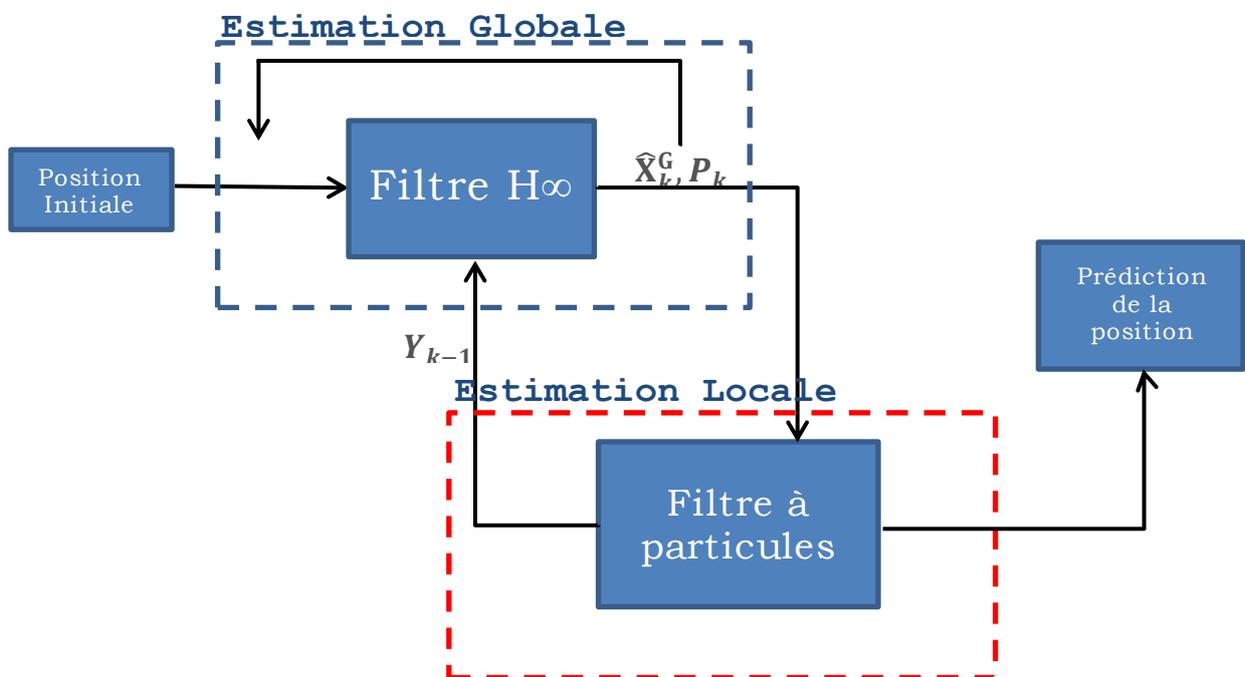


Figure II-7. Diagramme du filtre H^∞ PF

Le filtre de Kalman a été largement appliqué dans différentes applications de prédiction en particulier dans le cadre du suivi visuel, c'est une méthode d'estimation optimale pour le cas de mouvement gaussien linéaire, en effet l'optimalité du filtre de Kalman repose sur la connaissance de la covariance du processus et du bruit de mesure, ce type d'information a priori n'est pas toujours disponible.

Dans ce contexte le filtre H^∞ est une alternative intéressante au célèbre filtre de Kalman dans la plupart des problèmes d'estimation, en effet le filtre H^∞ ne fait aucune hypothèse sur les perturbations et minimise l'erreur d'estimation du pire

cas. Notre stratégie vise à incorporer le filtre H^∞ dans l'algorithme filtre à particules pour estimer l'état dynamique dans le cas du suivi visuel. L'utilisation du filtre H^∞ permet d'atténuer le problème d'appauvrissement des particules surtout dans le cas où la cible a un comportement dynamique complexe ou de manœuvre. En effet, le filtre H^∞ peut pleinement prendre en compte les mesures actuelles permettant ainsi au filtre particulaire de générer de nouvelles particules, en tirant parti des mesures courantes pour prédire effectivement l'état du système. L'essentiel des étapes du filtre H^∞ PF est présenté par Algorithme II-4.

Algorithme II.4 : Filtre H^∞ PF

Initialisation :

Initialiser le vecteur d'état \hat{X}_0^G .

Initialiser la matrice M_0 .

Initialiser γ .

Itérations :**-Prédiction globale**

Calculer la covariance de l'erreur P_k .

Calculer l'estimation du vecteur d'état \hat{X}_k^G .

-Prédiction locale

Générer aléatoirement NP particules $\{X_k^s, s = 1, \dots, NP\}$ à partir de l'estimation globale et \hat{X}_k^G la covariance de l'erreur P_k .

Calculer la distribution de densité de l'observation pour chaque particule.

Affecter les poids normalisés à chaque particule ξ_k^s

-Estimation

L'état est mis à jour par la moyenne des particules : $\hat{X}_k = \sum_{s=1}^{NP} \xi_k^s X_k^s$

II.4. Evaluation

Pour évaluer les performances de la méthode de prédiction proposée pour le suivi, nous avons mené une série d'expérimentations et de comparaisons, deux cas sont considérés, la première expérience est réalisée sur une cible qui se déplace uniformément sans manœuvre alors que la deuxième expérience concerne une cible qui réalise un mouvement de manœuvre. Nous allons considérer deux métriques, l'erreur de localisation du centre (ELC) donnée pour chaque pas de temps k par la formule suivante:

$$ELC_k = \sqrt{(x_r^k - x_e^k)^2 + (y_r^k - y_e^k)^2} \quad (\text{II.22})$$

Où x_r^k, y_r^k sont respectivement les coordonnées x et y à l'instant k de la vérité terrain. Alors que x_e^k, y_e^k représente la position estimée à l'instant k .

Pour une meilleure cohérence les expérimentations sont répétées vingt fois pour les différentes méthodes de comparaison à savoir PF, KF, H ∞ F, KPF et H ∞ PF. Ainsi l'erreur de localisation moyenne normalisée (ELMN) donnée par Equation II-24 est calculée.

$$ELMN = \frac{1}{20} \sum_{n=1}^{20} \frac{\sum_k ELC_k}{\sum_k} \quad (\text{II.23})$$

Dans les deux expérimentations $\gamma = 4$ pour les filtres H ∞ et H ∞ PF, pour les filtres KPF et H ∞ PF le nombre de particules est fixé à 60 alors que pour PF il est fixé à 200.

Dans la première partie de cette évaluation la cible se déplace uniformément avec une vitesse constante, le trajet réel réalisé est donné par la Figure II-8. Le système dynamique utilisé par les différentes méthodes de comparaison est le modèle à vitesse quasi constante (NCV) [Ron, 2004]. La matrice de covariance P est initialisée à I_4 , la covariance du bruit système $Q = 0.01 * I_4$, et la covariance du bruit de mesure $R = 2 * I_2$.

La Figure II-9 présente le trajet réel ainsi que les trajets estimés par les différentes méthodes de comparaison. Tous les algorithmes arrivent à suivre la cible, néanmoins les méthodes KPF et H ∞ PF réalise un suivi plus précis avec une

meilleure concession pour notre méthode malgré le nombre de particules moins petit par rapport au PF. Ce résultat est justifié de façon plus crédible par le calcul de ELC, la comparaison de l'erreur de localisation du centre à chaque pas de temps k est affichée par la Figure II-10, dans l'ensemble $H_{\infty}PF$ réalise une erreur de localisation plus petite.

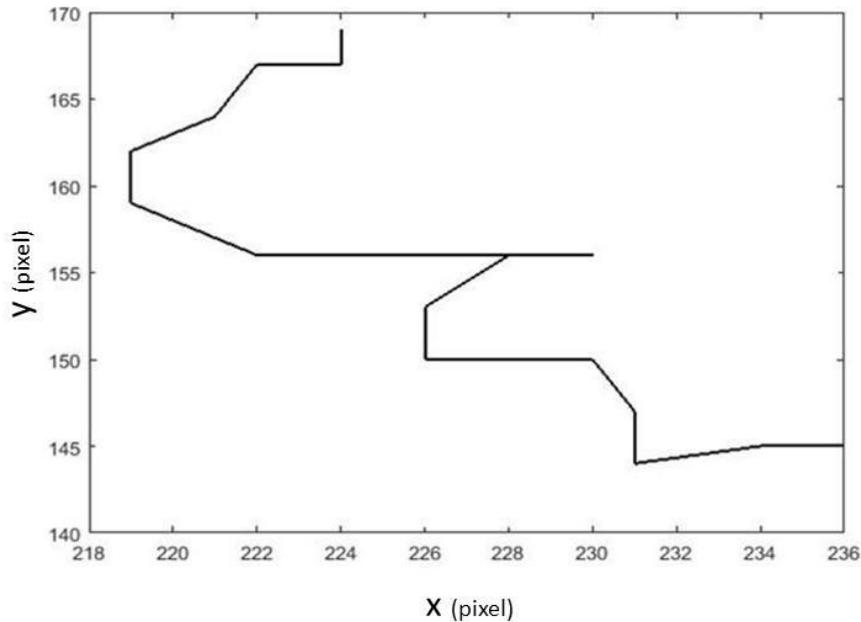


Figure II-8. Trajet réel

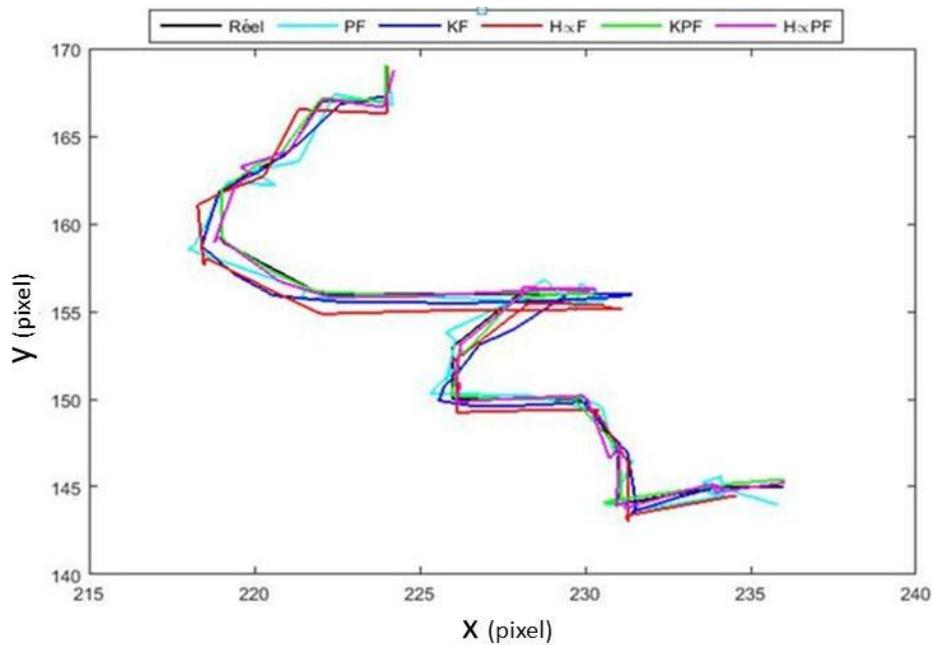


Figure II-9. Trajectoires estimées par les méthodes de comparaison et vérité du terrain

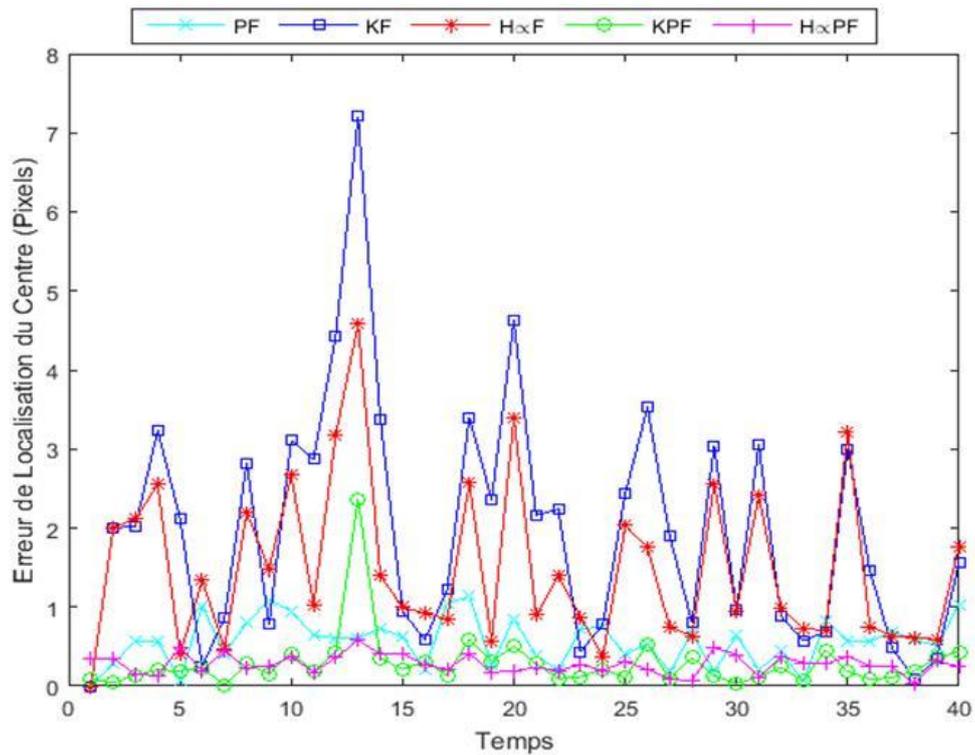


Figure II-10. Comparaison de l'erreur de localisation du centre (en pixels)

Pour analyser l'effet du nombre de particules, on a calculé l'erreur de localisation moyenne normalisée pour les méthodes PF, KPF et H_{∞} PF en fonction du nombre de particules. Le résultat de cette expérience est affiché par la Figure II-11, il est évident que la méthode PF est imprécise pour un nombre de particule petit alors que les méthodes KPF et H_{∞} PF sont plus précis, même lorsque le nombre de particules est petit.

On va terminer cette première partie d'expérimentation par calculer ELC ainsi que ELMN pour différentes valeurs de γ . La Figure II-12 présente ELC pour les méthodes PF, KPF et H_{∞} PF pour différentes valeurs de γ , les résultats obtenu montrent que pour toutes les valeurs de γ la méthode H_{∞} PF est la plus précise. La Figure II-13 affiche L'ELMN, il est bien claire que la méthode H_{∞} PF réalise le plus petit ELMN pour les différentes valeurs de γ en particulier quand $\gamma = 4$.

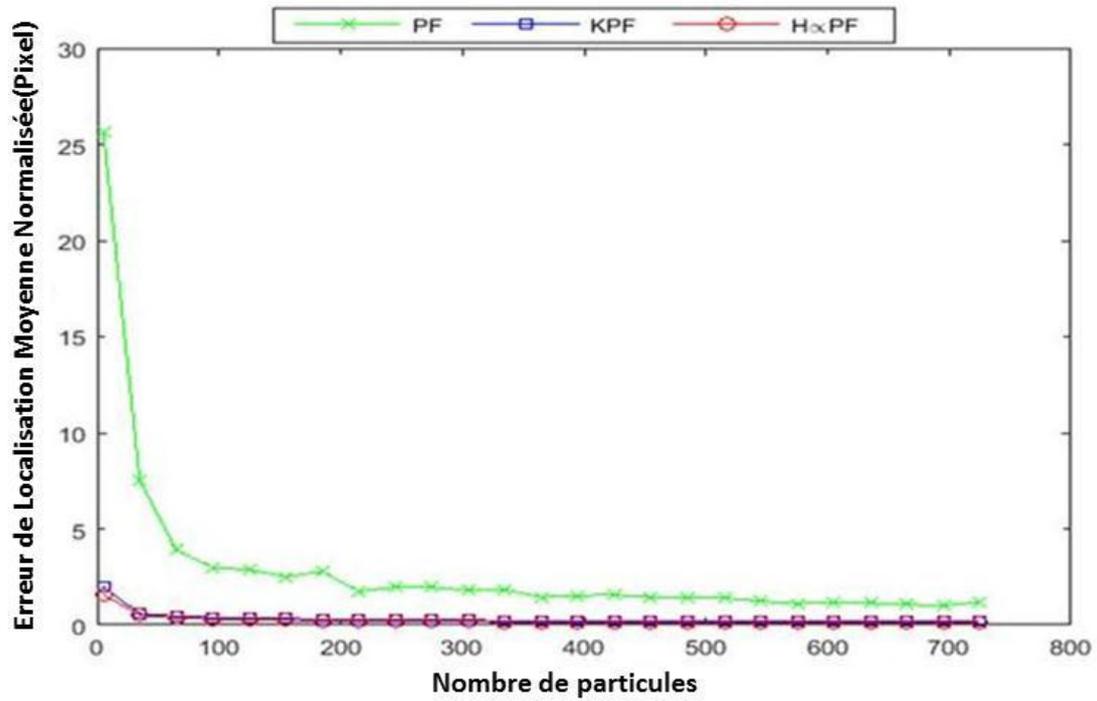


Figure II-11. Comparaison de l'Erreur de Localisation Moyenne Normalisée pour différents nombres de particules

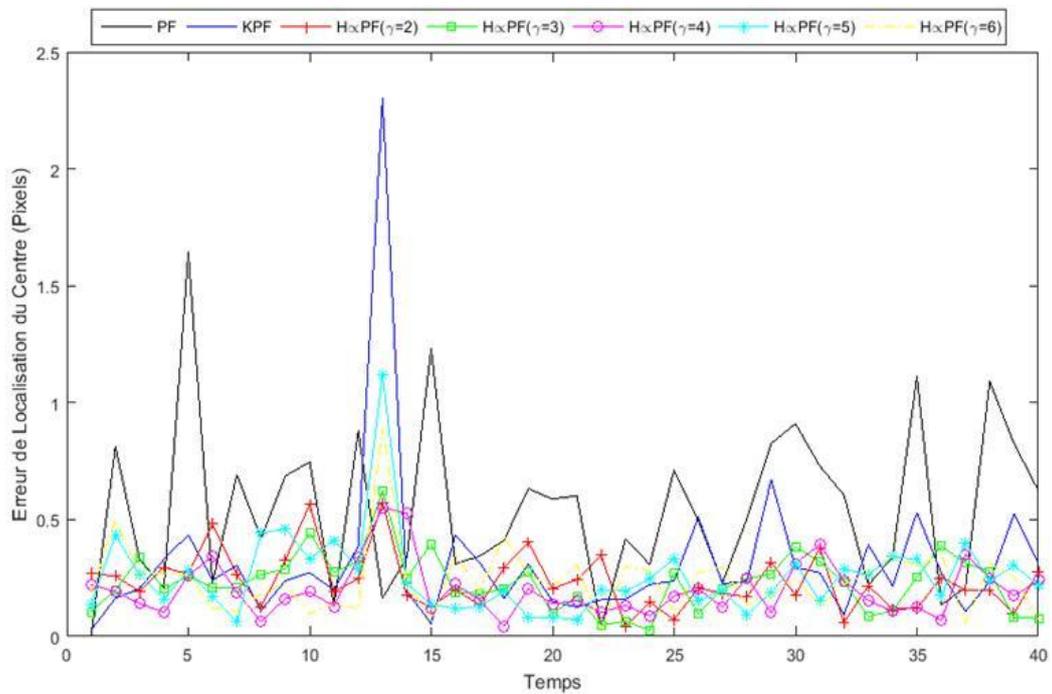


Figure II-12. Comparaison de l'Erreur de Localisation du centre pour différentes valeurs de γ

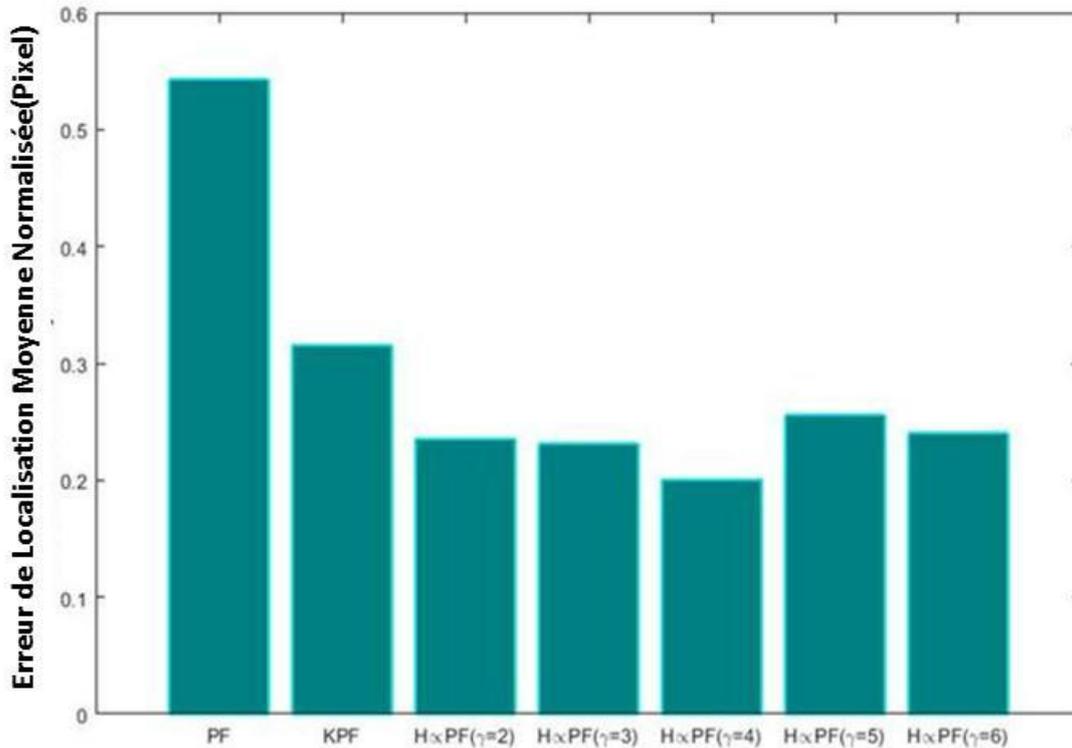


Figure II-13. Comparaison de l'Erreur de Localisation Moyenne Normalisée différentes valeurs de γ

Dans la deuxième partie de l'évaluation, la cible se déplace avec une vitesse variable et en changeant de direction de façon rapide réalisant ainsi un mouvement de manœuvre, la Figure II-14 affiche le trajet réel. Le système dynamique utilisé par les méthodes de comparaison est le modèle à accélération quasi constante (NCA) [Ron, 2004]. La matrice de covariance \mathbf{P} est initialisée à \mathbf{I}_6 , la covariance du bruit système $\mathbf{Q} = 0.5 * \mathbf{I}_6$, et la covariance du bruit de mesure $\mathbf{R} = 2 * \mathbf{I}_2$.

Les trajets estimés par les différentes méthodes de prédiction sont présentés par la Figure II-15. On peut voir que le trajet estimé par la méthode $H_{\infty}PF$ est très proche à celui de la vérité terrain réalisant ainsi un suivi le plus précis. En plus à partir des résultats affichés par la Figure II-16, $H_{\infty}PF$ marque dans l'ensemble une erreur de localisation meilleure que les méthodes $H_{\infty}KPF$, alors que les méthodes KF et PF ont marqué une erreur de localisation plus élevée.

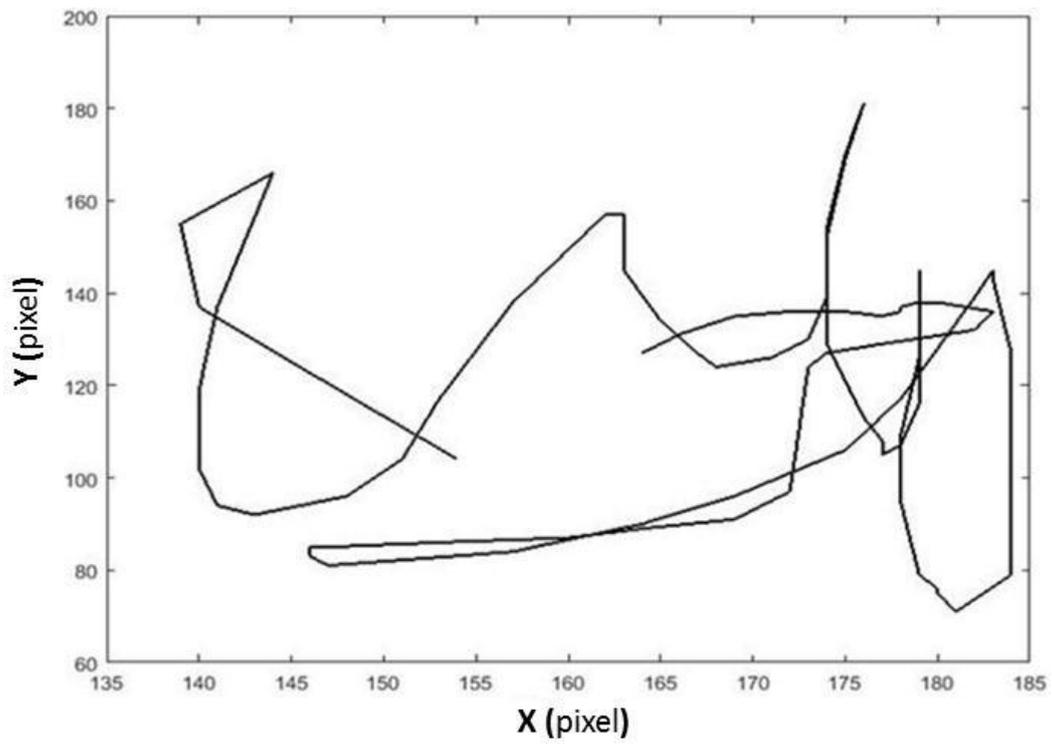


Figure II-14. Trajet réel

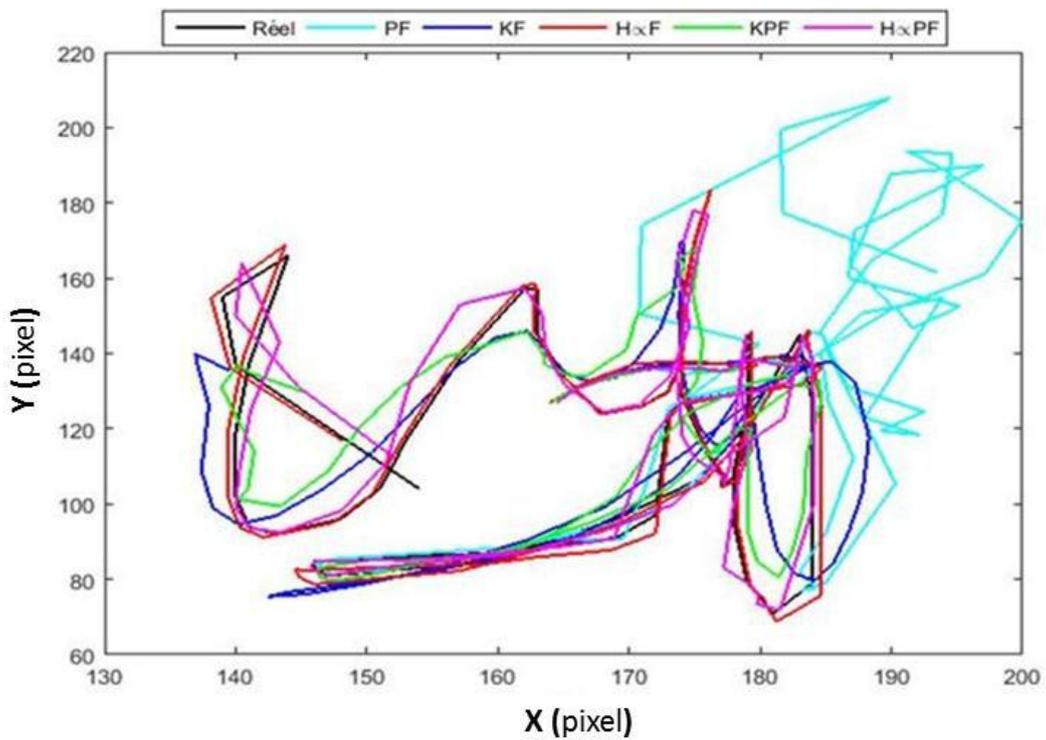


Figure II-15. Trajectoires estimées par les méthodes de comparaison et vérité du terrain

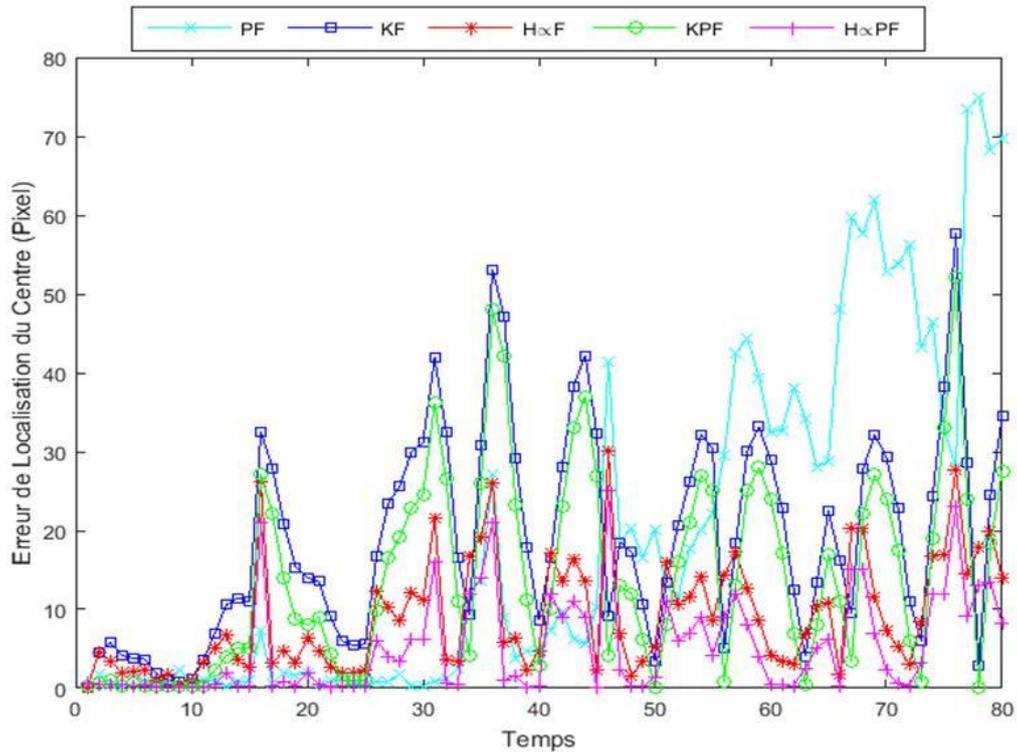


Figure II-16. Comparaison de l'erreur de localisation du centre (en pixels)

La Figure II-17 affiche ELMN pour les méthodes PF, KPF et H_{∞} PF pour différents nombres de particules, il est évident que H_{∞} PF réalise une erreur de localisation moyenne faible pour les différents nombres de particules, tandis que la méthode PF montre des faiblesses quand le nombre de particules est petit.

Nous terminons cette section d'évaluation par la présentation des résultats de suivi obtenus pour différentes valeurs de γ . La Figure II-18 affiche ELC, la méthode H_{∞} PF est meilleure en terme de ELC par rapport aux méthodes PF et KPF pour différentes valeurs de γ . En plus les résultats affichés par la Figure II-19 montrent de bonne performance en terme de ELMN de la méthode H_{∞} PF pour toutes les valeurs de γ en particulier pour $\gamma = 4$.

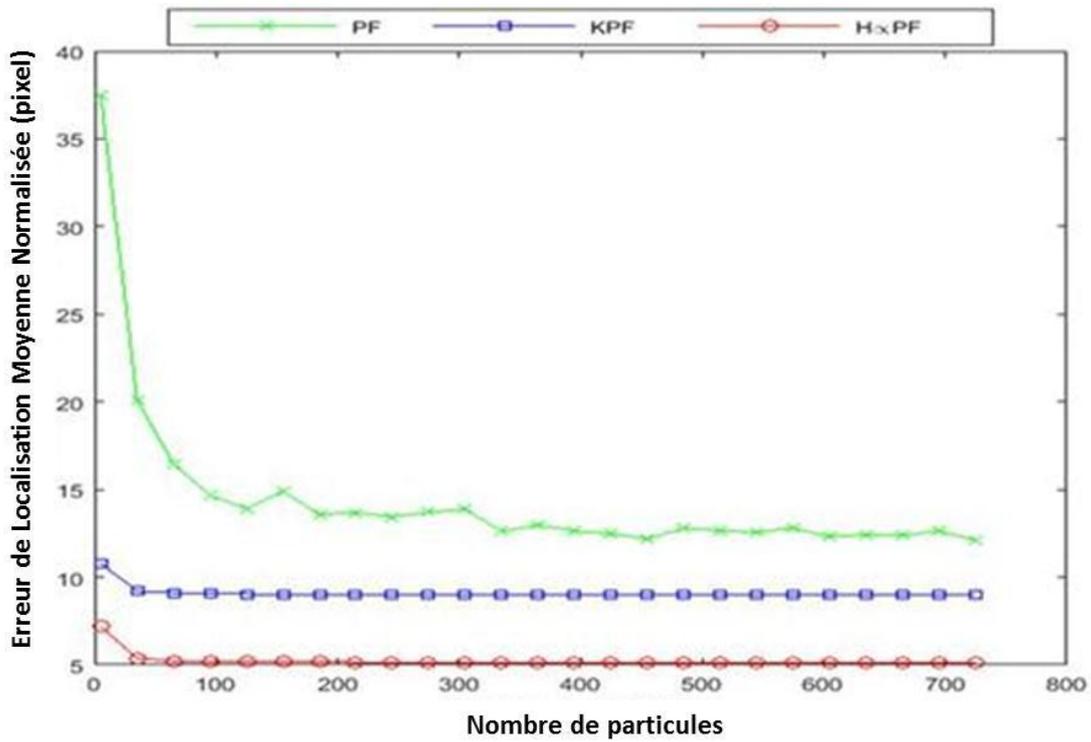


Figure II-17. Comparaison de l'Erreur de Localisation Moyenne Normalisée pour différents nombres de particules

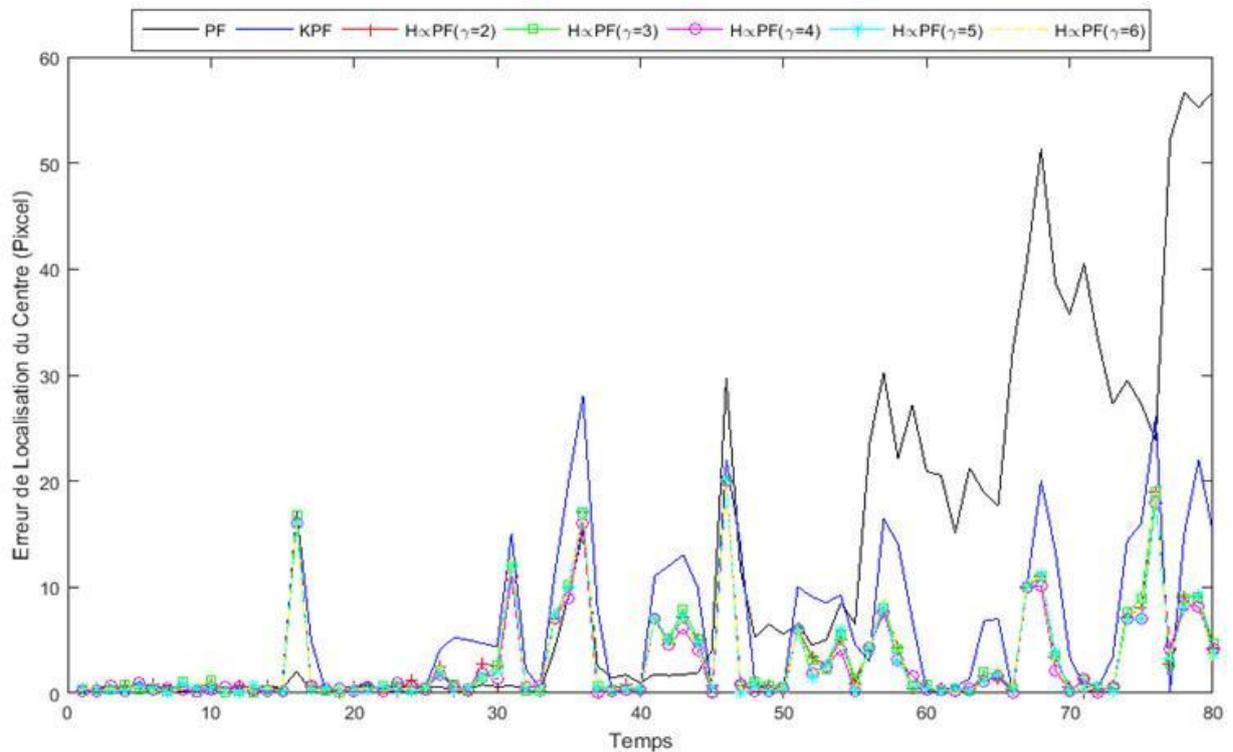


Figure II-18. Comparaison de l'Erreur de Localisation du centre pour différentes valeurs de γ

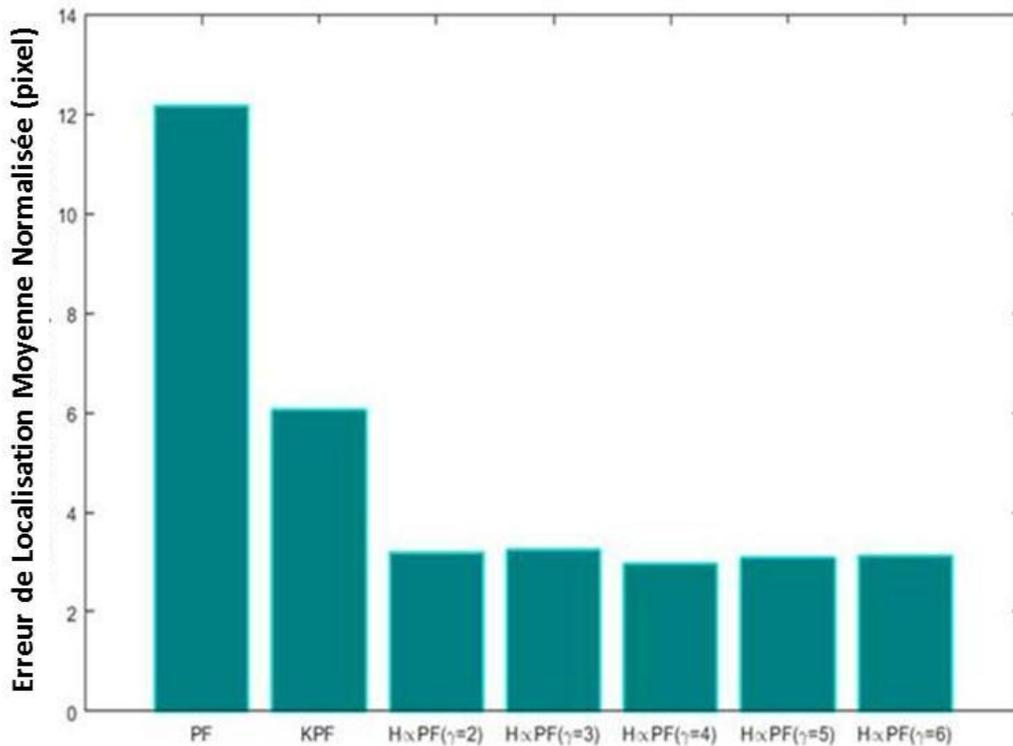


Figure II-19. Comparaison de l'Erreur de Localisation Moyenne Normalisée différentes valeurs de γ

II.5. Conclusion

Dans ce chapitre nous avons présenté l'essentiel de la théorie en rapport avec le filtrage Bayésien, pour les problèmes d'estimations d'état des systèmes sous différentes hypothèse à savoir modèles linéaires-gaussiens et modèles non-linéaires et / ou non-gaussiens. Trois filtres de prédiction ont été examiné à savoir le filtre de Kalman, le filtre H^∞ et le filtre à particules, dont chacun traitant différents types de modèles, la théorie de base et l'algorithme de chaque filtre a été présenter, finalement nous avons proposé une méthode de prédiction basé sur le filtre H^∞ et le filtre particulaire, la méthode proposée est adopté pour estimer l'état de la cible dans le cadre du suivi visuel. Une série d'expérimentations sont réalisées. D'après les résultats des évaluations, le filtre H^∞ -PF s'avère efficace et pratique pour suivre la cible dans différentes conditions dynamiques.

Chapitre III : Modélisation de l'apparence de l'objet suivi

La modélisation de l'apparence de la cible constitue une étape essentielle pour le suivi visuel, un modèle d'apparence faible conduit à un suivi non stable et incertain surtout dans des séquences sous des conditions réelles. Ce chapitre présente l'approche proposée qui consiste à fusionner les caractéristiques de couleur et de texture pour créer un modèle d'apparence robuste. Dans la première partie, nous passons en revue les différentes représentations de l'objet suivi : géométriques, par contour et par les points locaux. Dans la deuxième partie nous présentons notre approche ; ainsi nous décrivons d'abord la modélisation de la distribution des caractéristiques de couleur par α -mélange de gaussiennes pour les composants H et S de l'espace HSV suivie de la modélisation de la distribution de la caractéristique texture par l'histogramme distinctif basé sur l'opérateur ULBP sur la composante V. A la fin nous présentons la mise à jour adaptative du modèle par une pondération des deux descripteurs résultant de cette modélisation dans le cadre du filtre à particules.

CHAPITRE III : MODELISATION DE L'APPARENCE DE L'OBJET SUIVI	81
III.1. INTRODUCTION	82
III.2. REPRESENTATION DE L'OBJET SUIVI	83
III.2.1. GEOMETRIQUE	84
III.2.2. LE PROFIL DE CONTOUR	85
III.2.3. POINTS LOCAUX	86
III.3. MODELE D'APPARENCE BASE SUR LA COULEUR ET LA TEXTURE	87
III.3.1. HISTOGRAMME PONDERE BASE SUR LE MODELE DE MELANGE A-GAUSSIEN.....	87
III.3.1.1. Le modèle de mélange Gaussien.....	87
III.3.1.2. Le modèle d'apparence proposé.....	89
III.3.2. L'HISTOGRAMME DISTINCTIF BASE SUR LE MOTIF BINAIRE LOCAL UNIFORME.....	95
III.4. MISE A JOUR DU MODELE D'APPARENCE	103
III.5. CONCLUSION	104

III.1. Introduction

De nombreux problèmes rendent le suivi robuste des objets assez difficiles, notamment les capteurs de caméra de basse qualité, les conditions difficiles, les exigences de traitement en temps réel, le suivi des objets acquis par plusieurs caméras, les variations d'apparence des objets, l'occlusion complète, les perturbations ou bruit, la déformation ou la rotation de l'objet dans le plan et la variation de la forme. Face à ces problèmes, le choix du modèle d'observation est décisif pour un du suivi robuste. Ainsi, une vaste gamme de modèles d'apparence utilisant différentes représentations visuelles et les techniques de modélisation statistique a été proposée.

Le suivi visuel utilisant plusieurs types de caractéristiques visuelles (intensité, couleur, information spatio-temporelle, gradient et texture, etc.) s'avère être une approche robuste à cause des informations issues de ces caractéristiques qui sont complémentaires.

Dans ce cadre, de nombreux algorithmes de suivi [Xi, 2013] utilisent plusieurs caractéristiques pour augmenter la précision de la représentation par rapport aux variations d'apparence et pour améliorer la discrimination entre la cible et son arrière-plan. De nombreux chercheurs ont proposé diverses méthodes utilisant l'histogramme commun de couleur-texture [Fou, 2013] [Jin, 2018] [Saa, 2018]. Ce dernier est plus fiable que l'histogramme des couleurs dont l'utilisation induit quelques erreurs dans le suivi des objets. Il provoque un suivi incorrect voire échoue facilement lorsque la couleur de l'objet est proche de la couleur de fond ou lorsque l'objet se déplace rapidement ainsi dans le cas de flou de mouvement en plus, l'information spatiale de la cible est perdue. D'autre part, Les caractéristiques de texture introduisent de nouvelles informations par rapport à l'histogramme de couleur, elles reflètent une structure spatiale stable de l'objet, habituellement non sujet à l'impact de la lumière et de la couleur de fond.

Dans notre travail, nous choisissons une représentation visuelle basée sur plusieurs indices qui reflète les caractéristiques statistiques globales et locales de l'apparence de l'objet afin de mieux gérer les changements d'apparence. Cela permet de coder plus d'informations et ainsi améliorer la robustesse de la représentation visuelle. Les deux caractéristiques considérées sont la couleur et la texture et

l'espace de couleur adopté est HSV dont les composantes H et S sont étroitement liées aux informations de couleur alors que la composante V quantifie la majorité des informations de texture.

Le reste de ce chapitre est organisé de la manière suivante. La section III.2 présente brièvement les différentes représentations de l'objet suivi. La section III.3 présente notre méthode pour modéliser l'apparence de la cible. La stratégie de mise à jour du modèle d'apparence est exposée dans la section III.4. La section III.5 conclut le chapitre.

III.2. Représentation de l'objet suivi

En général, la cible est représentée par une zone de délimitation de la cible (Figure III-1), dont la forme est rectangulaire ou elliptique. Dans des applications plus spécifiques tels que le suivi des piétons ou les applications biomédicales, une représentation basée sur le contour est privilégiée. Lorsque l'objet suivi à une forme dynamique ou une petite taille ou lorsque la séquence présente des occlusions fréquentes, il est avantageux de suivre l'objet par les points locaux. Dans la suite de cette section nous examinons ces différentes représentations.

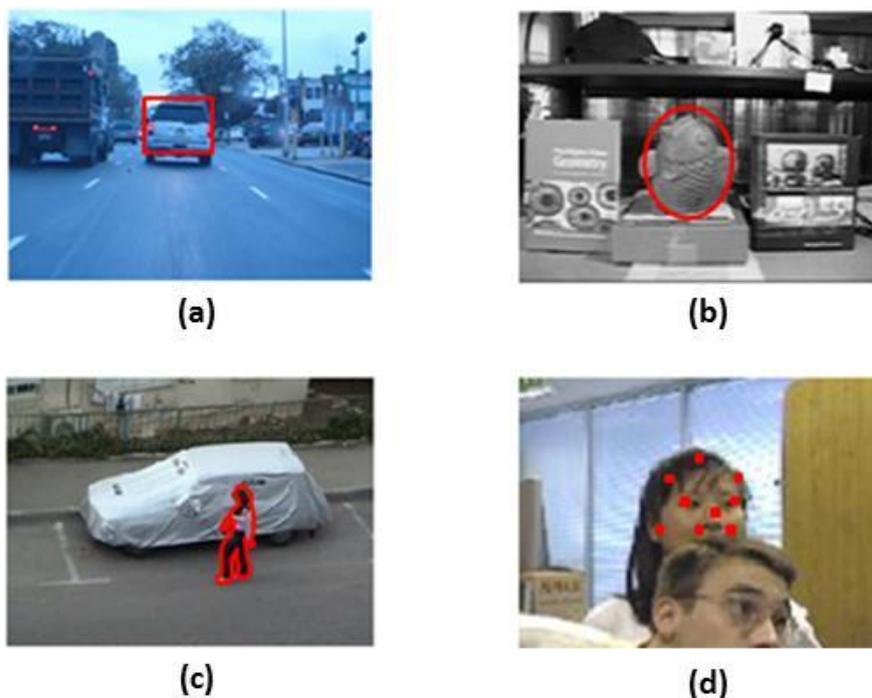


Figure III-1. Différentes représentations de la cible. (a) boîte, (b) ellipse, (c) contour, (d) point d'intérêt

III.2.1. Géométrie

Dans ce type de représentation, l'objet suivi est considéré comme un bloc d'image appelé modèle cible. La cible peut être exprimée par une forme géométrique simple généralement une forme rectangulaire ou elliptique, qui est prise comme modèle cible. Le modèle cible peut être extrait par une méthode de détection de cible ou une désignation artificielle. Ensuite, la similarité est calculée entre la cible et le modèle. Un seuil est appliqué pour déterminer s'il y a une cible dans la scène de suivi.

Ce type de modèle est largement appliqué pour suivre la cible dans une séquence d'images. Dans [Dav, 2008] l'objet suivi est représenté par une forme géométrique rectangulaire, la zone suivie est modélisée en utilisant une interprétation probabiliste de l'analyse en composantes principales, l'estimation de la zone cible dans des images consécutives est réalisée par un algorithme d'échantillonnage avec l'estimateur de vraisemblance. Dans [Jung, 2014] le modèle d'objet est représenté par une forme rectangulaire, le suivi est effectué en trouvant la transformation affine appropriée qui transforme le modèle d'objet en la région d'image la plus similaire de l'image suivante. Dans [Sha, 2010], la forme de l'objet cible est modélisée comme un petit nombre de blocs rectangulaires, dont les positions dans la fenêtre de suivi sont déterminées de manière adaptative, tandis que son apparence est représentée par les distributions d'intensité et les poids de ces blocs. Le suivi est réalisé par un algorithme de segmentation rapide par la suite la forme cible est mise à jour en ajustant d'une manière adaptative ces blocs localement afin qu'ils fournissent une couverture maximale de la cible de premier plan. Dans [Mat, 2010], la cible est modélisée par une forme elliptique, et son apparence est représentée par un histogramme de couleur. Le suivi est obtenu en estimant l'état optimal qui représente la cible elliptique en utilisant un dispositif de poursuite probabiliste à deux étages basé sur le filtre à particules.

Ce type de représentation conduit à un suivi précis car il utilise généralement des informations globales sur la cible, tels que la couleur et le gradient ou la texture. Cependant la précision du suivi diminue en présence d'une déformation ou d'une large occlusion.

III.2.2. Le profil de contour

La représentation de la cible par le profil de contour est une recherche plus récente. Ce type de représentation nécessite une position générale de la cible, qui est fournie par l'humain. La récursivité par équation différentielle est ensuite résolue pour faire converger le profil du contour en valeur minimale locale par la méthode de la fonction énergétique. Une fonction d'énergie est généralement construite selon la caractéristique de l'image et la douceur du contour, comme le bord et la courbure. L'algorithme Snake est un exemple de méthode. Il adopte l'action combinée de l'image, force de contrainte interne et externe pour contrôler le mouvement du contour, où la force de l'image pousse la courbe au bord de l'image, la force interne contraint la douceur locale du contour, la force externe pousse la courbe du Snake.

Dans [Kana, 2018], le suivi d'objet basé sur le contour est effectué en localisant la région d'objet dans chaque image à travers le modèle d'objet créé par les images précédentes. Les formes d'objets sont considérées comme des silhouettes de délimitation et les résultats de suivi obtenus sont mis à jour dynamiquement dans les images vidéo. La détection de mouvement de cible est effectuée sur la base de la méthode de flux optique pour calculer le champ d'écoulement, en fonction des caractéristiques de distribution du flux optique. Une fois que le champ d'écoulement a été déterminé, il est utilisé pour l'analyse du mouvement et la soustraction d'arrière-plan par l'algorithme de maximisation basée sur le modèle gaussien efficace est effectuée pour obtenir les pixels d'avant-plan. Dans [Han, 2018], le contour est spécifié dans l'image initiale artificiellement, le suivi est réalisé en adoptant la divergence de Kullback-Leibler pour mesurer la similarité entre les distributions de la région candidate cible et celles de leur modèle, la distance Bhattacharyya est utilisée pour mesurer les similarités entre la région candidate de premier plan et de fond pour acquérir la localisation et le contour d'une cible en mouvement. Dans [Seo, 2017], le contour est représenté par un modèle d'espace d'état, qui peut gérer les mouvements rigides et non rigides des contours indépendamment. Le contour est généré par les points de contrôle et les fonctions de base. Le modèle permet de se concentrer sur les mouvements non-rigides au cours de la formation, et le modèle fonctionne pour des scénarios de mouvements rigides difficiles. Pour le suivi, une fonction de mesure qui exploite une variété de repères est proposée, la fonction de mesure qui considère trois facteurs

complémentaires à savoir le contraste, la cohérence temporelle et les couleurs. Dans [Pen, 2017], pour extraire les informations de régions rugueuses discriminantes pour le modèle de contour actif, le modèle d'apparence dynamique est proposé à base de super-pixels utilisant à la fois des couches globales et locales pour extraire la région cible rugueuse discriminante. Après l'initialisation manuelle du contour de la cible, la nouvelle position de la cible est localisée par l'algorithme MeanShift en combinant les caractéristiques de couleur et de texture. Pour capturer les changements d'apparence et extraire la région cible, un modèle d'apparence discriminative est adopté en combinant les informations cibles globales et locales. Pendant le suivi, le modèle d'apparence est mis à jours en utilisant le résultat de segmentation dans chaque frame.

III.2.3. Points locaux

La cible est représentée par un ensemble de points locaux, les caractéristiques locales en rapport avec ces points sont extraites, puis le suivi de la cible est réalisé en fusionnant ces caractéristiques. Le suivi avec des caractéristiques locales peut être appliqué efficacement lorsqu'il y a une occlusion partielle de la cible. Par contre la précision du suivi à l'aide de ce type de représentation baisse dans le cas d'un changement d'échelle ou d'une rotation de la cible.

Dans [Dee, 2017], les caractéristiques géométriques du visage sous trois formes différentes (point, ligne et triangle) sont étudiées. Les caractéristiques sont extraites à l'aide d'AdaBoost multi-classe, sélectif et le classifieur par apprentissage extrême. Le suivi est effectué en utilisant l'algorithme de Kanade-Lucas-Tomasi (KLT).

Dans [Xia, 2015], les auteures ont proposé une approche de détection de points d'intérêts pour le suivi d'objets. Un apprentissage basé sur des filtres linéaire est réalisé. Ensuite, une mise en commun maximale est effectuée pour détecter les points d'intérêts discriminants, qui sont appliqués dans le suivi par la méthode de correspondance structurées.

Dans [Ven, 2015] l'objet suivi est représenté par un ensemble de points d'intérêts obtenus à partir de l'image initiale par le détecteur de Harris et un dictionnaire est construit à partir du petit carré d'image entourant ces points clés. Les points candidats obtenus à partir de la fenêtre de recherche de l'image courante sont mis

en correspondance avec des points de l'objet en représentant de manière parcimonieuse les correctifs de coin candidats en termes de correctifs de dictionnaire. La correspondance entre les points d'intérêt cible et candidat est établi via la valeur maximale des coefficients parcimonieux.

III.3. Modèle d'apparence basé sur la couleur et la texture

Dans cette section nous allons présenter notre approche pour modéliser l'apparence de l'objet suivi. Le principe de notre modélisation est basé sur une description globale et locale à la fois, la caractéristique de couleur est considérée dans la description globale, basé sur les deux composante H et S du mode HSV l'histogramme pondéré est construit basé sur le modèle de mélange α -Gaussien. Tandis que la caractéristique de texture est utilisée pour la description locale, ainsi l'histogramme distinctif basé sur le motif binaire local uniforme est construit sur la composante V du mode HSV.

III.3.1. Histogramme pondéré basé sur le modèle de mélange α -Gaussien

III.3.1.1. Le modèle de mélange Gaussien

Les mélanges gaussiens ont été largement utilisés dans la vision par ordinateur pour la segmentation d'image, la soustraction de fond, la classification d'image et l'estimation de mouvement. Dans le suivi visuel, GMM a été utilisé pour modéliser l'apparence de la cible ou comme support à la procédure de suivi. Les modèles de mélange gaussien utilisent un ensemble de distributions gaussiennes pour approcher la fonction de densité sous-jacente de l'apparence de l'objet. L'apparence de la cible peut être représentée par un modèle de mélange gaussien au lieu de l'histogramme présente certains avantages. En effet, GMM fournit une représentation plus compacte de l'espace caractéristique car un faible nombre de paramètres est généralement suffisant pour modéliser la distribution de couleur de la cible. Cependant, si des caractéristiques de grande dimension sont utilisées, la dimension de l'histogramme standard augmente exponentiellement, tandis que le nombre de composants GMM reste relativement faible.

La densité conditionnelle d'un pixel, appartenant à la zone objet suivi, est modélisée comme un mélange de gaussiennes de K composantes, avec les proportions des

différentes composantes π_k , tel que $\sum_{k=1}^K \pi_k = 1$, avec des vecteurs moyens η_k et des matrices de covariance Σ_k , pour $k = 1, \dots, K$.

$$p_\alpha(x|\theta) = \sum_{k=1}^K \pi_k \mathcal{N}(I_n|\theta_k) \quad (\text{III.1})$$

$\theta_k = \{\pi_k, \eta_k, \Sigma_k\}$ est l'ensemble de paramètres lié à une mixture donnée. $\mathcal{N}(I_n|\theta_k)$ est la fonction de densité de probabilité de la distribution gaussienne. I_n est le vecteur caractéristique associé au pixel n de la zone objet. Le nombre de pixels dans cette zone est noté N . Une distribution gaussienne multi variables a une fonction de densité de probabilité donnée par :

$$\mathcal{N}(I_n|\theta_k) = \frac{1}{2\pi|\Sigma_k|^{\frac{1}{2}}} e^{-\frac{1}{2}(I_n - \eta_k)^t \Sigma_k^{-1} (I_n - \eta_k)} \quad (\text{III.2})$$

Les paramètres du GMM sont estimés en partitionnant les valeurs de densité des pixels de l'objet en utilisant l'algorithme itératif EM, les paramètres de mélange mis à jour à chaque itération. Le processus d'estimation est organisé en deux étapes: étape E et étape M.

La fonction de vraisemblance logarithmique pour la zone objet est :

$$\mathcal{L}(I|\theta) = \sum_{n=1}^N \log p(I_n|\theta) \quad (\text{III.3})$$

E-Step: La fonction auxiliaire est donnée comme suit:

$$\varphi(\theta, \theta^{t-1}) = \sum_{k=1}^K \sum_{n=1}^N [\log(\pi_k \mathcal{N}(I_n|\eta_k, \Sigma_k))] z_{nk} \quad (\text{III.4})$$

Où z_{nk} est une probabilité a posteriori définie par:

$$z_{nk} = \frac{\pi_k \mathcal{N}(I_n|\eta_k, \Sigma_k)}{\sum_{k=1}^K \pi_k \mathcal{N}(I_n|\eta_k, \Sigma_k)} \quad (\text{III.5})$$

M-step: La fonction auxiliaire est maximisée par rapport aux paramètres $\theta_k = \{\pi_k, \eta_k, \Sigma_k\}$ conduisant aux mises à jour suivantes:

$$\pi_k = \frac{1}{N} \sum_{n=1}^N z_{nk} \quad (\text{III.6})$$

$$\eta_k = \frac{\sum_{n=1}^N I_n z_{nk}}{\sum_{n=1}^N z_{nk}} \quad (\text{III.7})$$

$$\Sigma_k = \frac{\sum_{n=1}^N z_{nk} (I_n - \eta_k)(I_n - \eta_k)^t}{\sum_{n=1}^N z_{nk}} \quad (\text{III.8})$$

Les itérations ci-dessus sont répétées jusqu'à convergence de la vraie semblance.

III.3.1.2. Le modèle d'apparence proposé

En utilisant les composantes H et S du mode HSV, on construit l'histogramme pondéré basé sur le modèle de mélange α -Gaussien pondéré en utilisant l'algorithme EM. Le nième pixel à l'intérieur de la zone rectangulaire est décrit par un α -GMM Figure III-2 de K composantes avec des proportions de mélange π_k , avec des vecteurs moyens η_k et des matrices de covariance Σ_k , pour $k = 1, \dots, K$.

La fonction de densité de probabilité de la distribution α -gaussienne est donnée par :

$$p_\alpha(x|\theta) = \begin{cases} c \left(\sum_{k=1}^K \pi_k \mathcal{N}(x|\theta_k)^{\frac{1-\alpha}{2}} \right)^{\frac{2}{1-\alpha}}, & \alpha \neq 1 \\ c \text{Exp} \left(\sum_{k=1}^K \pi_k \text{Log}(\mathcal{N}(x|\theta_k)) \right), & \alpha = 1 \end{cases} \quad (\text{III.9})$$

Avec c est une constante de normalisation et $\theta_k = \{\pi_k, \eta_k, \Sigma_k\}$ est le jeu de paramètres de la distribution α -GMM. $\mathcal{N}(x|\eta_k, \Sigma_k)$ est la fonction de densité de probabilité du même distribution Gaussienne.

α -GMM a une grande puissance dans la modélisation stochastique, qui intègre un nouveau paramètre α à chaque composant, cette intégration rend la modélisation du mélange flexible et permet à la fonction de probabilité de posséder une distribution plus plane Figure III-3, la flexibilité introduite par le facteur α permet de modéliser une plus large gamme de distribution de données, également le

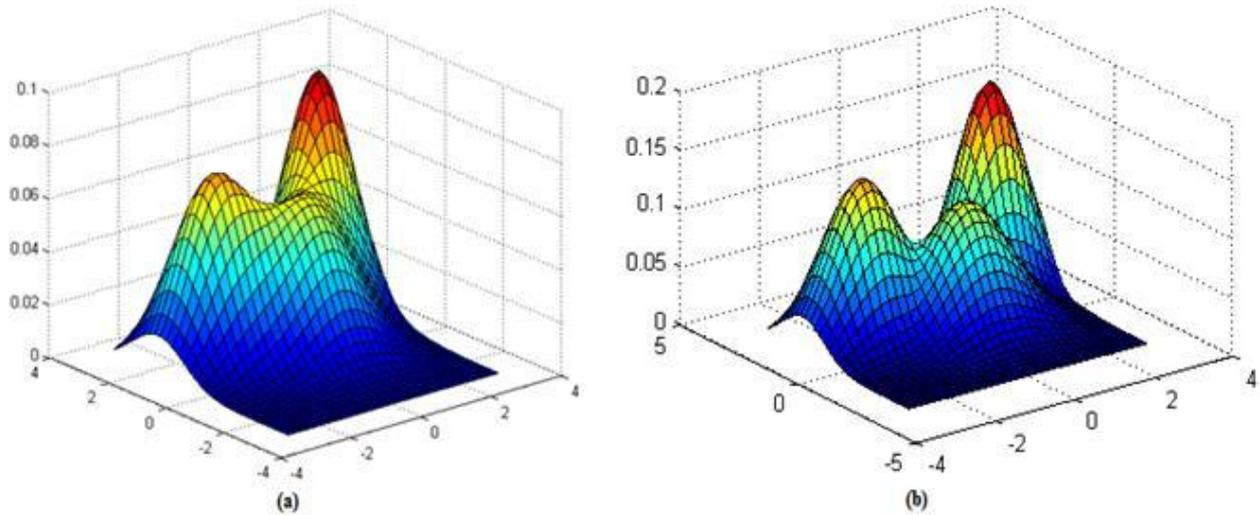


Figure III-2. Tracé de surface pour (a) GMM et (b) α -GMM avec $\alpha=-5$

processus de cette méthode de modélisation est considéré comme similaire au système de perception humaine. De plus, pour intégrer les informations spatiales et augmenter la fiabilité, les distributions sont pondérées en utilisant un profil de noyau isotrope, en effet les pixels proches du centre de la zone de la cible contiennent plus d'informations de premier plan, par conséquent leur contribution est plus importantes.

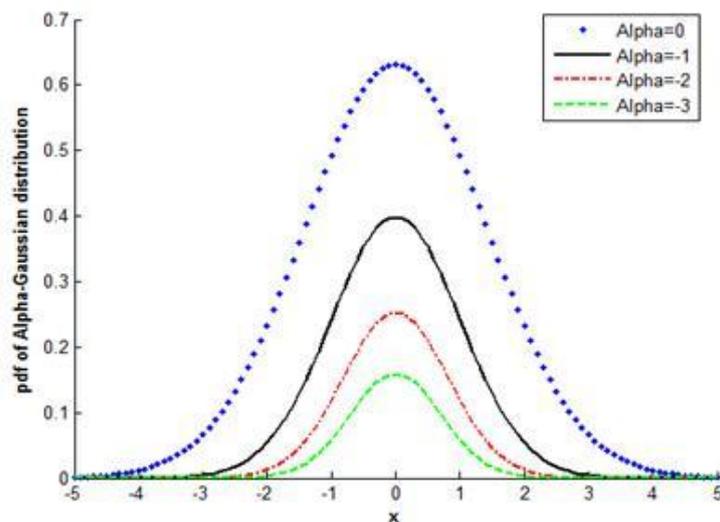


Figure III-3. La distribution α -Gaussien pour différentes valeurs de α .

Pour comparer la qualité de la modélisation basé respectivement sur GMM et α -GMM nous avons générer un ensemble de donnée de façons aléatoire, les données générées forme trois classes, La Figure III-4 expose la répartition réelle des données et le centre de chaque classe.

Par la suite nous avons estimé les paramètres des deux mixtures par l'algorithme EM, la Figure III-5 expose la représentation des données par GMM ainsi que les centres estimés. La Figure III-6 expose la modélisation des données par α -GMM et les centres estimés pour différent valeur du paramètre α .

La modification du paramètre α permet une meilleure modélisation, le bon choix de ce paramètre assure une bonne couverture des données réels, pour la valeur de $\alpha = -0,5$ la mixture résultante est loin d'être représentatif et la discrimination est mauvaise, en diminuant la valeur de α la discrimination s'améliore ainsi pour $\alpha = -5$ la couverture est plus flexible que pour GMM, en effet pour cette valeur le résultat de la modélisation est meilleur.

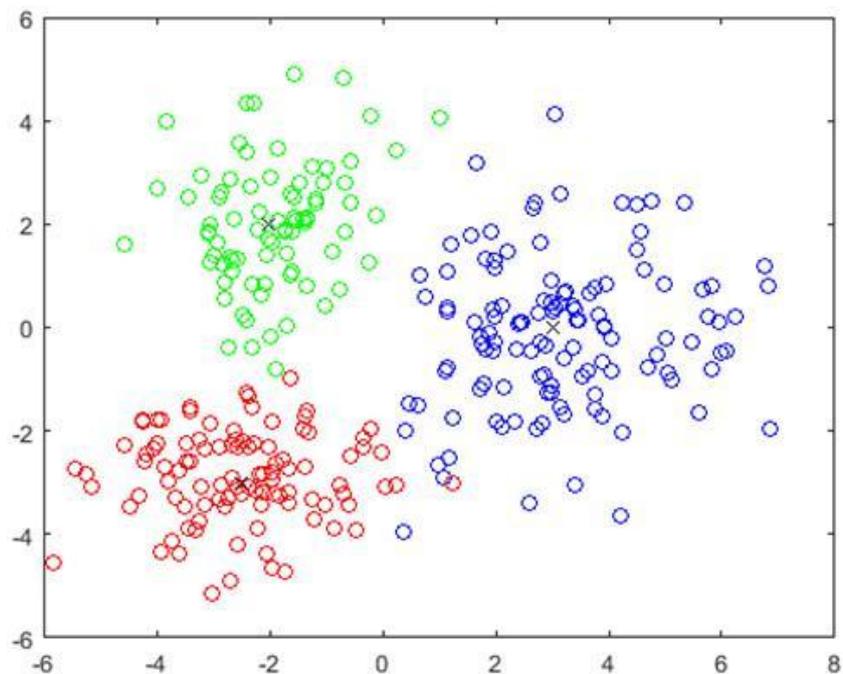


Figure III-4. Répartition spatiale des données générées aléatoirement

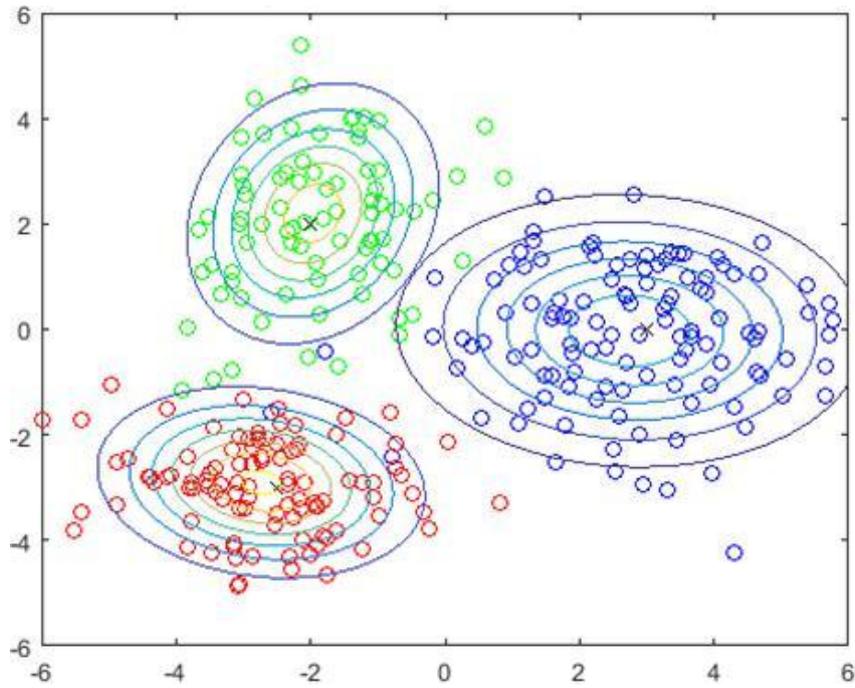


Figure III-5. Données originales et fonctions de distribution de probabilité estimées du GMM

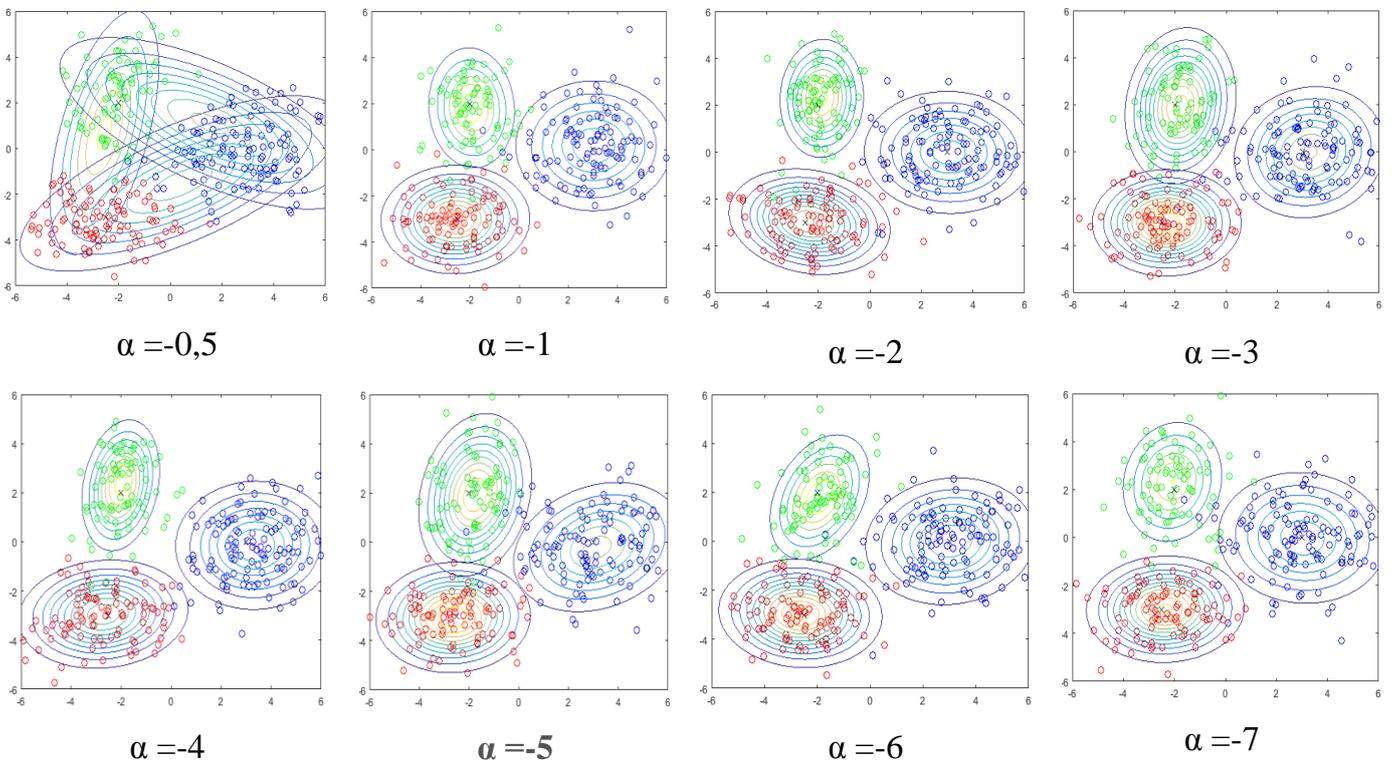


Figure III-6. Données originales et fonctions de distribution de probabilité estimées du α -GMM

Soit h et w la hauteur et la largeur de la région de la cible rectangulaire, et soit x_n, y_n les coordonnées du nième pixel dans la zone rectangulaire et x_c, y_c est son centre, un poids w_n est assigné à chaque pixel en masquant la zone rectangulaire avec un profil de noyau isotrope Ker Figure III-7 et Figure III-8:

$$w_n = Ker([x_n - x_c]^t \mathcal{M}^{-1} [y_n - y_c]) \quad (III.10)$$

Avec
$$\mathcal{M} = \begin{bmatrix} h & 0 \\ 0 & w \end{bmatrix}$$

L'algorithme EM est utilisé pour estimer les paramètres de α -GMM dans le cas de $\alpha \neq 1$, La fonction de log-vraisemblance pour la zone rectangulaire est:

$$\mathcal{L}(I|\theta) = \sum_{n=1}^N \log p_{\alpha \neq 1}(I|\theta) \quad (III.11)$$

Où N est le nombre de pixels dans la zone rectangulaire, $I = \{I_n\}_{n=1 \dots N}$ et $I_n = [H_n S_n]^t$.

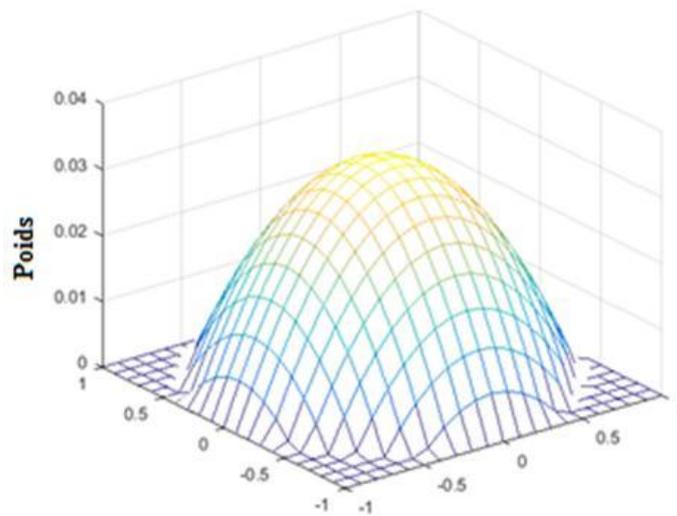


Figure III-7. Le profil du noyau Epanechnikov pour les positions de pixel normalisées.

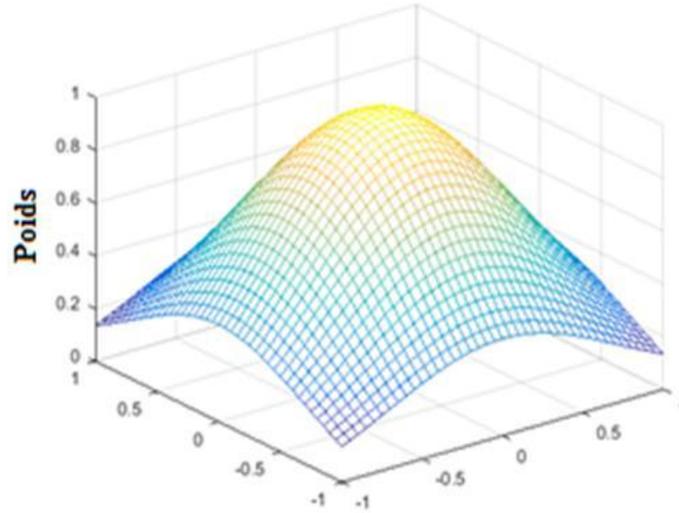


Figure III-8. Le profil du noyau Gaussien pour les positions de pixel normalisées.

E-Step: La fonction auxiliaire est donnée comme suit:

$$\varphi(\theta, \theta^{t-1}) = \frac{2}{1-\alpha} \sum_{k=1}^K \sum_{n=1}^N \left[\log \left(\pi_k \mathcal{N}(I_n | \eta_k, \Sigma_k)^{\frac{1-\alpha}{2}} \right) z_{nk} \right] \quad (\text{III.12})$$

La probabilité a posteriori est donnée par:

$$z_{nk} = \frac{\pi_k \mathcal{N}(I_n | \eta_k, \Sigma_k)^{\frac{1-\alpha}{2}}}{\sum_{k=1}^K \pi_k \mathcal{N}(I_n | \eta_k, \Sigma_k)^{\frac{1-\alpha}{2}}} \quad (\text{III.13})$$

M-step: La fonction auxiliaire est maximisée par rapport aux paramètres $\theta_k = \{\pi_k, \eta_k, \Sigma_k\}$ conduisant aux mises à jour suivantes:

$$\pi_k = \frac{1}{N} \sum_{n=1}^N z_{nk} \quad (\text{III.14})$$

$$\eta_k = \frac{\sum_{n=1}^N I_n z_{nk}}{\sum_{n=1}^N z_{nk}} \quad (\text{III.15})$$

$$\Sigma_k = \frac{\sum_{n=1}^N z_{nk} (I_n - \eta_k)(I_n - \eta_k)^t}{\sum_{n=1}^N z_{nk}} \quad (\text{III.16})$$

Ensuite, nous créons l'histogramme des caractéristiques de bas niveau Figure III-9 à partir d'un ensemble de caractéristiques de bas niveau $I = \{I_n\}_{n=1\dots N}$ extraites de la zone rectangulaire de l'objet cible comme suit:

$$H_{\alpha-GMM}^k = \frac{\sum_{n=1}^N \omega_n z_{nk}}{\sum_{n=1}^N \omega_n} \quad (\text{III.17})$$

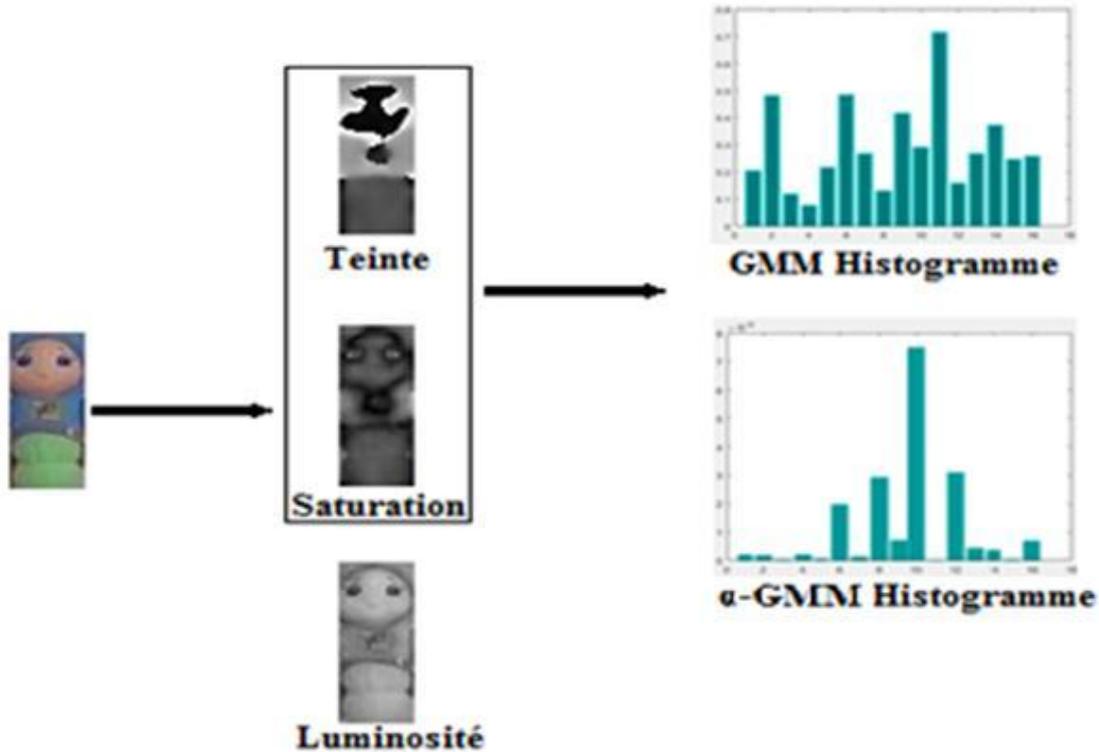


Figure III-9. GMM Histogramme et α -GMM Histogramme basés sur les caractéristiques H et S du mode HSV.

III.3.2. L'histogramme distinctif basé sur le motif binaire local uniforme

L'opérateur de texture le motif local binaire a été introduit pour la classification de texture invariante par rotation, et a eu beaucoup de succès dans la vision et la reconnaissance de formes par son pouvoir discriminant et son invariance par illumination. L'opérateur LBP d'origine marque les pixels d'une image avec des nombres décimaux, appelés motif local binaire, qui codent la structure locale autour de chaque pixel. Il procède, comme illustré sur la figure III-10, chaque pixel est comparé à ses voisins dans un voisinage en soustrayant la valeur de pixel central; Les valeurs strictement négatives résultantes sont codées avec 0 et les autres avec 1; Un nombre binaire est obtenu en concaténant tous ces codes binaires dans le sens des aiguilles d'une montre en commençant par celui en haut à

gauche et la valeur décimale correspondante est utilisée pour l'étiquetage. Les nombres binaires dérivés sont appelés modèles binaires locaux ou codes LBP.

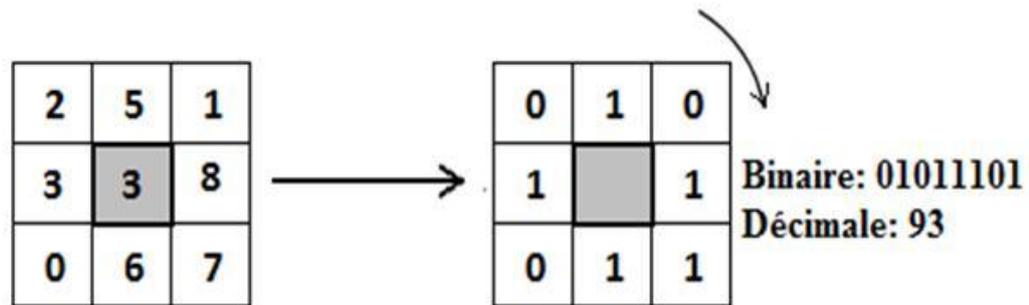


Figure III-10. LBP opérateur appliqué sur un voisinage 3×3.

Un voisinage local est défini comme un ensemble de points d'échantillonnage régulièrement espacés sur un cercle qui est centré sur le pixel à marquer, et les points d'échantillonnage qui ne tombent pas dans les pixels sont interpolés en utilisant une interpolation bilinéaire, permettant ainsi n'importe quel rayon et n'importe quel nombre de points d'échantillonnage dans le voisinage. La Figure III-11, montre quelques exemples d'opérateur LBP étendu, où la notation (P,R) indique un voisinage de P points d'échantillonnage sur un cercle de rayon R.

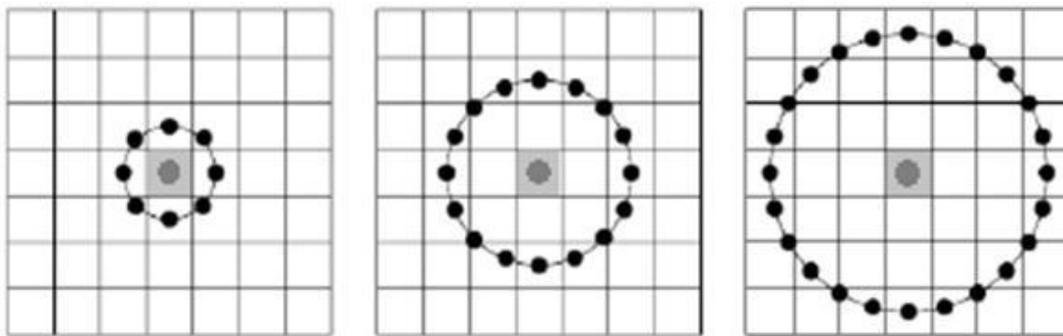


Figure III-11. Exemples de l'opérateur LBP étendu: les voisinages circulaires (8,1), (16,2) et (24,3).

Formellement, étant donné un pixel en (x_c, y_c) , le LBP résultant peut être exprimé sous forme décimale comme:

$$LBP(x_c, y_c) = \sum_{p=0}^{P-1} s(I_p - I_c) 2^p \tag{III.18}$$

Où I_c et I_p sont respectivement des valeurs de niveau de gris du pixel central et P des pixels environnants dans le voisinage de cercle de rayon R , et la fonction $s(x)$ est définie comme:

$$s(x) = \begin{cases} 1 & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases} \quad (\text{III.19})$$

L'opérateur LBP de base est invariant par rapport aux transformations d'échelle de gris monotones, préservant l'ordre d'intensité des pixels dans les voisinages locaux.

Le LBP génère un total de 256 modèles allant de 0 à 255. Considérer une telle taille de modèles est une tâche coûteuse en termes de temps de calcul. L'ULBP Figure III-12 permet de surmonter ce problème, en effet un motif est considéré comme uniforme si le nombre de transitions dans la séquence entre 0 à 1 ou 1 à 0 est inférieur ou égal à deux, en effet la complexité de calcul est réduite, car le sous-ensemble ne contiendra que $P(P-1) + 3$ modèles, au lieu de 2^P , comme la version originale avait. Dans ce cas, lors de la construction de l'histogramme correspondant, chaque motif uniforme unique est assigné à une case séparée, alors que tous les codes non-uniformes sont affectés à une seule case (qui est généralement la dernière).

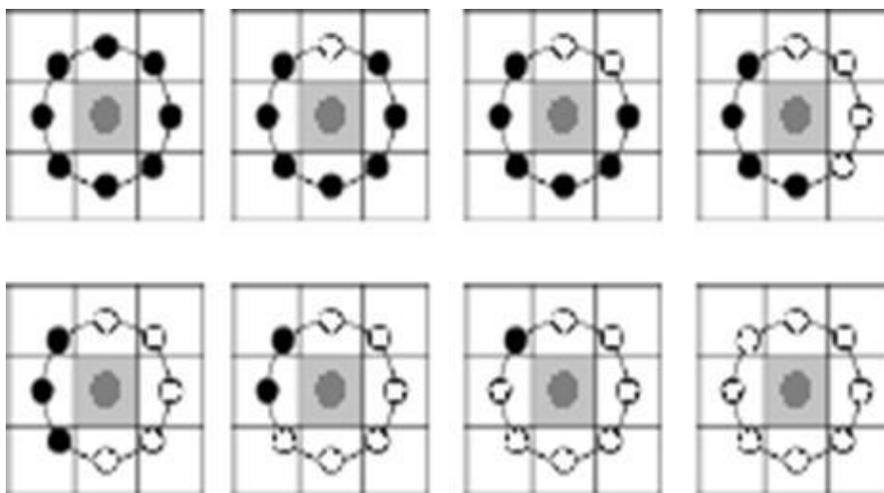


Figure III-12. Exemples de motifs uniformes dans le cas d'un voisinage circulaires (8, 1).

L'ULBP ne forme que 23% du nombre total de LBP avec 8 pixels voisins alors que la majorité des caractéristiques LBP peuvent être catégorisées pour être uniformes vu que la plupart des modèles binaires locaux dans les images naturelles sont uniformes, en effet les textures contiennent en moyenne 75% à 90% des motifs en

ULBP seulement, par conséquent la taille est réduite sans perdre trop d'informations utiles. A partir de là, il est bien compris qu'en considérant seulement les ULBP, on peut facilement reconnaître ou analyser les problèmes liés aux divers domaines du traitement d'image et de la reconnaissance de formes.

Basé sur le composant V, nous caractérisons la structure spatiale de la texture de l'objet local en utilisant l'opérateur ULBP calculé comme suit :

$$ULBP(x_c, y_c) = \begin{cases} \sum_{p=0}^{P-1} s(I_p^V - I_c^V) 2^p & \text{si } U \leq 2 \\ P(P-1) + 2 & \text{sinon} \end{cases} \quad (\text{III.20})$$

Dont

$$U = \left| s(I_{p-1}^V - I_c^V) - s(I_0^V - I_c^V) \right| + \sum_{p=1}^{P-1} \left| s(I_p^V - I_c^V) - s(I_{p-1}^V - I_c^V) \right| \quad (\text{III.21})$$

$$s(x) = \begin{cases} 1 & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases} \quad (\text{III.22})$$

Différents niveau de luminosité pour l'objet Doll et leurs opérateurs LBP ainsi que les histogrammes correspondants sont présentés par la Figure III-13. Les résultats obtenus pour l'opérateur ULBP sont affichés par la Figure III-14. On peut conclure que l'opérateur ULBP montre une bonne stabilité vis-à-vis des changements de la luminosité en effet les histogrammes associés aux différents opérateurs ULBP présentent une grande similarité et invariabilité.

D'après ces évaluations de base, on peut conclure que l'utilisation des ULBP conserve presque les propriétés fondamentales ou significatives de la texture, ainsi ULBP préserve la stabilité et l'invariance vis-à-vis des changements de la luminosité et de rotation qui caractérisent l'opérateur LBP, avec l'avantage de réduire la dimensionnalité globale.

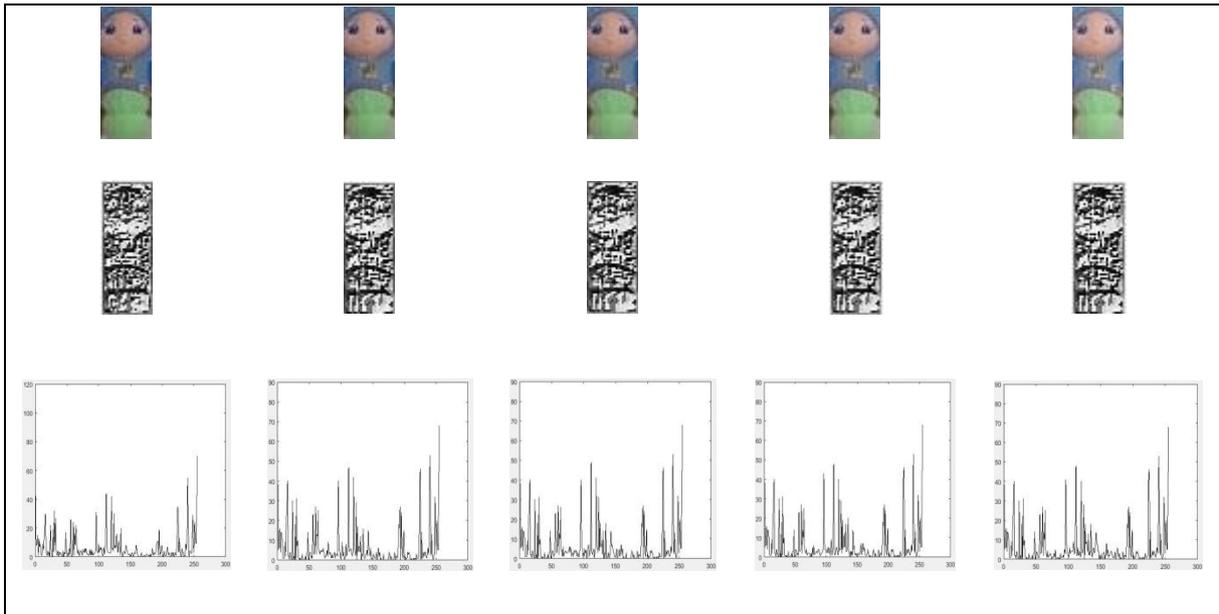


Figure III-13. Image LBP et histogramme LBP pour le même objet sous différents niveau de luminosité.

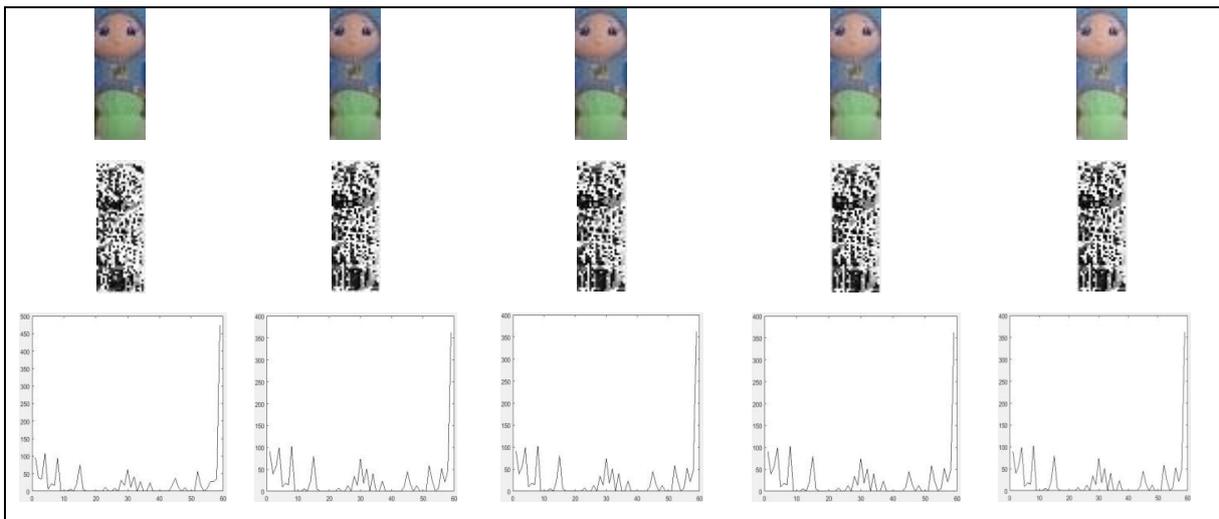


Figure III-14. Image ULBP et histogramme ULBP pour le même objet sous différents niveau de luminosité.

En plus pour voir le comportement de l'opérateur ULBP vis-à-vis à la transformation de rotation, nous avons exposé l'objet Doll aux différents degrés de rotation, les résultats obtenus pour les opérateurs LBP et ULBP sont présentés respectivement par la Figure III-15 et la Figure III-16. Pour les différents degrés de rotation, les histogrammes associés aux objets ULBP montrent une stabilité encourageante et une invariabilité acceptable.

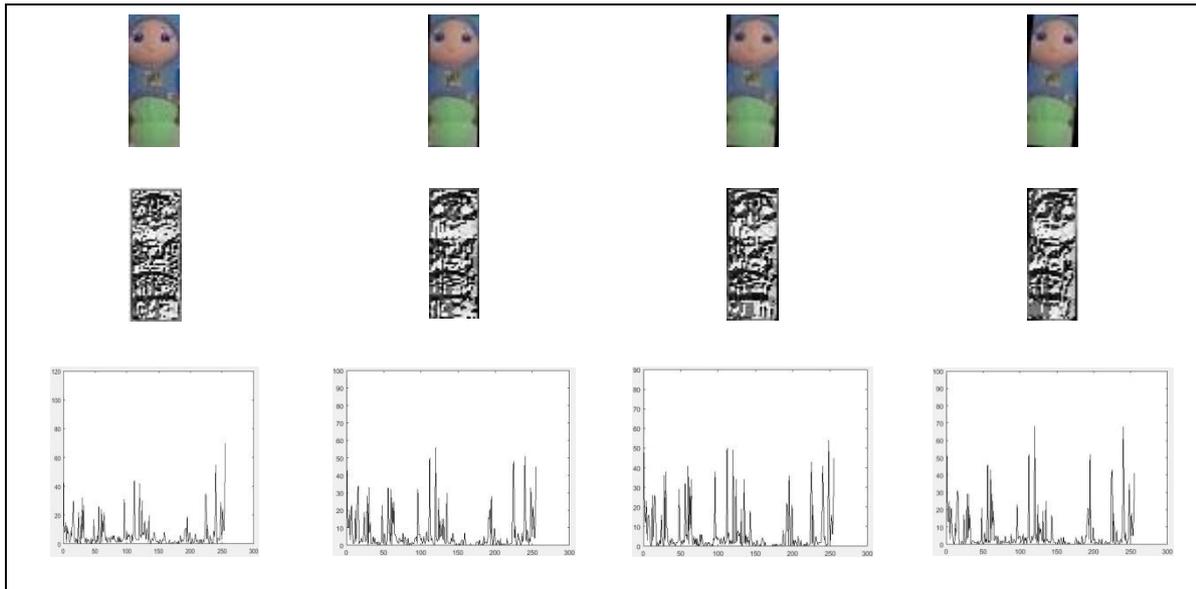


Figure III-15. Image LBP et histogramme LBP pour le même objet sous différentes orientations

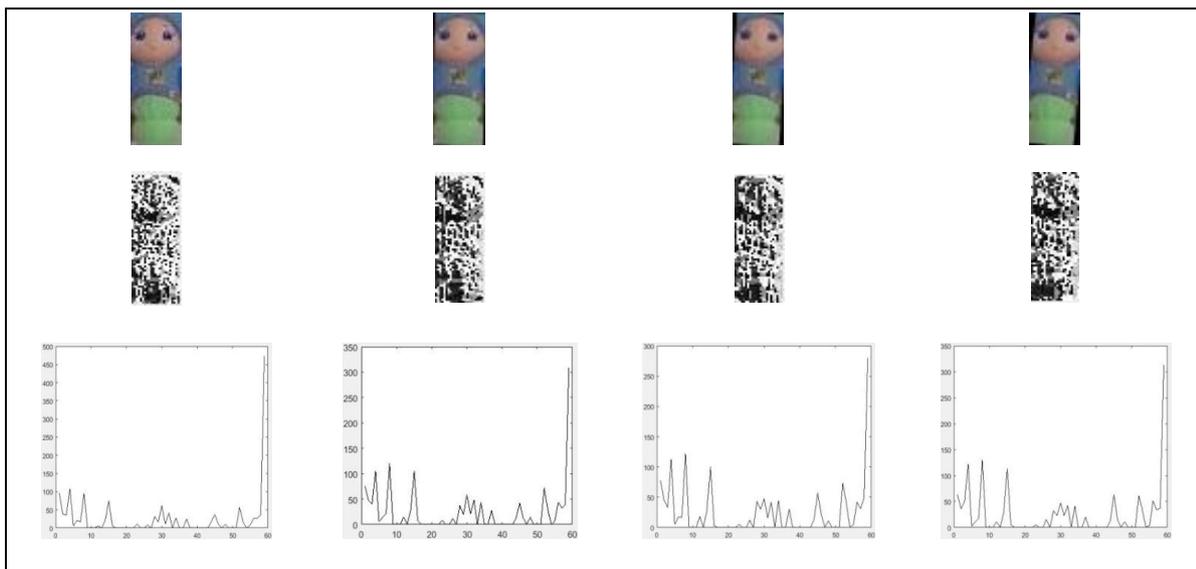


Figure III-16. Image ULBP et histogramme ULBP pour le même objet sous différentes orientations.

Basé sur le composant V, nous caractérisons la structure spatiale de la texture de l'objet local en utilisant l'opérateur ULBP Figure III-18. Ensuite, nous calculons l'historgramme ULBP avec 59 bins, au lieu de 256. Les 58 premières dimensions contiendront le nombre d'occurrences de motifs uniformes, et le dernière contiendra le nombre d'occurrences de tous les motifs non uniformes Figure III-19, ainsi la taille est réduite sans perdre trop d'informations utiles.

Ensuite, la cible est divisée en 4×4 blocs sans chevauchement et un bloc central Figure III-20, ainsi chaque bloc est caractérisé par un ULBP-histogramme calculé sur la composante V du mode HSV Figure III-21.

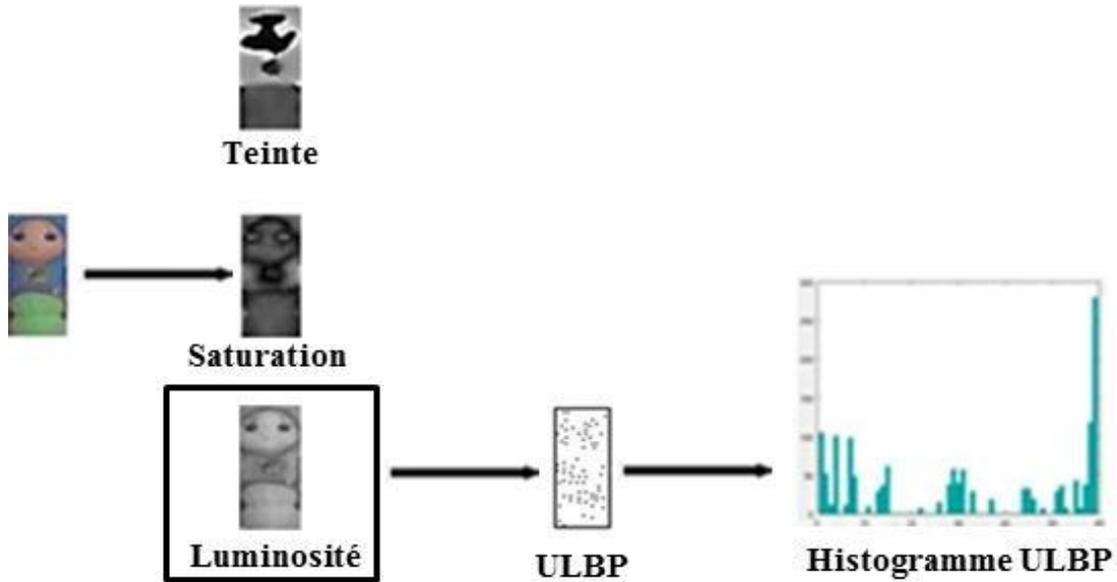


Figure III-17. Représentation de la composante V du mode HSV par l'histogramme ULBP.

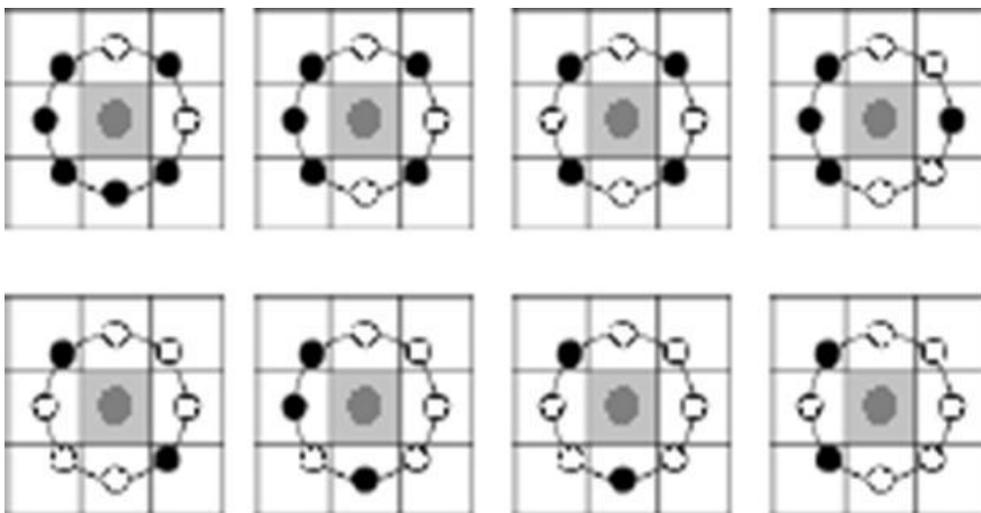


Figure III-18. Exemples de motifs non uniformes dans le cas d'un voisinage circulaires (8,1).

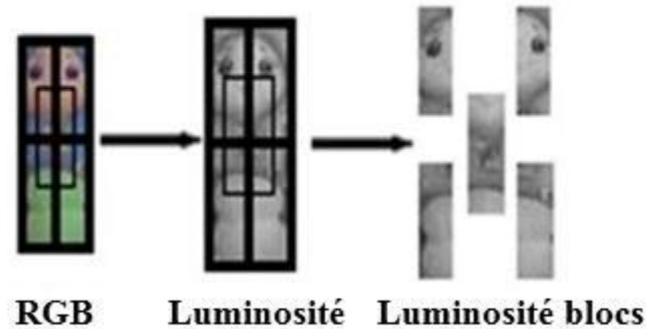


Figure III-19. Division de l'objet en blocs (exemple de division en $2 \times 2 + 1$ blocs).

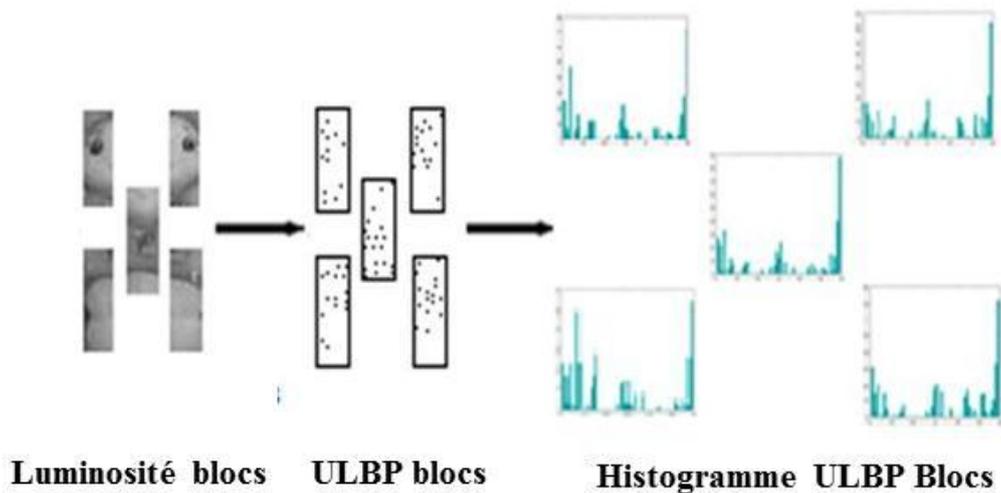


Figure III-20. Représentation des blocs par l'histogramme ULBP.

Pour incorporer la relation entre les blocs, nous calculons ULBP histogrammes distinctifs qui est défini comme la différence entre le bloc central et ceux dans le voisinage comme suit :

$$DH_{ULBP}^j(i) = |H_{ULBP}^c(i) - H_{ULBP}^j(i)| \quad (III.23)$$

Où $H_{ULBP}^c(i)$ est le $i^{\text{ème}}$ élément du ULBP histogramme du bloc central, $H_{ULBP}^j(i)$ est le $i^{\text{ème}}$ élément du ULBP histogramme du $j^{\text{ème}}$ bloc, et $DH_{ULBP}^j(i)$ est le $i^{\text{ème}}$ élément du DULBP histogramme DULBP du $j^{\text{ème}}$ bloc, par la suite tous les ULBP histogrammes distinctifs sont concaténés pour former un seul descripteur de texture DH_{ULBP} pour l'objet suivi.

III.4. Mise à jour du modèle d'apparence

Durant le processus de suivi, l'apparence de la cible peut changer à cause des différentes conditions de l'environnement, ce qui peut rendre le modèle d'apparence faible, par conséquent le suivi de la cible peut dévier. Pour résoudre ce problème, l'idée principale est de permettre au modèle de s'adapter aux différents changements de l'environnement, ainsi la mise à jour de manière dynamique du modèle d'apparence de la cible est indispensable. Dans cet objectif, pour le modèle de mixture α -Gaussien de nouvelles composantes sont insérées dans le modèle de mixture en utilisant des pixels proches de la cible qui ont une probabilité faible. De plus, si l'importance d'un composant devient suffisamment faible, le composant est éliminé de la mixture.

A une certaine fréquence, à chaque nombre de frames M , nous fixons M égale à 50, par conséquent la mise à jour aura lieu toutes les deux secondes pour 25 frames par seconde. On initialise la nouvelle composante avec des paramètres calculés à partir des pixels qui représentent le quantile inférieur de la vraisemblance et avec un faible poids. Par la suite l'algorithme EM est utilisé pour estimer la moyenne et la covariance correcte du nouveau composant. Les paramètres du modèle de mixture de base, la moyenne et la covariance, sont fixés pour éviter le problème de dérive, seules leurs proportions de mélange changent en raison de l'insertion du nouveau composant. Finalement, si l'importance d'un composant est inférieure à un seuil (inférieure à $0.1/K$), le composant est supprimé du la mixture.

Pour le modèle basé sur le motif binaire local uniforme, nous adoptons la méthode de mise à jour présentée dans [Yaw, 2017], dans ce but le modèle d'apparence est mis à jour de façon adaptative ni conservative ni libre. La stratégie conservative permet de favoriser le modèle de base tandis que la méthode libre privilège le modèle estimé.

Pour le modèle basé sur le motif binaire local uniforme, nous adoptons la méthode de mise à jour présentée dans [Yaw, 2017], dans ce but le modèle d'apparence est mis à jour de façon adaptative ni conservative ni libre. La stratégie conservative permet de favoriser le modèle de base tandis que la méthode libre privilège le modèle estimé.

Le modèle conservative est construit comme suivant :

$$DH_{C-ULBP}^* = (1 - \zeta_C)DH_{C-ULBP} + \zeta_C \widehat{DH}_{ULBP} \quad (\text{III.24})$$

Où $\zeta_C \in [0, 1]$ est le facteur d'apprentissage qui contribue au modèle mis à jour conservatif DH_{C-ULBP}^* , DH_{C-ULBP} est le modèle conservatif actuel et \widehat{DH}_{ULBP} est le modèle correspondant à l'état estimé.

Et le modèle libéral est construit comme suivant :

$$DH_{L-ULBP}^* = (1 - \zeta_L)DH_{L-ULBP} + \zeta_L \widehat{DH}_{ULBP} \quad (\text{III.25})$$

Où $\zeta_L \in [0, 1]$ est le facteur d'apprentissage qui contribue au modèle mis à jour libre DH_{L-ULBP}^* , ζ_L est choisi plus grande que ζ_C . DH_{L-ULBP} est le modèle libre actuel et \widehat{DH}_{ULBP} est le modèle correspondant à l'état estimé.

Le modèle approuvé est celui pour lequel la distance de Bhattacharyya avec le modèle candidat est minimale. Dans [Yaw, 2017], il a été approuvé que lorsque ζ_C est mis à 0.1 et que ζ_L est réglé sur 0.5, la performance de la méthode de mise à jour est bonne.

III.5. Conclusion

Dans ce chapitre nous avons présenté notre modèle d'apparence basé sur deux caractéristiques, la couleur et la texture, telle représentation reflète les caractéristiques statistiques globales et locales de l'apparence de l'objet. Ainsi l'incorporation de la couleur et la texture dans le modèle d'apparence permet de coder plus d'informations tout en tirons profit de la complémentarité des deux indices, par conséquent le modèle proposé reste stable et robuste vis-à-vis des différentes conditions de suivi en particulier les changements de la luminosité, des objets similaires et d'encombrement de l'arrière-plan.

Nous avons représenté la distribution des caractéristiques de couleur par α -GMM sur les composants H et S. Alors que, la distribution des caractéristiques de texture est représentée par l'histogramme distinctif ULBP basé sur l'opérateur ULBP pour la composante V. par la suite, les deux caractéristiques sont fusionnées pour construire le model d'apparence de la cible. Nous avons conclu ce chapitre par présenter la stratégie de mise à jour assurant ainsi au modèle proposé une bonne stabilité et adaptabilité

Chapitre IV : Suivi visuel par le filtre adaptatif H_∞ PF

Dans ce chapitre nous présentons notre approche de suivi visuel réparti sur deux étapes, étape d'estimation globale dans laquelle l'état globale de la cible est calculé via une stratégie adaptative basé sur le filtre H_∞ et un système dynamique paramétré, les paramètres du système sont automatiquement ajuster en fonction des modèles de probabilités. L'étape d'estimation est effectuée par le filtre particulaire dont les particules sont générées et pondérées à partir du modèle d'apparence présenté dans le chapitre III. Pour rendre la contribution de chaque caractéristique efficace leurs poids sont mis à jours automatiquement. Les deux estimations sont combinées dans le cadre de l'algorithme IMM pour déterminer l'état de l'objet suivi. Des expérimentations qualitatives et quantitatives sont réalisées pour évaluer les performances ne notre méthode, ainsi les résultats des tests et des comparaisons sont exposés et discutés.

CHAPITRE IV : SUIVI VISUEL PAR LE FILTRE ADAPTATIF H_∞PF	104
IV.1. INTRODUCTION	105
IV.2. ESTIMATION GLOBALE.....	106
IV.2.1. MODELISATION	106
IV.2.2. ESTIMATION	109
IV.2.3. MISE A JOUR ADAPTATIVE.....	111
IV.2.4. ETUDES DES PERFORMANCES.....	112
IV.3. ESTIMATION LOCALE	115
IV.3.1. ECHANTILLONNAGE ET PONDERATION	115
IV.3.2. ESTIMATION	116
IV.3.3. MISE A JOUR DES COEFFICIENTS DE FUSION	117
IV.4. ALGORITHME DE SUIVI	118
IV.5. EXPERIMENTATIONS ET DISCUSSIONS	122
IV.5.1. SEQUENCES VIDEO	122
IV.5.2. CRITERES DE PERFORMANCE.....	124
IV.5.3. RESULTATS QUALITATIFS.....	125
IV.5.4. RESULTATS QUANTITATIFS	138
IV.6. CONCLUSION	145
CONCLUSION ET PERSPECTIVES	146

IV.1. Introduction

Deux types de problèmes peuvent rendre assez complexe le développement de méthodes de suivi robuste. Le premier se pose lorsque le mouvement de la cible est rapide et complexe et le second concerne le changement de l'apparence de l'objet tout au long de la séquence. Dans l'objectif d'élaborer une approche de suivi robuste et stable vis à vis de telles situations, l'approche de suivi que nous proposons suppose que le mouvement global de la cible est linéaire et le mouvement local est non linéaire. Ainsi l'approche s'appuie sur la coopération de deux modules (Figure IV-1). Le premier traite l'aspect global du système et permet de gérer son comportement dynamique supposé linéaire alors que le second module gère le comportement local de la cible supposé non linéaire. Le module global renferme trois fonctions : la modélisation, l'estimation de l'état du système et la mise à jour.



Figure IV-1. Structure de notre Algorithme de suivi visuel

La modélisation consiste à élaborer un modèle dynamique pour décrire le mouvement de la cible. Dans notre approche, le système dynamique adopté est composé de deux sous-systèmes paramétrés dont chacun peut couvrir un

comportement dynamique différent de la cible. La fonction d'estimation de l'état du système est assurée par le filtre H ∞ adaptatif. Une stratégie adaptative assure la mise à jour des paramètres du système dynamique. Pour le module local, l'estimation locale est effectuée par le filtre à particules. Basé sur les caractéristiques de couleur et de texture, le modèle d'apparence est construit et utilisé par la suite pour mesurer la vraisemblance d'observation des particules et estimer l'état locale associé à chaque sous modèle dynamique. La fusion des deux caractéristiques est effectuée de façon adaptative de manière que la contribution de chaque caractéristique est ajustée automatiquement en fonction de son poids. Finalement les deux estimations sont combinées dans le cadre de l'algorithme IMM pour déterminer l'état de l'objet suivi.

Le reste de ce chapitre est organisé comme suit. Section IV 2 présentes les étapes de l'estimation globale basée sur un système dynamique adaptative. L'essentiel de l'estimation locale par le filtre particulaire est expliqué dans la section IV 3. Dans la section IV 4 notre approche de suivi adaptative basé sur le filtre H ∞ PF est présentée. Les résultats expérimentaux et l'analyse de la performance sont mis en évidence dans la section IV 5. Dans la section IV 6 nous terminons ce chapitre par une conclusion.

IV.2. Estimation globale

IV.2.1. Modélisation

Le choix du modèle dynamique pour décrire le mouvement de la cible affecte largement la robustesse de l'algorithme de suivi, notamment dans le cas du suivi d'une cible dont le mouvement est complexe. Les mouvements d'une cible sont normalement classés en deux classes: manœuvre et non-manœuvre. Un mouvement non-manœuvre est un mouvement rectiligne à vitesse constante, appelé aussi mouvement uniforme. Alors que les autres mouvements appartiennent au mode de manœuvre. Les modèles dynamiques les plus courants sont les modèles à vitesse quasi constante NCV et à accélération quasi constante (NCA) [Rong, 2004]. Comme leur nom l'indique, ils décrivent le mouvement d'une cible se déplaçant sur une trajectoire rectiligne respectivement à vitesse constante ou à une accélération constante. Ces modèles sont très populaires en raison de leur simplicité. Ils sont

adaptés aux cas où la cible suit typiquement des trajectoires déterministes, avec de faibles écarts dus à des perturbations aléatoires affectant respectivement les composantes de vitesse ou d'accélération.

Lorsque la cible se déplace dans une certaine direction, le mouvement est mieux décrit par le modèle NCV. Le model NCV est décrit par le système discret suivant :

$$X_{k+1} = A_{NCV}X_k + w_k \quad (\text{IV.1})$$

Où w_t est le bruit blanc gaussien et A_{NCV} est la matrice de transition à temps discret, et qui est donnée par:

$$A_{NCV} = \begin{bmatrix} 1 & 0 & T & 0 \\ 0 & 1 & 0 & T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (\text{IV.2})$$

Dans le cas d'un mouvement de manœuvre d'une cible, le modèle d'accélération à bruit blanc est le plus simple à utiliser. Il ne diffère du modèle NCV que par le niveau de bruit, le processus de bruit blanc w_t a une intensité plus élevée que celle utilisée dans un modèle non-manœuvre. Il est parfois utilisé lorsque la manœuvre est assez petite ou aléatoire.

Un autre modèle qui décrit de façon simple un mouvement de manœuvre est le modèle d'accélération appelé processus de Wiener. Il suppose que l'accélération soit un processus de Wiener. Il est également appelé simplement le modèle à accélération constante (CA) ou plus précisément le modèle à accélération quasi constante (NCA). Le modèle NCA est décrit par le système discret suivant :

$$X_{k+1} = A_{NCA}X_k + w_k \quad (\text{IV.3})$$

Ou

$$A_{NCA} = \begin{bmatrix} 1 & 0 & T & 0 & T^2/2 & 0 \\ 0 & 1 & 0 & T & 0 & T^2/2 \\ 0 & 0 & 1 & 0 & T & 0 \\ 0 & 0 & 0 & 1 & 0 & T \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (\text{IV.4})$$

Un modèle usuel pour le suivi de cible qui présente un mouvement de manœuvre est le modèle d'accélération de Singer (SA) qui est à la base de nombreux modèles de manœuvre efficaces. Il suppose que l'accélération cible est un processus de Markov stationnaire de premier ordre et de moyenne nulle. Il est aussi par essence un modèle adaptatif à travers un ajustement de ses paramètres.

Le modèle d'accélération de Singer (SA) est décrit par l'équation suivante :

$$\dot{a}(t) = -\beta a(t) + w(t), \beta > 0 \quad (\text{IV.5})$$

Lorsque $w(t)$ est un bruit blanc de moyenne nulle, la représentation en espace d'état du modèle de Singer en temps discret pour la composante x est :

$$x_{k+1} = A_{SA}x_k + w_k \quad (\text{IV.6})$$

Avec

$$A_{SA} = \begin{bmatrix} 1 & T & (\beta T - 1 + e^{-\beta T})/\beta^2 \\ 0 & 1 & (1 - e^{-\beta T})/\beta \\ 0 & 0 & e^{-\beta T} \end{bmatrix} \quad (\text{IV.7})$$

Le point fort du modèle Singer repose sur une détermination précise du paramètre β . Ce paramètre est l'inverse de la constante de temps de manœuvre et dépend donc de la durée de la manœuvre.

L'utilisation d'un seul modèle risque de ne pas garder un suivi assez proche voire perdre tôt ou tard la cible. Cependant, l'utilisation d'un grand nombre de sous-modèles fournit une description précise des mouvements cibles du monde réel, en revanche, elle augmente la complexité des calculs. Par contre, l'utilisation d'un petit ensemble de sous-modèles peut aboutir à des résultats non précis et en conséquence les performances de l'algorithme de suivi peuvent diminuer.

Le problème peut être résolu en utilisant un nombre de sous modèles très réduit et qui couvrent les différents styles de mouvements de la cible, sans en aucun cas augmenter la complexité du système. Pour cette raison, nous proposons une approche basée sur le modèle multiple adaptatif dont le principe consiste à utiliser deux sous modèles d'accélération de Singer paramétrés. Le premier sous modèle peut couvrir les différents modes de mouvement uniforme et le deuxième modèle représente les différents styles de mouvements de manœuvre.

Le système de sous modèles paramétré adopté est à l'itération k $M_k = \{\beta_k^l, \beta_k^c, \beta_k^r\}$

Où

- β_k^l est le sous-modèle gauche du modèle de manœuvre
- β_k^r est le sous-modèle droit du modèle uniforme, et
- β_k^c le paramètre central du système.

Le choix du modèle de Singer est basé sur le fait qu'il possède une couverture assez large, pour un choix de $0 < \beta \leq \beta_{max}$, et se place entre les modèles NCV et NCA. En effet, au fur et à mesure que la constante de temps de manœuvre augmente β devient petite et le modèle de Singer se réduit au modèle NCA puisque la partie déterministe de l'accélération dans le modèle de Singer devient constante à la limite lorsque la constante de temps de manœuvre augmente. D'autre part, lorsque la constante de temps de manœuvre diminue (β augmente) l'accélération devient un bruit blanc et le modèle d'accélération de Singer se réduit alors au modèle NCV.

Le paramètre β de chaque sous-modèle est ajusté suivant la probabilité du modèle à chaque itération. Pour le premier sous-modèle, le paramètre β est auto-ajusté dans un intervalle qui correspond à divers types de manœuvre, et pour le second sous-modèle, β varie automatiquement dans un intervalle qui correspond à divers types de mouvement uniforme.

IV.2.2. Estimation

Le système adaptatif présenté sera utilisé comme modèle dynamique de base pour estimer le mouvement global de l'objet, en effet la plupart des mouvements naturels correspondent au modèle linéaire, Sur la base de cette idée, les travaux antérieures se base sur le filtre de Kalman pour effectuer l'estimation globale. Le filtre de Kalman est une méthode d'estimation optimale pour le cas gaussien linéaire ce qui n'est pas toujours le cas. En vue globale, notre stratégie consiste à utiliser le filtre H ∞ pour estimer la position de la cible. Le filtre H ∞ est une alternative intéressante au célèbre filtre de Kalman dans la plupart des problèmes d'estimation, en effet l'optimalité du filtre de Kalman repose sur la connaissance de la covariance du processus et du bruit de mesure. Dans la plupart des applications du monde réel,

ce type d'information a priori n'est pas disponible. D'autre part, le filtre H ∞ ne fait aucune hypothèse sur les perturbations et minimise l'erreur d'estimation du pire cas.

Notre approche d'estimation globale est basée sur le filtre H ∞ (section II.2.3). Le système $M_k = \{\beta_k^l, \beta_k^c, \beta_k^r\}$ étant modélisé par un sous-modèle gauche de manœuvre β_k^l et un sous-modèle droit uniforme β_k^r de type modèle de Singer, deux filtres H ∞ sont utilisés dans le cadre du filtre IMM, chacun est associés respectivement aux sous modèles dynamique gauche et droite. Les deux filtres H ∞ sont exécutés en parallèle, pour estimer les états \hat{X}_k^j et les covariances \hat{P}_k^j associés à chaque sous modèles, à base des états des modes interactifs \bar{X}_k^j Equation IV-37, en utilisant les équations itératifs suivantes:

$$M_k^j = A_j M_{k-1}^j A_j^t - A_j M_{k-1}^j \bar{H}^t (\bar{R} + \bar{H} M_{k-1}^j \bar{H}^t)^{-1} \bar{H} M_{k-1}^j A_j^t + G G^t \quad (\text{IV.8})$$

$$\hat{P}_k^j = \left((M_k^j)^{-1} - \gamma^{-2} L^t L \right)^{-1} \quad (\text{IV.9})$$

$$\bar{S}_k^j = H \hat{P}_k^j H^t + I, \quad (\text{IV.10})$$

$$K_k^j = A_j \hat{P}_k^j H^t (\bar{S}_k^j)^{-1} \quad (\text{IV.11})$$

$$\bar{X}_k^j = A_j \bar{X}_{k-1}^j, \quad (\text{IV.12})$$

$$\bar{Z}_k^j = Y_k - H \bar{X}_k^j, \quad (\text{IV.13})$$

$$\hat{X}_k^j = \hat{X}_{k-1}^j + K_k^j \bar{Z}_k^j. \quad (\text{IV.14})$$

Où A_j est la matrice d'état associée au sous-modèle

Avec

$j=1$ correspond aux sous modèle gauche et $j=2$ correspond aux sous modèle droite ;

H est la matrice de mesure; K_k^j est le gain du filtre H ∞ du sous-modèle j ; \bar{S}_k^j est la covariance d'innovation du sous-modèle j et \bar{Z}_k^j est l'innovation résiduelle du sous-modèle j . Les matrices \bar{R} et \bar{H} sont définies dans la section II.2.3.

IV.2.3. Mise à jour adaptative

Le système de modèles initial est sélectionné comme suit, $M_0 = \{\beta_0^l = 0,05, \beta_0^r = 7\}$. L'auto-ajustement des paramètres du système des sous modèles est effectué selon les étapes logiques suivantes.

- Ajustement du paramètre central du système :

$$\beta_{k+1}^c = \mu_k^l \beta_k^l + \mu_k^r \beta_k^r \quad (\text{IV.15})$$

Où $\{\mu_k^l, \mu_k^r\}$ représente les probabilités modèle pour chaque sous-modèle j, respectivement le sous modèle gauche et droite, et qui sont calculés en utilisant Equation IV-42.

- Ajustement des paramètres des sous-modèles:

Cas 1: Lorsque le comportement dynamique de la cible ne montre pas de changement significatif, l'ensemble du système des sous modèles restera stable, alors :

$$\beta_{k+1}^l = \begin{cases} \beta_{k+1}^c - \Delta\beta_k^l/2 & \text{si } \mu_k^l < t_1 \\ \beta_{k+1}^c - \Delta\beta_k^l & \text{sinon} \end{cases} \quad (\text{IV.16})$$

$$\beta_{k+1}^r = \begin{cases} \beta_{k+1}^c + \Delta\beta_k^r/2 & \text{si } \mu_k^r < t_1 \\ \beta_{k+1}^c + \Delta\beta_k^r & \text{sinon} \end{cases} \quad (\text{IV.17})$$

Où $\Delta\beta_k^l = \max(\beta_k^c - \beta_k^l, \Delta\beta)$, $\Delta\beta_k^r = \max(\beta_k^r - \beta_k^c, \Delta\beta)$, $t_1 < 0.1$ est un seuil qui permet de détecter un mode invalide, et $\Delta\beta$ est le pas de modification de β .

Cas 2: Lorsque le comportement dynamique de la cible passe en mode uniforme et que l'ensemble de modèles passe de gauche à droite, dans ce cas μ_k^r est supérieur à μ_k^l , alors :

$$\beta_{k+1}^l = \beta_{k+1}^c - \Delta\beta_k^l \quad (\text{IV.18})$$

$$\beta_{k+1}^r = \begin{cases} \beta_{k+1}^c + 2\Delta\beta_k^r & \text{si } \mu_k^r > t_2 \\ \beta_{k+1}^c + \Delta\beta_k^r & \text{sinon} \end{cases} \quad (\text{IV.19})$$

Où $t_2 = 0.9$ est un seuil qui permet de détecter un modèle significatif.

Cas 3: Lorsque le comportement dynamique de la cible bascule vers le mode de complexe et que l'ensemble de modèles passe de droite à gauche, dans ce cas μ_k^l est supérieur à μ_k^r , alors :

$$\beta_{k+1}^l = \begin{cases} \beta_{k+1}^c - 2\Delta\beta_k^l & \text{si } \mu_k^l > t_2 \\ \beta_{k+1}^c - \Delta\beta_k^l & \text{sinon} \end{cases} \quad (\text{IV.20})$$

$$\beta_{k+1}^r = \beta_{k+1}^c + \Delta\beta_k^r \quad (\text{IV.21})$$

IV.2.4. Etudes des performances

Dans le but de comparer les performances de notre méthode, nous avons conduit une série d'expérimentation sur des séquences vidéo qui présentent différents challenges et dont l'objet est caractérisé par un comportement dynamique diversifié. Les vidéos sont Ping-Pong, Fish, Jumping et David. Les expérimentations sont réalisées dans les mêmes conditions, trois algorithmes de suivi ont fait sujet de comparaison, le premier algorithme est le filtre particulaire (PF), le second est le filtre particulaire combiné au filtre de Kalman (KPF) et le troisième algorithme est le filtre particulaire combiné au filtre adaptative H ∞ (AH ∞ PF).

IV.2.4.1. Critères de performance

La Figure IV-2 expose les résultats des expérimentations en terme d'erreur de localisation du centre (en pixels) telle que définie par :

$$ELC_k = \sqrt{(GO_x^k - O_x^k)^2 + (GO_y^k - O_y^k)^2} \quad (\text{IV.22})$$

Où O_x^k, O_y^k sont respectivement la position x et la position y du centre de la zone de délimitation de l'objet dans l'image k, et GO_x^k, GO_y^k sont respectivement la position x et la position y du centre de la vérité terrain.

Pour assurer l'équité dans nos expériences, chacun des trois algorithmes de suivi est exécuté vingt fois pour chaque séquence, par conséquent l'erreur de localisation moyenne normalisée définie par Equation IV-23 est calculée. Les résultats numériques sont exposés dans le Tableau IV-1

$$ELMN = \frac{1}{20} \sum_{n=1}^{20} \frac{1}{N_{frames}} \sum_k ELC_k \quad (IV.23)$$

IV.2.4.2. Résultats expérimentaux

D'après les résultats des expérimentations, on peut conclure que l'utilisation du filtre H ∞ adaptative pour estimer le comportement dynamique global de l'objet en mouvement permet une meilleure précision que le filtre de Kalman, en effet pour toutes les séquences vidéo utilisées dans les expérimentations l'erreur de localisation est la plus petite en faveur de notre méthode. En plus le fait d'estimer l'état global du mouvement de l'objet et par la suite calculer son état local par le filtre particulière, améliore la précision du suivi surtout avec une estimation globale basé sur le filtre H ∞ adaptative. En effet cette stratégie permet de prendre en compte les mesures actuelles et rendre la distribution des particules plus proche de la distribution postérieure ce qui améliore la précision du suivi.

Nous terminons cette partie d'évaluation en présentant les temps moyens d'exécution pour les trois algorithmes. D'après les résultats exposés dans le Tableau IV-2, on peut conclure que la méthode basée sur la combinaison du filtre H ∞ adaptative avec le filtre particulière réalise un suivi avec un temps de calcul réduit par rapport aux autres méthodes utilisées pour la comparaison. Ainsi L'utilisation du filtre H ∞ adaptative pour estimer l'état dynamique globale de l'objet garanti un compromis entre, précision du suivi et coût de calcul réduit.

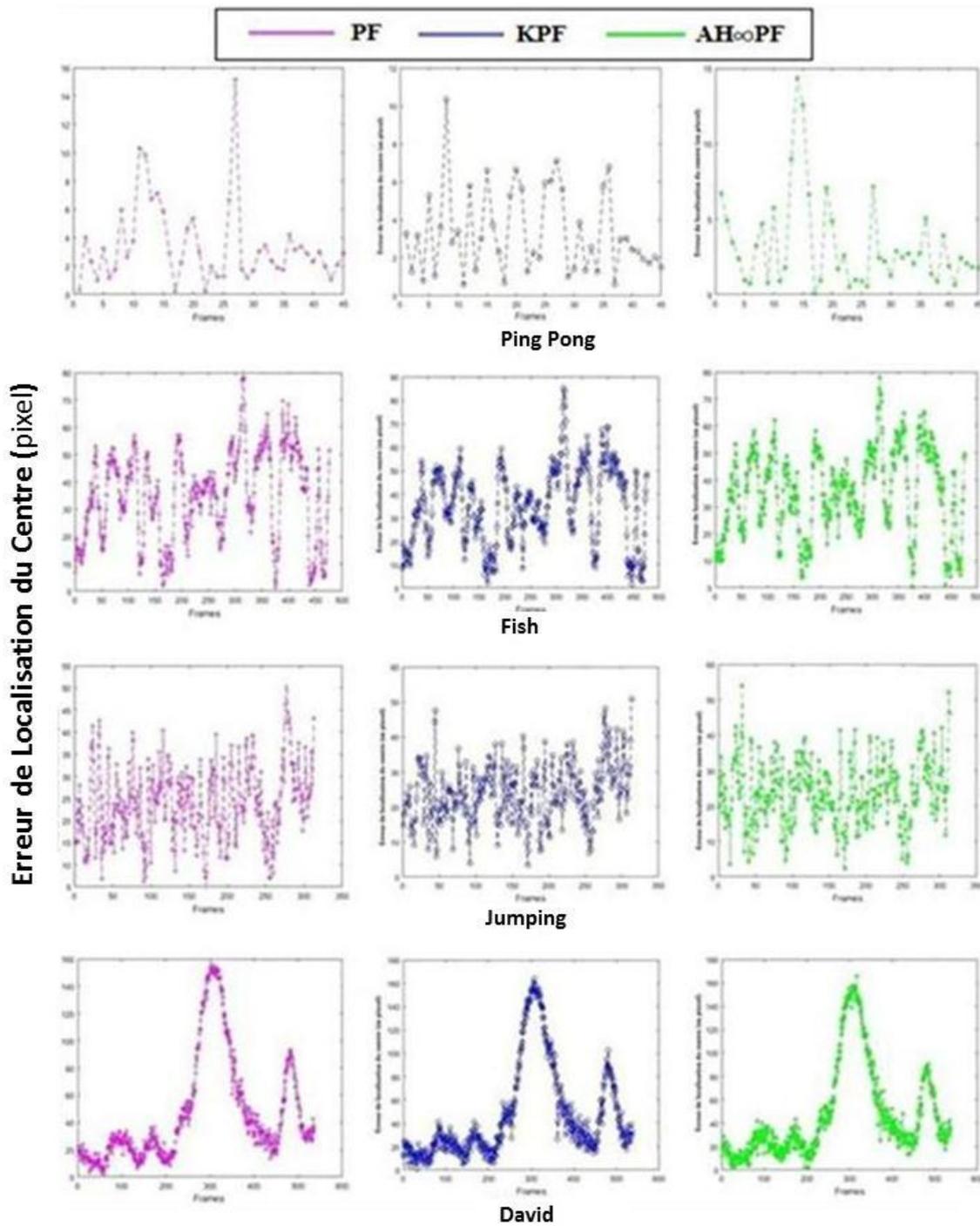


Figure IV-2. Comparaison entre PF, KPF et $AH_{\infty}PF$ en termes d'erreur de localisation du centre (en pixels)

Tableau IV-1. Erreur de localisation moyenne normalisée (en pixels).

Sequence	N_{Frames}	PF	KPF	AH ∞ PF
Ping-Pong	86	3.8691	3.3199	3.2612
Fish	476	35.1754	35.1614	34.4474
Jumping	313	24.0145	24.0139	23.9409
David	537	47.9861	47.4573	47.4540

Tableau IV-2. Temps d'exécution moyen (en second).

Sequence	N_{Frames}	PF	KPF	AH ∞ PF
Ping-Pong	86	7.3942	6.3446	6.2325
Fish	476	35.1754	35.1614	34.4474
Jumping	313	24.0145	24.0139	23.9409
David	537	47.9861	47.4573	47.4540

IV.3. Estimation locale

Dans la section précédente nous avons estimé l'état globales qui correspond à la dynamique globale de la cible, ainsi deux filtres H ∞ parallèles sont exécutés, chaque filtre est basé sur un sous modèle de Single qui correspond chacun à un type de comportement dynamique différent, le paramètre de chaque sous-modèle est auto-ajusté par une stratégie adaptative. Dans cette section nous allons présenter les détails de l'estimation locale basée sur le filtre particulière.

IV.3.1. Echantillonnage et pondération

L'état global calculé dans la section précédente est utilisé comme entré pour le filtre particulière. Ainsi un ensemble d'échantillons X_j^s est généré à partir de chaque fonction de densité de probabilité antérieure correspondant à chaque estimation globale et sa covariance. Chaque particule est pondérée en fonction de la probabilité d'observation suivant le modèle d'apparence introduit dans le chapitre III. Basé sur le fait que l'apparence de l'objet ne présente pas des changements radical entre deux images consécutifs, et la stabilité du mode HSV vis-à-vis des changements de

la luminosité, le modèle d'observation de l'objet reste stable dans le temps, donc les centres μ_k et les variances Σ_k du modèle α -GMM restent inchangés, donc nous exécutons l'algorithme EM juste pour la première image, et à chaque itération, seulement les proportions de mélange sont calculées comme suit :

$$\pi_k^s = \frac{1}{N} \sum_{n=1}^N \frac{\pi_k \mathcal{N}(I_n^s | \eta_k, \Sigma_k)^{\frac{1-\alpha}{2}}}{\sum_{k=1}^K \pi_k \mathcal{N}(I_n^s | \eta_k, \Sigma_k)^{\frac{1-\alpha}{2}}} \quad (\text{IV.24})$$

Où $I_n^s = [H_n S_n]^t$ est le vecteur caractéristique formé par les composantes H et S du mode HSV du nième pixel dans la zone échantillon.

IV.3.2. Estimation

Afin de calculer la vraisemblance du modèle cible et du modèle de base, une métrique basée sur le coefficient de Bhattacharyya est utilisée [Bha, 943], en effet les deux vecteurs caractéristiques $H_{\alpha\text{-GMM}}^s$ et DH_{ULBP}^s associés au model d'apparence introduit dans le chapitre III sont calculés pour chaque particule, puis les similitudes entre les vecteurs particules $H_{\alpha\text{-GMM}}^s$, DH_{ULBP}^s et les vecteurs models $H_{\alpha\text{-GMM}}$, DH_{ULBP} sont calculés en utilisant la distance de Bhattacharya comme suit:

$$d_{\alpha\text{-GMM}}^s = \sqrt{1 - \rho_{\alpha\text{-GMM}}^s} \quad (\text{IV.25})$$

Et

$$d_{ULBP}^s = \sqrt{1 - \rho_{ULBP}^s} \quad (\text{IV.26})$$

où $\rho_{\alpha\text{-GMM}}^s$ est le coefficient de Bhattacharyya qui correspond au produit scalaire $\langle H_{\alpha\text{-GMM}}^s, H_{\alpha\text{-GMM}} \rangle$ calculé en utilisant Equation IV-27, tandis que ρ_{ULBP}^s est le coefficient de Bhattacharyya qui correspond au produit scalaire $\langle DH_{ULBP}^s, DH_{ULBP} \rangle$ calculé en utilisant Equation IV-28.

$$\rho_{\alpha\text{-GMM}}^s = \sum_{u=1}^K \sqrt{H_{\alpha\text{-GMM}}^s(u) \times H_{\alpha\text{-GMM}}(u)} \quad (\text{IV.27})$$

Et

$$\rho_{ULBP}^s = \sum_{u=1}^m \sqrt{DH_{ULBP}^s(u) \times DH_{ULBP}(u)} \quad (\text{IV.28})$$

On obtient ainsi la similarité entre les distributions discrètes des candidats et le modèle, par la suite chaque particule est pondérée en fonction de la vraisemblance de l'observation calculée en intégrant les deux caractéristiques comme suit:

$$\xi^s = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(\lambda_1 d_{\alpha-GMM}^s + \lambda_2 d_{ULBP}^s)^2}{\sigma^2}\right) \quad (IV.29)$$

Où σ est l'écart type gaussien, dans nos expériences, σ est fixé à 0.2. λ_1 correspond au poids attribué au model d'apparence basé sur la couleur et λ_2 correspond au poids attribué au model d'apparence basé sur la texture, tel que :

$$\lambda_1 + \lambda_2 = 1 \quad (IV.30)$$

Enfin la prédiction locale de l'état cible à partir du modèle de fusion multi-caractéristiques pour chaque sous-modèle dynamique j est calculée comme suit:

$$\hat{X}_j = \sum_{s=1}^{NP} \xi^s X_j^s \quad (IV.31)$$

IV.3.3. Mise à jour des coefficients de fusion

Les poids de chaque caractéristique sont ajustés dynamiquement. La stratégie de mise à jour repose sur la contribution de la caractéristique pour la fiabilité du suivi, en effet les particules montrent des distributions différentes observées sur différentes caractéristiques, donc les particules avec des poids élevés observés sur une caractéristique avec une discrimination majeure, sont plus concentrées près de la position réelle de la cible. La mise à jour des poids des caractéristiques est basée sur l'analyse de la distribution spatiale des particules, pour cela un paramètre relatif de dispersion qui est une mesure de l'écart relatif des valeurs d'une distribution à sa valeur centrale est considéré. Dans cet objectif, des paramètres statistiques à savoir la moyenne et la déviation standard sont calculés.

$$m_{\{c,t\}} = \frac{1}{NP} \sum_{s=1}^{NP} \xi_{\{c,t\}}^s \quad (IV.32)$$

$$\sigma_{\{c,t\}} = \sqrt{\frac{1}{NP-1} \sum_{s=1}^{NP} (\xi_{\{c,t\}}^s - m_{\{c,t\}})^2} \quad (IV.33)$$

Où $\xi_{\{c,t\}}^s$ est le poids d'apparence associé à chaque caractéristique {couleur , texture} pour chaque particule, donné par la formule suivante :

$$\xi_{\{c,t\}}^s = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(1-\rho_{\{c,t\}}^s)^2}{\sigma^2}\right) \quad (\text{IV.34})$$

Avec $\rho_{\{c,t\}}^s$ est le coefficient de Bhattacharyya associé à chaque caractéristique, par la suite, en tant que paramètre de dispersion, le coefficient de variation est donnée par :

$$\tau_{\{c,t\}} = \frac{\sigma_{\{c,t\}}}{m_{\{c,t\}}} \quad (\text{IV.35})$$

Ainsi on obtient les poids associés à chaque caractéristique :

$$\lambda_{\{c,t\}} = \frac{1}{\tau_{\{c,t\}}} \quad (\text{IV.36})$$

IV.4. Algorithme de suivi

Notre approche consiste à utiliser le filtre interactif multiple modèle pour le processus de suivi visuel. L'organigramme de la méthode de suivi visuel proposée est représenté par la Figure IV-3. Elle est composée de cinq étapes: étape de mélange et d'interaction, étape de filtrage H ∞ , étape de filtre particulière, étape de mise à jour du model de probabilité, étape de mise à jour du système dynamique, étape d'ajustement automatique des coefficients de fusion et enfin étape de combinaison.

On suppose que la position de l'objet à suivre dans l'image initiale est connue, le modèle d'apparence basé sur les deux caractéristiques de couleur et de texture dans le mode HSV est calculé comme expliqué respectivement dans section III.3.1 et section III.3.2.

Dans l'étape d'interaction et de mélange l'état du mode interactif et sa covariance pour chaque sous modèle dynamique sont calculés à base des probabilités des modèles et les probabilités des modèles de transition en utilisons les formules suivante :

$$\bar{X}_{k-1}^j = \sum_{i=1}^2 \bar{X}_{k-1}^i \mu_{ij}(k-1) \quad (\text{IV.37})$$

$$\bar{P}_{k-1}^j = \sum_{i=1}^2 ((\hat{X}_{k-1}^i - \bar{X}_{k-1}^i)(\hat{X}_{k-1}^i - \bar{X}_{k-1}^i)^T + \hat{P}_{k-1}^i) \mu_{ij}(k-1) \quad (\text{IV.38})$$

Avec μ_{ij} est la probabilité de mélange donnée par :

$$\mu_{ij}(k-1) = \frac{1}{c_{k-1}^j} p_{ij} \mu_i(k-1) \quad (\text{IV.39})$$

Et μ_i est la probabilité des modèles donnée par Equation IV-42 et c_{k-1}^j est un facteur de normalisation donnée par :

$$c_{k-1}^j = \sum_{i=1}^2 p_{ij} \mu_i(k-1) \quad (\text{IV.40})$$

Et p_{ij} est les probabilités des modèles de transition donnée par :

$$p_{ij} = \begin{bmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{bmatrix} \quad (\text{IV.41})$$

Dans l'étape de filtrage, les états \hat{X}_k^j associés respectivement aux sous modèle dynamique gauche et droite sont estimés par le filtre H ∞ comme expliqué dans la section IV.2.2

Par la suite le filtre particulaire est exécuter pour estimer l'état local de chaque sous modèle comme expliqué dans la section IV.3.2

Dans l'étape de mise à jours des probabilités des modèles, la probabilité de chaque sous modèle est mise à jour, en fonction de la covariance d'innovation de la covariance, l'innovation résiduelle et les probabilités des modèles de transition selon les équations suivantes:

$$\mu_j(k) = \frac{\Lambda_j \sum_{i=1}^2 p_{ij} \mu_i(k-1)}{\sum_{i=1}^2 \Lambda_i \sum_{i=1}^2 p_{ij} \mu_i(k-1)} \quad (\text{IV.42})$$

Avec Λ_j est une fonction de vraisemblance définie par :

$$\Lambda_j = |2\pi \bar{S}_k^j|^{-1/2} \times \exp \left\{ \frac{-1}{2} [\bar{Z}_k^j]^t [\bar{S}_k^j]^{-1} [\bar{Z}_k^j] \right\} \quad (\text{IV.43})$$

Après, le système dynamique $M_k = \{\beta_k^l, \beta_k^r\}$ est ajusté automatiquement utilisant les règles logiques expliquées dans de la section IV.2.3.

Dans l'étape de mise à jour des coefficients de fusion des indices, Les poids de chaque caractéristique sont ajustés dynamiquement suivant la stratégie expliquée dans de la section IV.3.3.

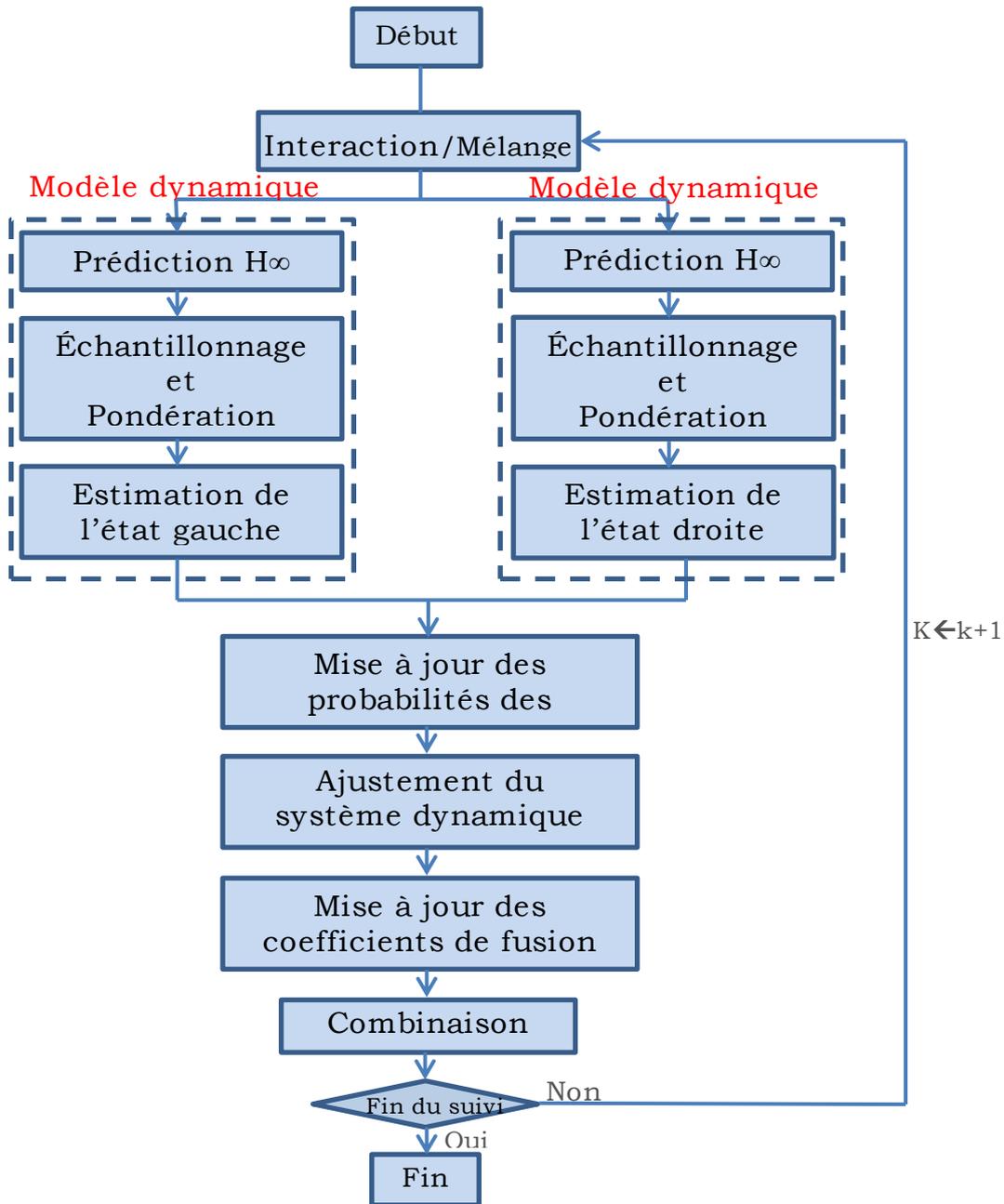


Figure IV-3. Organigramme de la méthode proposée

Finalement dans l'étape de combinaison, les estimations associées à chaque sous modèles sont combiner Equation IV-44 selon leurs probabilités des modèles. Le pseudo code de la méthode proposé est résumé par Algorithme IV-1.

$$\hat{X}_k = \sum_{j=1}^2 \hat{X}_k^j \mu_k^j \quad (\text{IV.44})$$

Algorithme IV.1 : Notre approche de suivi visuel

Initialisation :

- Sélectionner l'objet et obtenir le cadre initiale $\{x_0, y_0, w, h\}$.
- Calculer le model d'apparence
 - Calculer la fonction densité de probabilité basée sur la couleur section III.3.1.
 - Calculer la fonction densité de probabilité basée sur la texture section III.3.2.
- Initialiser le système dynamique $M_k = \{\beta_k^l, \beta_k^r\}$ basé sur le modèle de Single :

$$M_0 = \{\beta_0^l = 0,05, \beta_0^r = 7\}$$

Itération:

Pour $k=2, \dots, NF$ **Faire**

- Interaction / Mélange
 - Calculer l'état interactif \bar{X}_{k-1}^j pour $j=\{l,r\}$ Equation IV-37.
 - Calculer la covariance de l'état interactif \bar{P}_{k-1}^j pour $j=\{l,r\}$ Equation IV-38.
- Filtrage H ∞
 - Calculer la prédiction globale pour le modèle gauche et le modèle droite \hat{X}_k^j Equation IV-14.
 - Calculer la covariance de l'état de prédiction \hat{P}_k^j Equation IV-9.
- Filtrage Particulaires
 - Échantillonner un NP candidats pour chaque sous modèle dynamique, autour de chaque état global \hat{X}_k^j .

Pour $s=1, \dots, NP$ **Faire**

- Calculer les modèles d'observation $H_{\alpha-GMM}^s$ et DH_{ULBP}^s
- Calculer la distance de Bhattacharya Equation IV-25 et Equation IV-26.
- Fusionner les deux modèles pour calculer le poids de la particule ξ^s Equation IV-29.

Fin

- Calculer la prédiction locale \hat{X}_k^j Equation IV-31.
- Mise à jour de probabilité de modèle μ_k Equation IV-42.

- Ajuster le système dynamique $M_k = \{\beta_k^l, \beta_k^r\}$ utilisant les règles logiques de la section IV.2.3.
- Ajuster le poids de chaque modèle d'apparence Equation IV-36.
- Calculer l'état estimé \hat{X}_k en combinant les prédictions locales \hat{X}_k^j associées respectivement aux modèles gauche et droit Equation IV-44.

Fin

IV.5. Expérimentations et discussions

IV.5.1. Séquences vidéo

Pour évaluer la performance de notre algorithme de suivi, une série de tests ont été effectués sur différentes séquences vidéos disponible sur des bases publiques: CAVIAR¹ et Benchmark². Douze séquences réelles ont été utilisées, qui représentent des aspects difficiles dans le suivi visuel Figure IV-4, à savoir variation d'éclairage, encombrements de fond, flou de mouvement, mouvement rapide et de manœuvre. La longueur des séquences varie entre 73 et 1490 images. Tableau IV-3 présente les différents aspects du suivi pour chaque séquence.

Dans toutes les expériences, le nombre de particules est fixé à 20, λ_1 et λ_2 sont initialisés à 0.5 et la valeur -5 pour le paramètre α s'est révélée être constamment optimale. La méthode proposée a été implémentée et testée sur un ordinateur portable dont la vitesse est de 2.10 GHz avec une mémoire RAM de 3 Go.

Les performances de notre méthode de suivi sont comparées à celles des méthodes discutées dans l'état de l'art: filtre à particules de Kalman hiérarchique (HKPF) [Yin, 2011], filtre à particule interactif multiple modèles (IMM-PF) [Dou, 2014] et suivi visuel par probabilité spatialement pondérée de mélanges gaussiens (WLTMS) [Vas, 2015].

¹ EC Funded CAVIAR project/IST 2001 37540 : <http://homepages.inf.ed.ac.uk/rbf/CAVIAR/>

² Visual Tracker Benchmark: http://cvlab.hanyang.ac.kr/tracker_benchmark/datasets.html

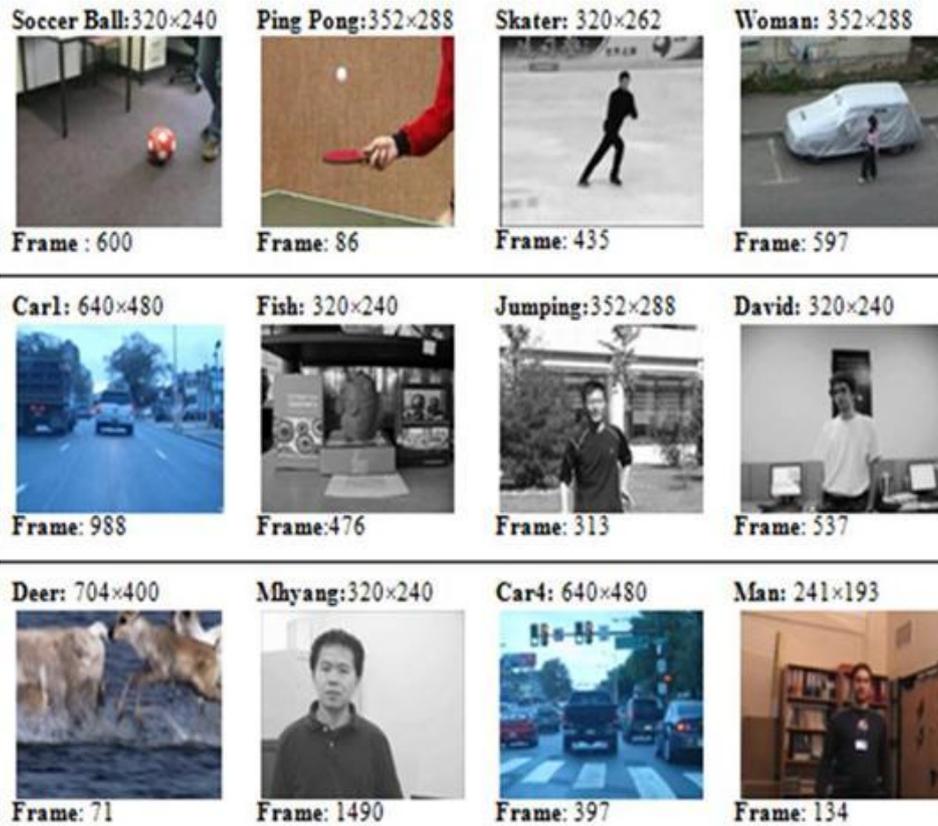


Figure IV-4. Séquences vidéo utilisées dans les tests des expérimentations

Tableau IV-3. Aspects de suivi des séquences vidéo utilisées dans les expérimentations

Séquence	Taille	Aspects de suivi
Soccer Ball	600	Changement brusque orientation, mouvement variable
Ping-Pong	86	Mouvement rapide, Changement brusque orientation
Skater	435	Déformation, mouvement complexe
Woman	597	Encombrement de fond, changement de luminosité, occlusion
Carl	988	Mouvement de camera, mouvement rapide
Fish	476	Changement de la luminosité
Jumping	313	Changement brusque d'orientation, flou, mouvement rapide
David	537	Encombrement, changement de luminosité
Deer	71	Mouvement rapide, confusion
Mhyang	1490	Changement de la luminosité, confusion
Car4	397	Mouvement de caméra, mouvement rapide, changement de luminosité
Man	134	Changement de luminosité, changement brusque d'orientation

IV.5.2. Critères de performance

De nombreuses mesures d'évaluation des performances de suivi ont été proposées [Bah, 2011], généralement avec une comparaison avec la vérité du terrain, en tenant compte de la présence et de la position, de la largeur et de la hauteur de la cible.

Le premier critère est l'erreur de localisation du centre (ELC) définie par Equation IV-22, et qui est calculée entre le centre estimé et celui de la vérité de terrain.

Le deuxième critère est l'erreur de localisation moyenne normalisée (ELMN) donné par Equation IV-23, qui est définie comme la moyenne normalisée des distances entre les centres estimés à chaque image et ceux de la vérité terrain.

Le troisième critère est la précision de suivi moyenne (PSM) définie par Equation IV-44, cette la métrique indique à quel point les zones de délimitations estimées chevauchent avec la zone vérité de terrain. Un objet est considéré comme correctement suivi dans une image si le rectangle estimé couvre au moins 25% de la surface de la cible dans la vérité terrain.

$$PSM = \frac{1}{N_{frames}} \sum_i \frac{|O^i \cup GO^i|}{|O^i \cap GO^i|} \quad (IV.45)$$

Où O^i désigne la zone de délimitation de l'objet suivi dans l'image i , et GO^i désigne la zone de délimitation de la vérité terrain.

Le quatrième critère est le score F1 définie par Equation IV-45, cette métrique permet de mesurer la couverture moyenne de la zone de délimitation estimée et la zone de délimitation de la vérité terrain.

$$F1 = \frac{1}{N_{frames}} \sum_i 2 \frac{|p^i * r^i|}{|p^i + r^i|} \quad (IV.46)$$

Lequel p^i et r^i sont définis par $p^i = \frac{|O^i \cap GO^i|}{|O^i|}$; $r^i = \frac{|O^i \cap GO^i|}{|GO^i|}$

Le cinquième critère est le taux de réussite (TR), cette métrique est mesurée en comptant le nombre d'images favorables. Une mesure du score S définie par Equation IV-46 est calculée, en effet une image est considérée comme favorable, lorsque la surface de chevauchement entre la zone de délimitation estimée et celle de la vérité terrain est supérieure à la moitié de la surface totale occupée par les deux zones.

$$S = \frac{|O^i \cap GO^i|}{|O^i \cup GO^i|} \quad (IV.47)$$

IV.5.3. Résultats qualitatifs

Les résultats du suivi sur plusieurs images sont illustrés par la Figure IV-5. $AH_{\infty}PF$ et WLMS peuvent suivre de près la balle alors que les autres algorithmes réalisent un suivi avec moins de précision. Comme on peut le voir sur la Figure IV-6, la méthode proposée peut conserver une stabilité de suivi et montre une précision intéressante mieux que les méthodes de comparaison.

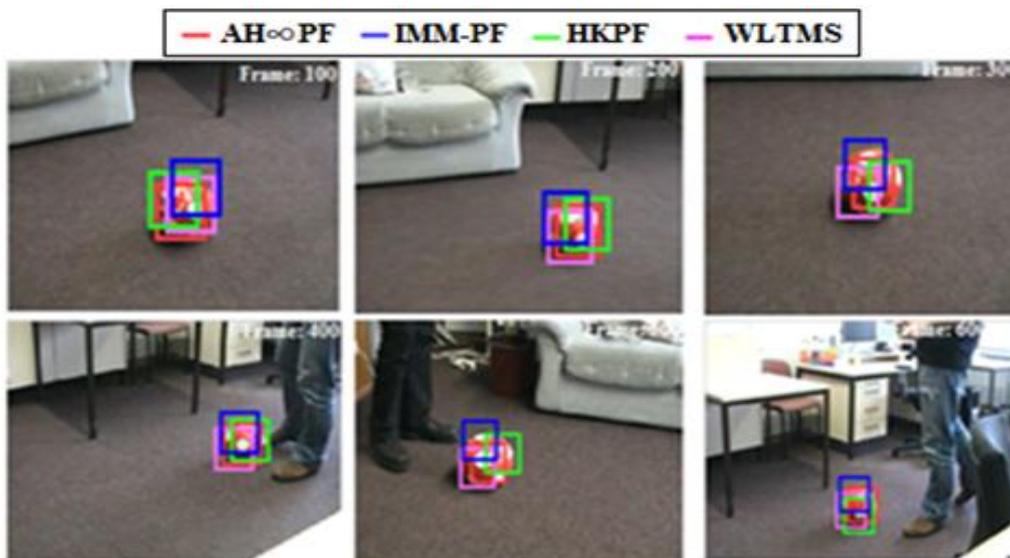


Figure IV-5. Résultats de suivi pour la séquence Soccer-Ball

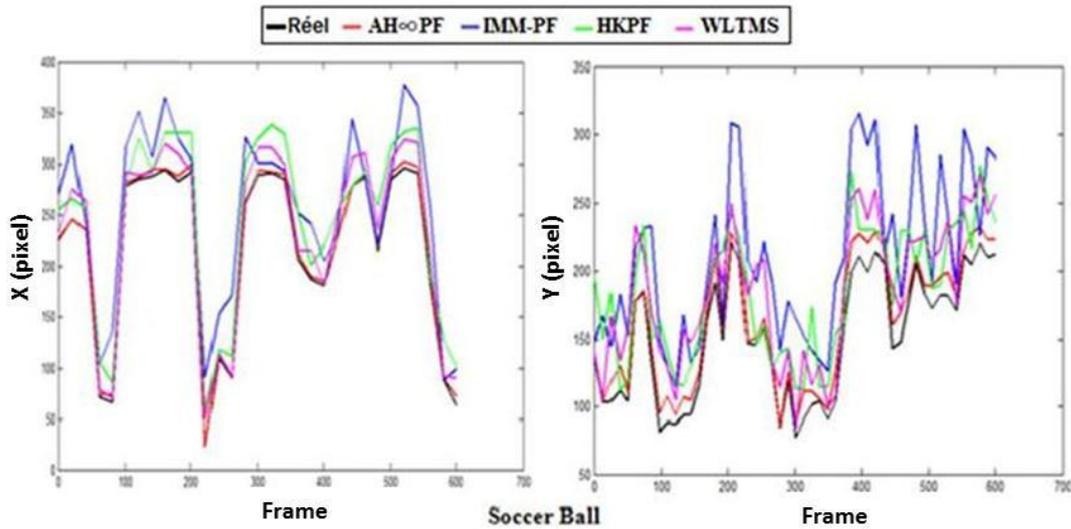


Figure IV-6. Trajectoires estimées et vérité du terrain pour la séquence Soccer-Ball

La deuxième séquence Ping-Pong, est caractérisée par un mouvement rapide et des changements soudains de direction, plusieurs images résultats sont visualisées par la Figure IV-7. La méthode proposée suit la balle avec précision tout au long de la séquence, tandis que les autres algorithmes n'arrivent pas à garder une trajectoire précise et surtout lorsque la cible change brusquement de direction. Dans la Figure IV-8, la méthode de suivi proposée peut conserver une trajectoire proche et stable par rapport à la vérité de terrain tandis que les trajectoires estimées par les méthodes de comparaison manquent de précision.

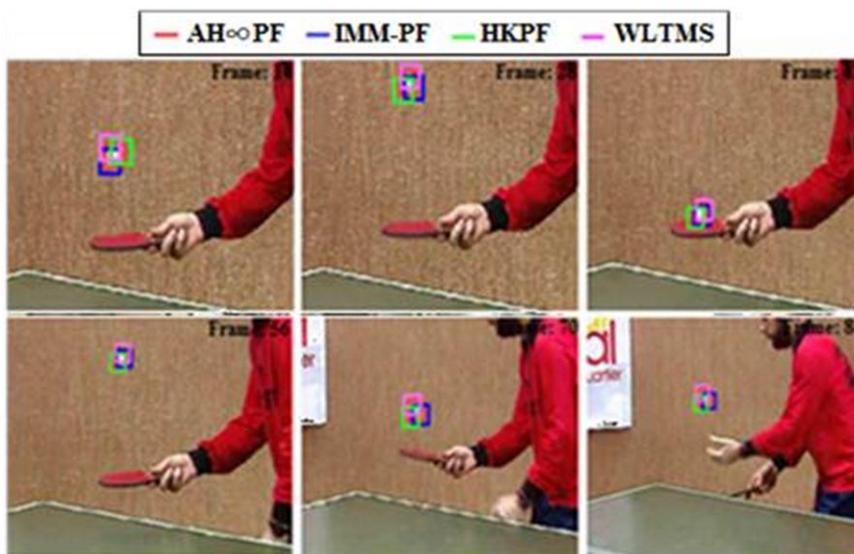


Figure IV-7. Résultats de suivi pour la séquence Ping-Pong

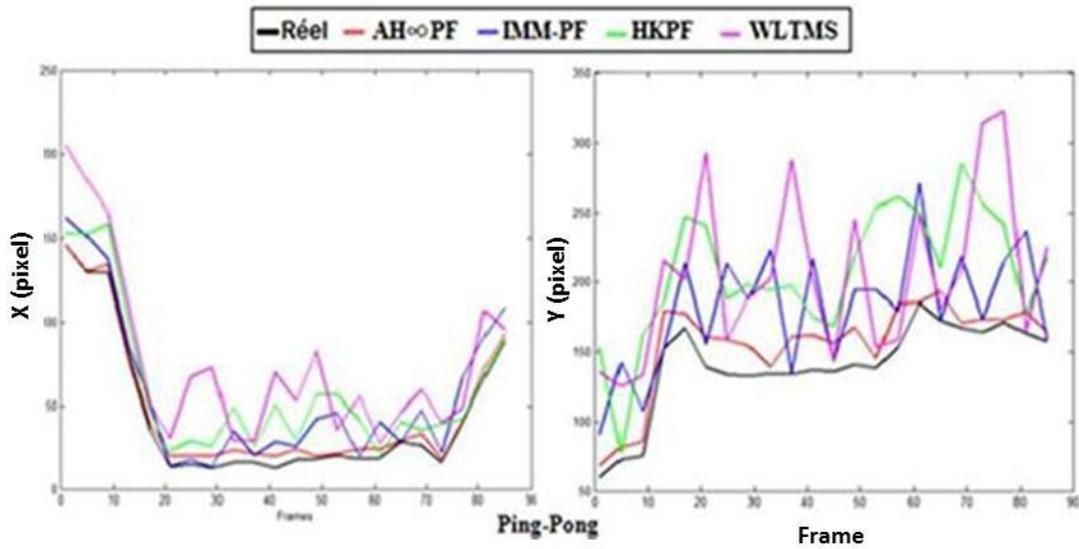


Figure IV-8. Trajectoires estimées et vérité du terrain pour la séquence Ping Pong

Plusieurs images résultats pour la troisième séquence sont visualisées par la Figure IV-9. La séquence Skater est sujet de plusieurs conditions difficiles de suivi à savoir déformation, mouvement rapide ainsi que changement brusque de direction, les méthodes de comparaison fonctionnent mal, alors que $AH_{\infty}PF$ opère favorablement, les trajectoires estimées et vérité de terrain sont visualisés par la Figure IV-10. On peut voir que la trajectoire estimée par notre algorithme a une bonne stabilité par rapport aux autres méthodes.

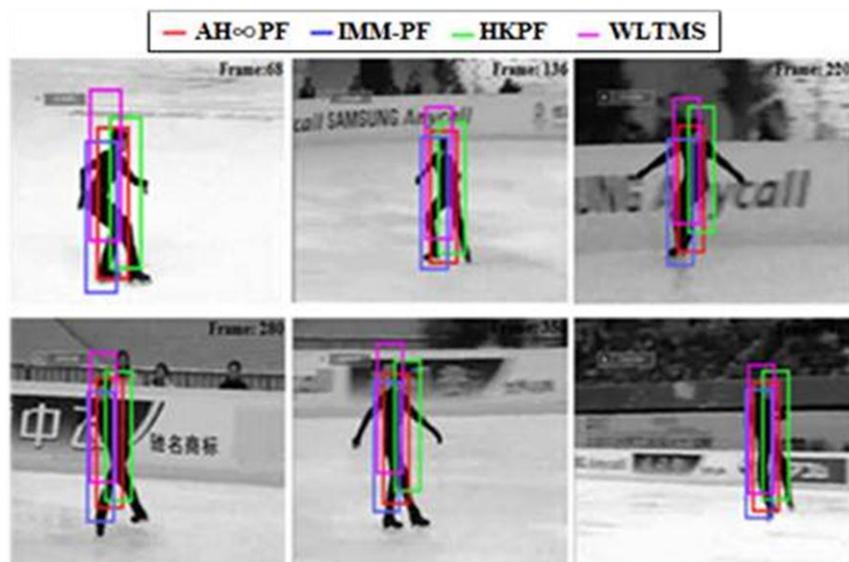


Figure IV-9. Résultats de suivi pour la séquence Skater

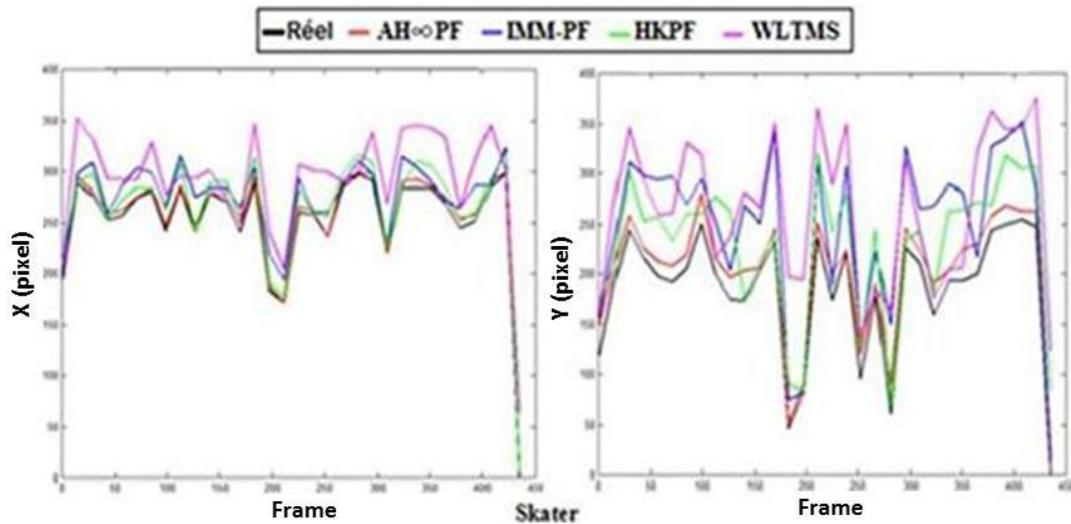


Figure IV-10. Trajectoires estimées et vérité du terrain pour la séquence Skater

Quatrième expérience est réalisée sur une séquence très difficile; Woman, la cible se déplace rapidement dans un environnement d'éclairage considérablement changeant, ainsi que les encombrements de fond, les images les plus significatives sont représentés sur la Figure IV-11, notre algorithme de suivi réalise un bon suivi mieux que WLTMS et IMM-PF, mais il montre moins de performance par rapport à HKPF; en effet, AH ∞ PF dérive légèrement de la trajectoire de vérité Figure IV-12, et surtout entre les images 90-200 et 400-550 du fait que l'objet est occlus, ce qui diminue la précision du suivi. Cependant, l'utilisation du filtre H ∞ adaptative a la capacité de récupérer le suivi après que la cible sort des occlusions.

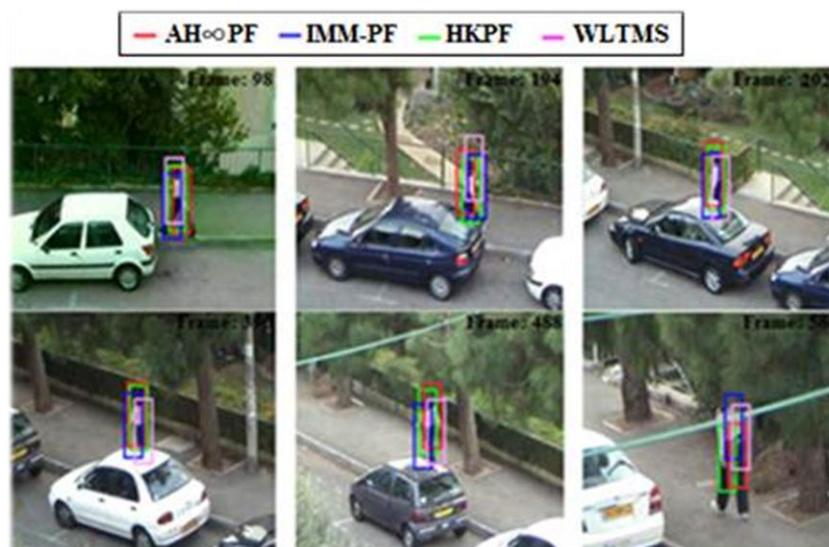


Figure IV-11. Résultats de suivi pour la séquence Woman

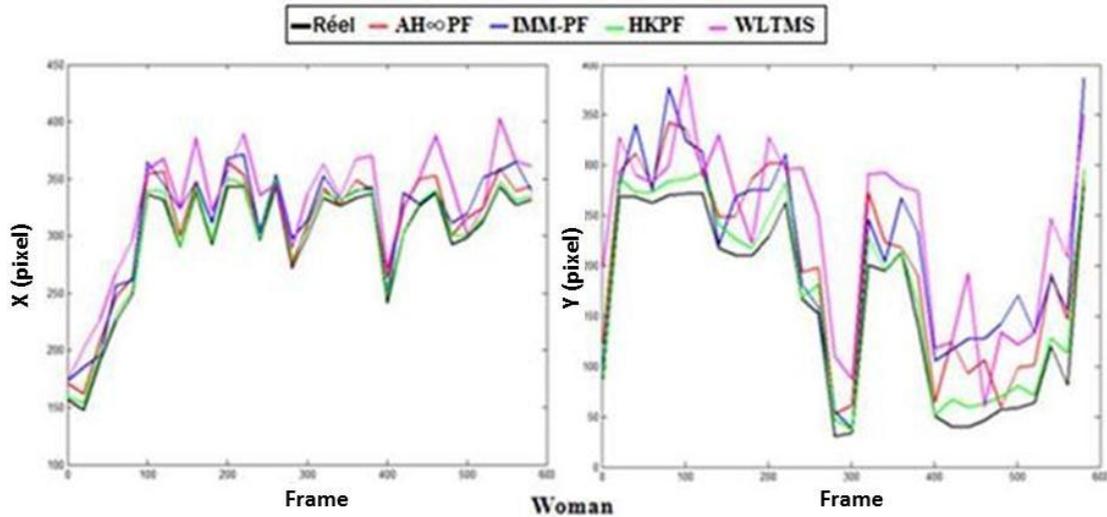


Figure IV-12. Trajectoires estimées et vérité du terrain pour la séquence Woman

Dans la cinquième séquence; Car1, une voiture roule vite, le déplacement de la cible en nombre de pixels est grand, et le mouvement de la cible est vraiment instable en raison du mouvement de la caméra, les résultats sur cette séquence sont visualisés par la Figure IV-13, notre algorithme de suivi gère favorablement le mouvement rapide et le changement fréquent de la position de la cible, alors que les méthodes de comparaison dérivent de la cible surtout lorsque la voiture change brusquement de position. La Figure IV-14 visualise les trajectoires estimées, notre estimation montre une trajectoire de suivi stable et proche de la vérité terrain, en effet $AH_{\infty}PF$ réalise un suivi meilleure que les autres méthodes.



Figure IV-13. Résultats de suivi pour la séquence Car1

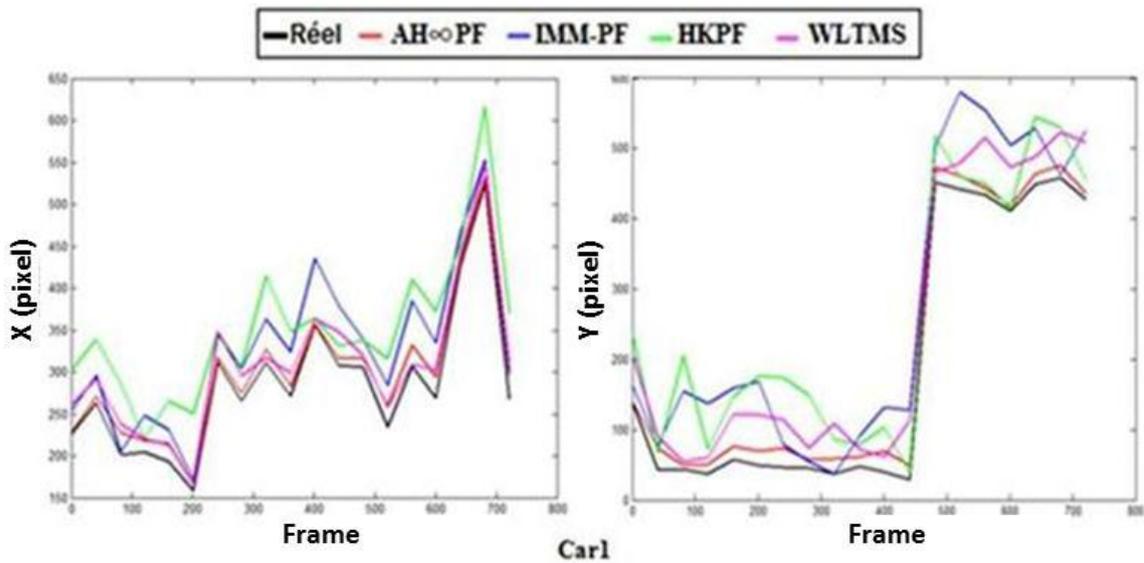


Figure IV-14. Trajectoires estimées et vérité du terrain pour la séquence Car1

Les résultats pour la sixième séquence; Fish, sont représentés par la Figure IV-15. Dans cette séquence, notre méthode est testée dans un environnement caractérisé par des variations d'illumination, notre algorithme de suivi montre une précision encourageante. Les différentes trajectoires estimées et la vérité de la terre sont visualisées par la Figure IV-16, la méthode proposée réalise un appariement favorable, tandis que les méthodes de comparaison n'arrivent pas à garder un suivi précis tout au long de la séquence.

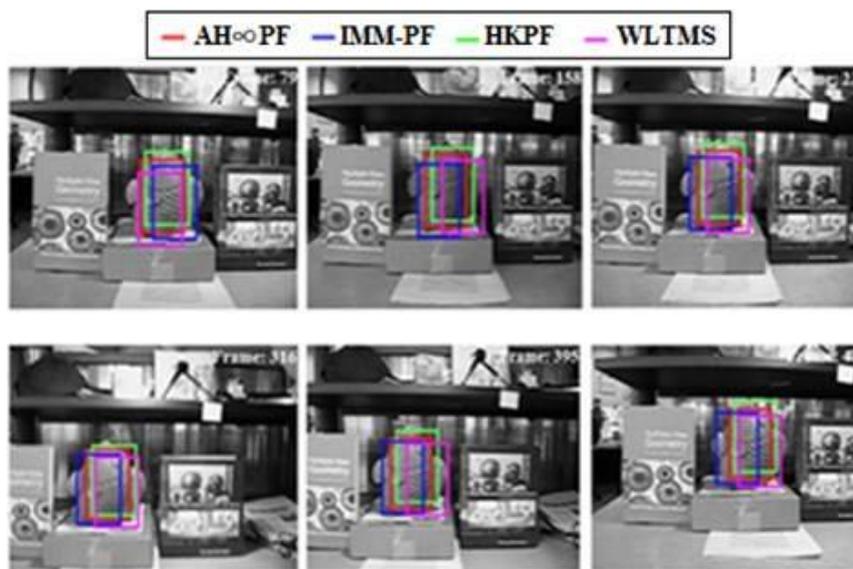


Figure IV-15. Résultats de suivi pour la séquence Fish

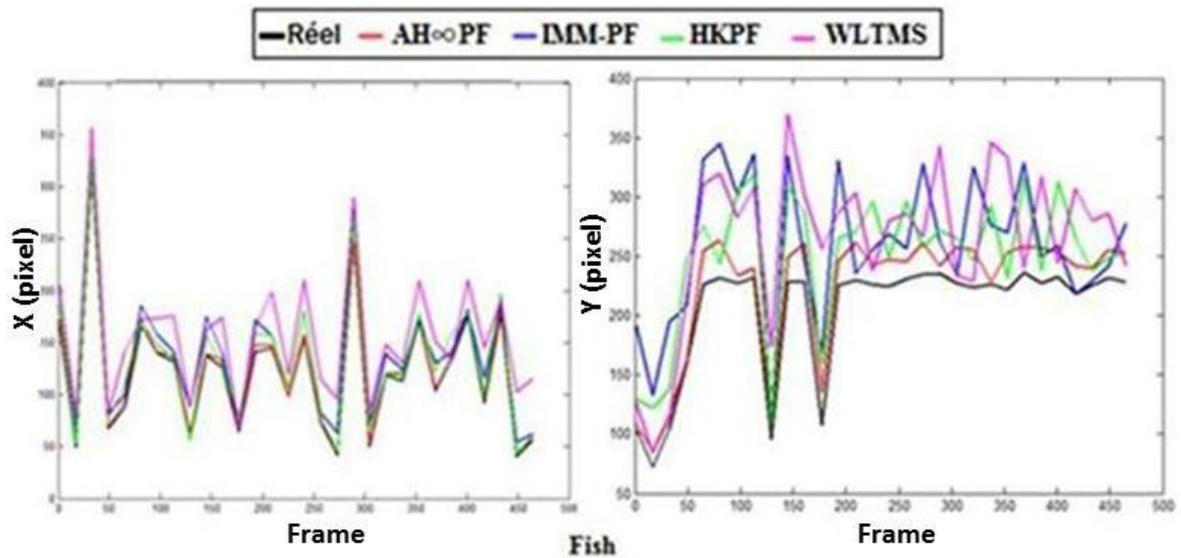


Figure IV-16. Trajectoires estimées et vérité du terrain pour la séquence Fish

La septième séquence; Jumping, une personne saute la corde, la séquence présente des situations de changement de direction fréquent et brusque ainsi que mouvement rapide et flou, la Figure IV-17 visualise les résultats du suivi, l'algorithme proposé est capable de garder une trajectoire précise par rapport aux autres méthodes qui dérivent loin de la position réelle, en particulier lorsque la cible change brusquement de direction, ces remarques sont clairement tirées des résultats de la comparaison des trajectoires avec la vérité de terrain présentés par la Figure IV-18.

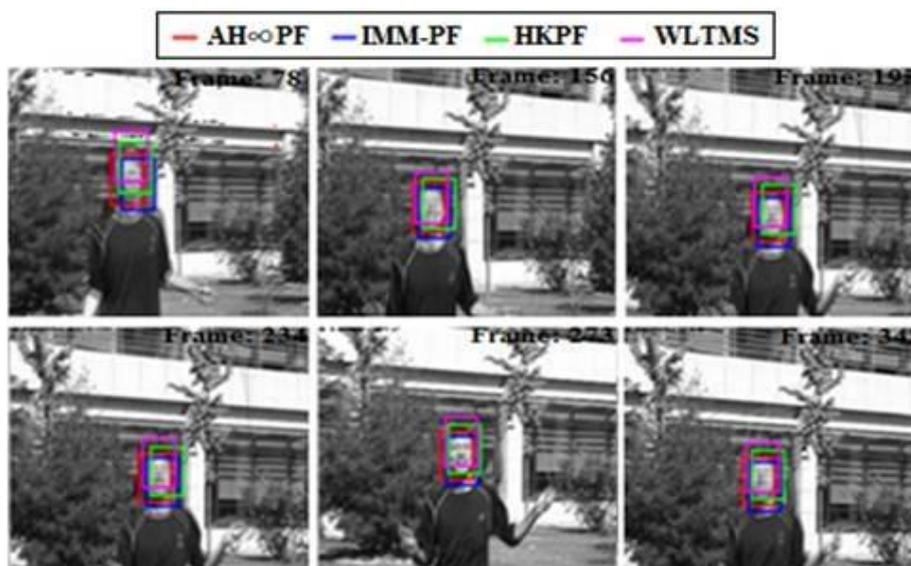


Figure IV-17. Résultats de suivi pour la séquence Jumping

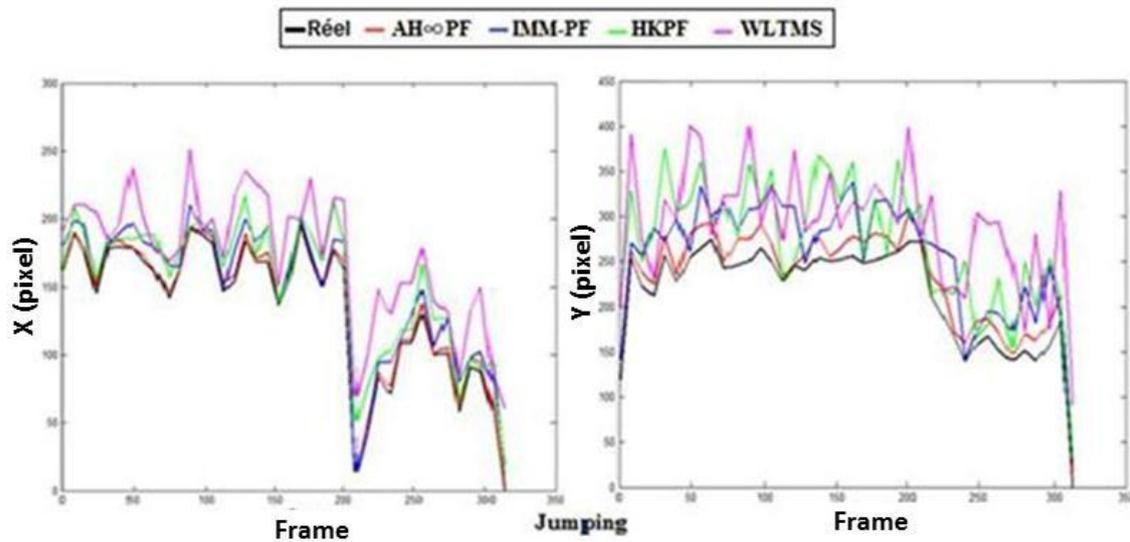


Figure IV-18. Trajectoires estimées et vérité du terrain pour la séquence Jumping

La Figure IV-19 présente les scénarios les plus significatifs dans la séquence; David. La séquence est caractérisée par de nombreuses conditions de suivi difficiles, à savoir des variations d'éclairage avec des échos de fond, ainsi que des changements brusque de direction. La maintenance cohérente de la trajectoire est assurée uniquement par HKPF et $AH_{\infty}PF$. Notre méthode montre une certaine faiblesse par rapport à HKPT, dans l'image 470 par exemple près d'un tiers de la zone réel est négligé et notre algorithme reste assez loin du centre de l'objet sans perdre le suivi, cette faiblesse est évidemment justifiée vue que la cible tourne dans le plan de l'image. La Figure IV-20, visualise les trajectoires estimées et celui de la vérité terrain. Dans l'ensemble, $AH_{\infty}PF$ garde une trajectoire stable et précise. Cependant, lorsque l'objet mobile commence à tourner dans le plan d'image, ce qui entraîne une déficience au niveau des informations de mouvement de l'objet mobile. Ce ci conduit à une déviation loin du vrai centroïde de la cible.

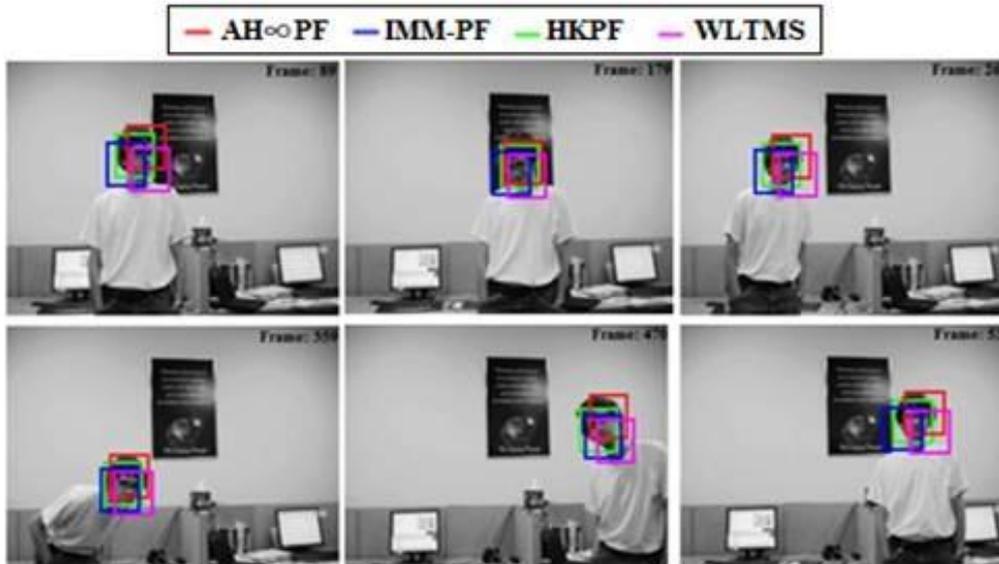


Figure IV-19. Résultats de suivi pour la séquence David

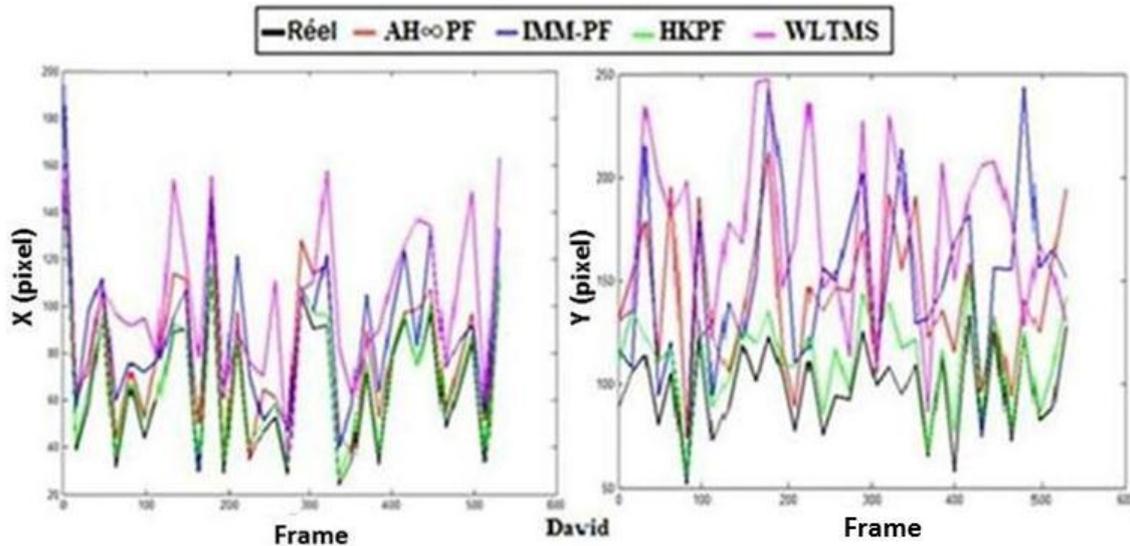


Figure IV-20. Trajectoires estimées et vérité du terrain pour la séquence David

Dans la neuvième séquence; Deer, Quelques images tirées du résultat du suivi sont visualisées par la Figure IV-21, la cible se déplace dans un environnement où l'arrière-plan au voisinage a la même apparence que l'objet suivi, le mouvement de la cible est rapide. Basé sur plusieurs modèles pour décrire l'état dynamique, $AH_{\infty}PF$ et IMM-PF réalisent un suivi précis, en comparaison avec les deux autres méthodes de comparaison qui n'arrivent pas à réaliser un suivi cohérent. De plus, la fusion de caractéristiques complémentaires permet à notre méthode d'être plus précise. Ceci est confirmé par les résultats visualisés par la Figure IV-22.

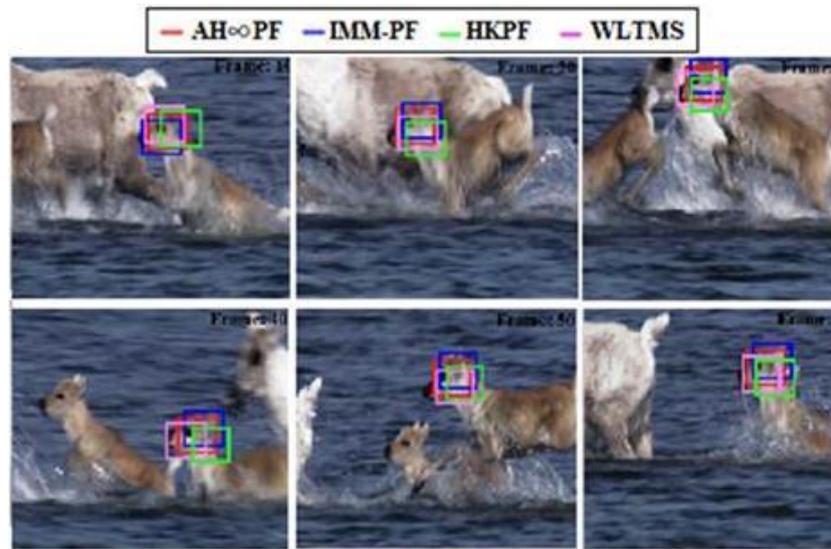


Figure IV-21. Résultats de suivi pour la séquence Deer

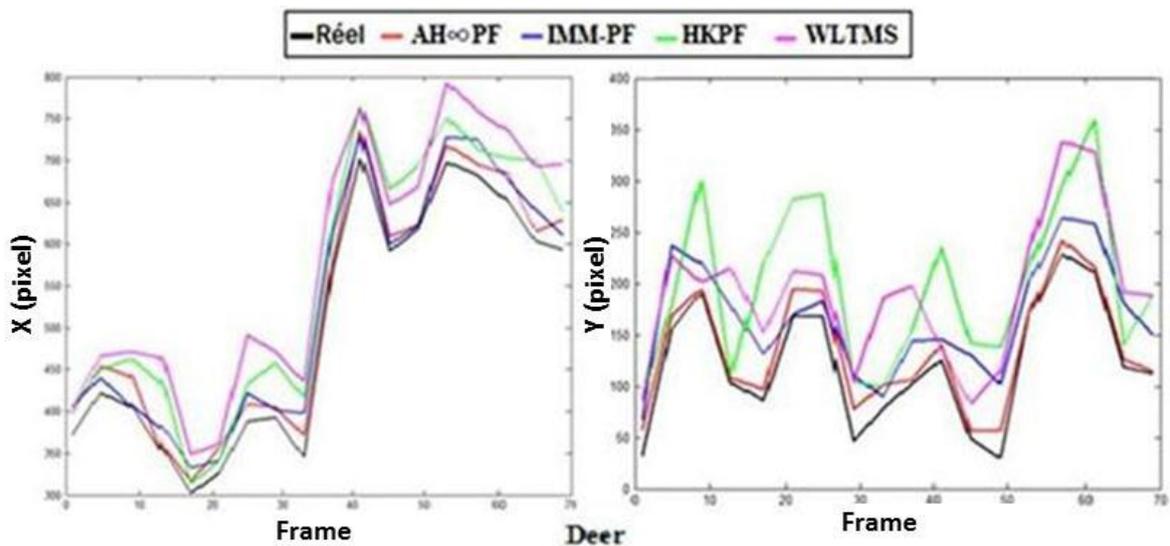


Figure IV-22. Trajectoires estimées et vérité du terrain pour la séquence Deer

La dixième séquence, Mhyang. Dans cette vidéo, la couleur de fond est similaire à la couleur du visage de la cible et la scène est caractérisée par des variations de la luminosité. Certaines images résultats sont visualisées par la Figure IV-23, les trajectoires estimées sont présentées par la Figure IV-24. Les algorithmes $AH_{\infty}PF$, IMM-PF sont capables de suivre la cible avec précision, mieux que les deux autres méthodes. Cependant, $AH_{\infty}PF$ commence à dériver loin du centroïde de la cible lorsque le visage de l'homme pivote hors du plan de l'image. Ceci est clairement visible sur la Figure IV-24, lorsque la courbe rouge représentant notre méthode commence à s'écarter loin de la courbe de vérité du terrain en noir.

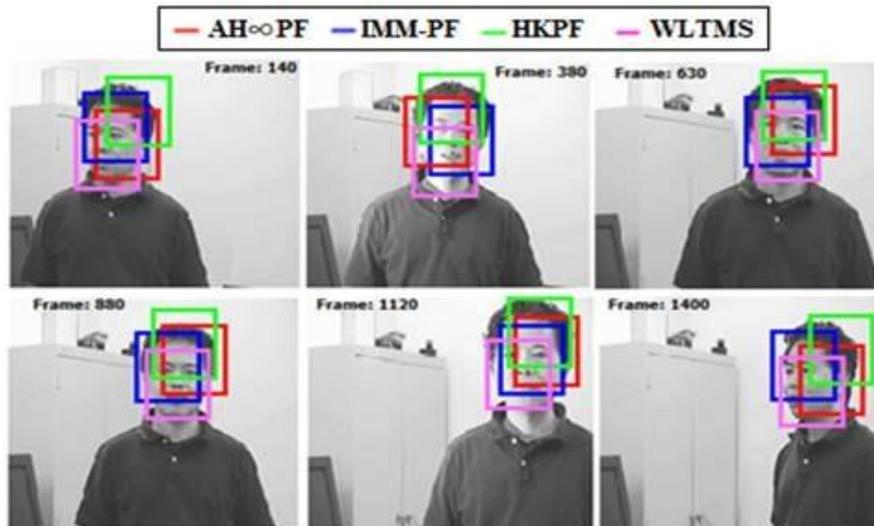


Figure IV-23. Résultats de suivi pour la séquence Mhyang

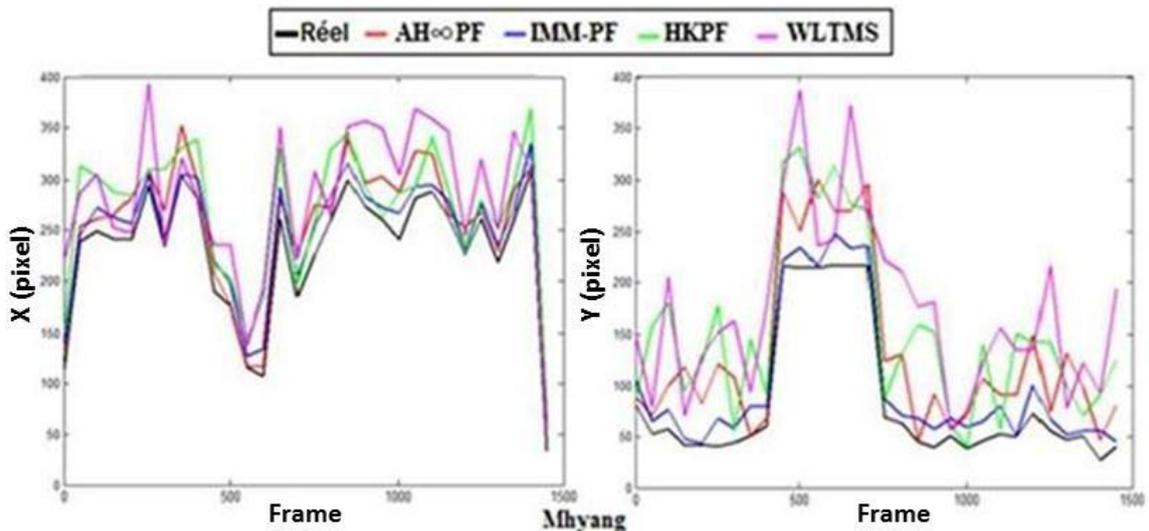


Figure IV-24. Trajectoires estimées et vérité du terrain pour la séquence Mhyang

Les images les plus significatives de la onzième séquence; Car4, sont représentés par la Figure IV-25, la voiture se déplace rapidement dans un environnement avec une variation significative de la luminosité. $AH_{\infty}PF$ réalise une meilleure performance par rapport aux autres algorithmes, en effet grâce au mode dynamique de commutation adopté et à l'ajustement automatique des paramètres de chaque sous-modèle, $AH_{\infty}PF$ a pu estimer la position de la cible malgré sa vitesse, en plus de la stabilité du modèle d'apparence permet de surmonter les conditions de variation de luminosité. Les trajectoires visualisées par la Figure IV-26 montrent que notre algorithme réalise un suivi assez proche durant toute la séquence, alors que les

méthodes de comparaison présentent une certaine imprécision, en particulier WMTLS et IMM-PF.



Figure IV-25. Résultats de suivi pour la séquence Car4

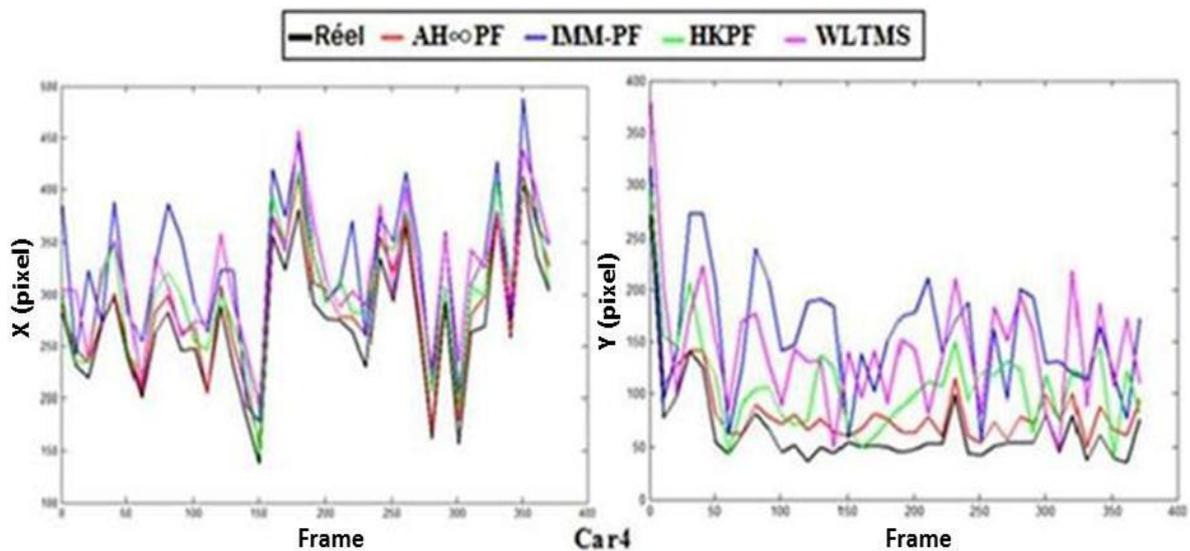


Figure IV-26. Trajectoires estimées et vérité du terrain pour la séquence Car4

La dernière séquence, Man. L'objet se déplace de façon clairsemée et brusque, l'éclairage de la région cible est considérablement varié, la Figure IV-27 montre quelques images représentatifs. $AH_{\infty}PF$ suit la cible de manière précise, mieux que les méthodes de comparaison, les méthodes HKPF et WLTMS montrent une certaine faiblesse surtout quand la cible change brusquement de direction comme le montre la Figure IV-28.

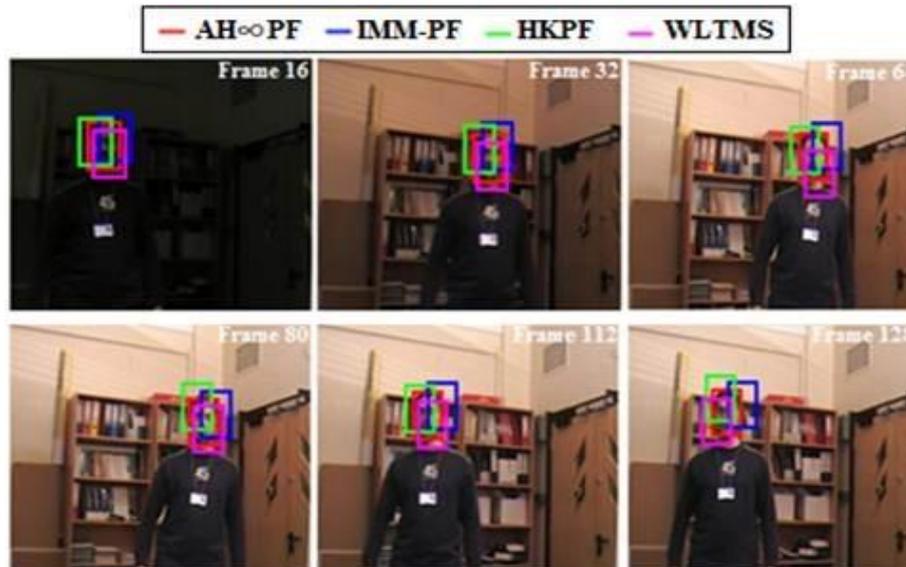


Figure IV-27. Résultats de suivi pour la séquence Man

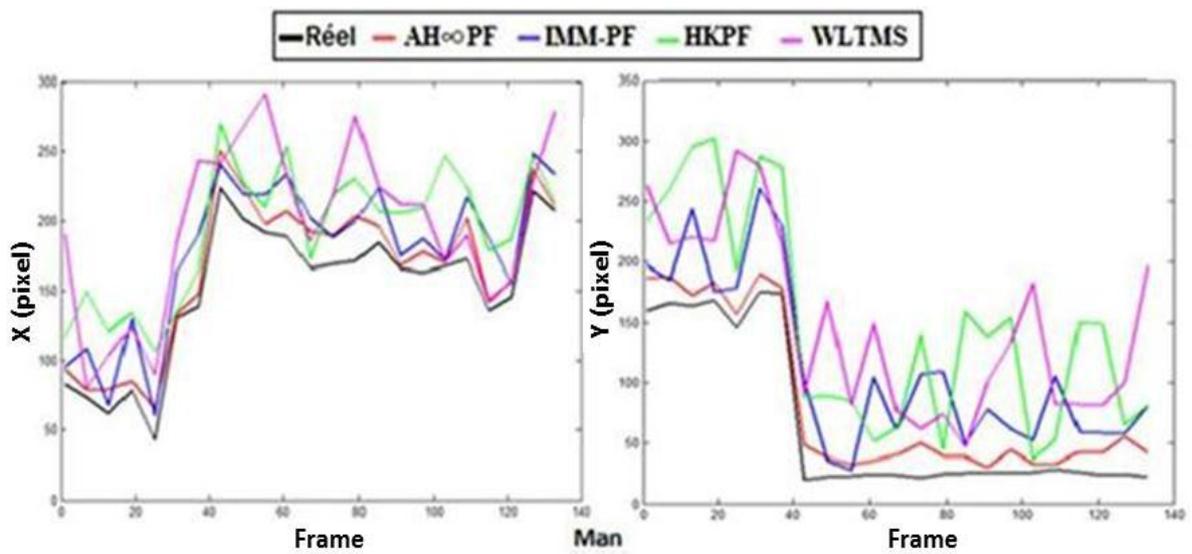


Figure IV-28. Trajectoires estimées et vérité du terrain pour la séquence Man

IV.5.4. Résultats quantitatifs

Les résultats représentant ELC sont visualisés par Figure IV-29 et Figure IV-30. Pour la plupart des séquences, notre méthode réalise un suivi précis avec une faible erreur, en effet la courbe d'erreur est sensiblement proche de l'axe zéro. Cependant, il y a des cas où notre méthode présente des faiblesses spécialement dans le cas des objets non rigides comme dans les séquences Woman, David et Mhyang, dans lesquels les positions relatives des pixels changent ainsi dans le cas où la cible tourne dans le plan de l'image. Dans ces conditions particulières, réacquérir la cible devient difficile par conséquent l'erreur augmente de manière significative.

Pour une meilleure évaluation des performances de notre algorithme. Nous avons réalisé les expériences 20 fois pour chaque séquence, ainsi la moyenne des résultats obtenus, est considérée pour calculer les critères : ELMN, PSM, F1 et TR. Les résultats quantitatifs obtenus pour les méthodes de comparaison sont présentés dans Tableau IV-3 à IV-6.

Dans toutes les séquences, $AH_{\infty}PF$ réalise un suivi précis. Dans l'ensemble, les mesures F1 sont supérieures à 50% et les mesures PSM sont inférieures à 2, ce qui signifie que les zones de délimitations estimées par la méthode proposée se chevauchent bien avec la vérité terrain, donc le centre de la cible estimée reste assez proche au centre de la vérité terrain, par conséquent une petite valeur en terme d'ELMN est obtenue ainsi qu'une valeur du TR proche de 1.

Dans la majorité des séquences, notre méthode a démontré une précision de suivi très intéressante mieux que les méthodes de comparaison. Dans les séquences où la cible se déplace rapidement, à savoir Car1, Deer et Car4, notre méthode réalise de meilleures performances en termes d'erreur de localisation et de précision de suivi moyenne.

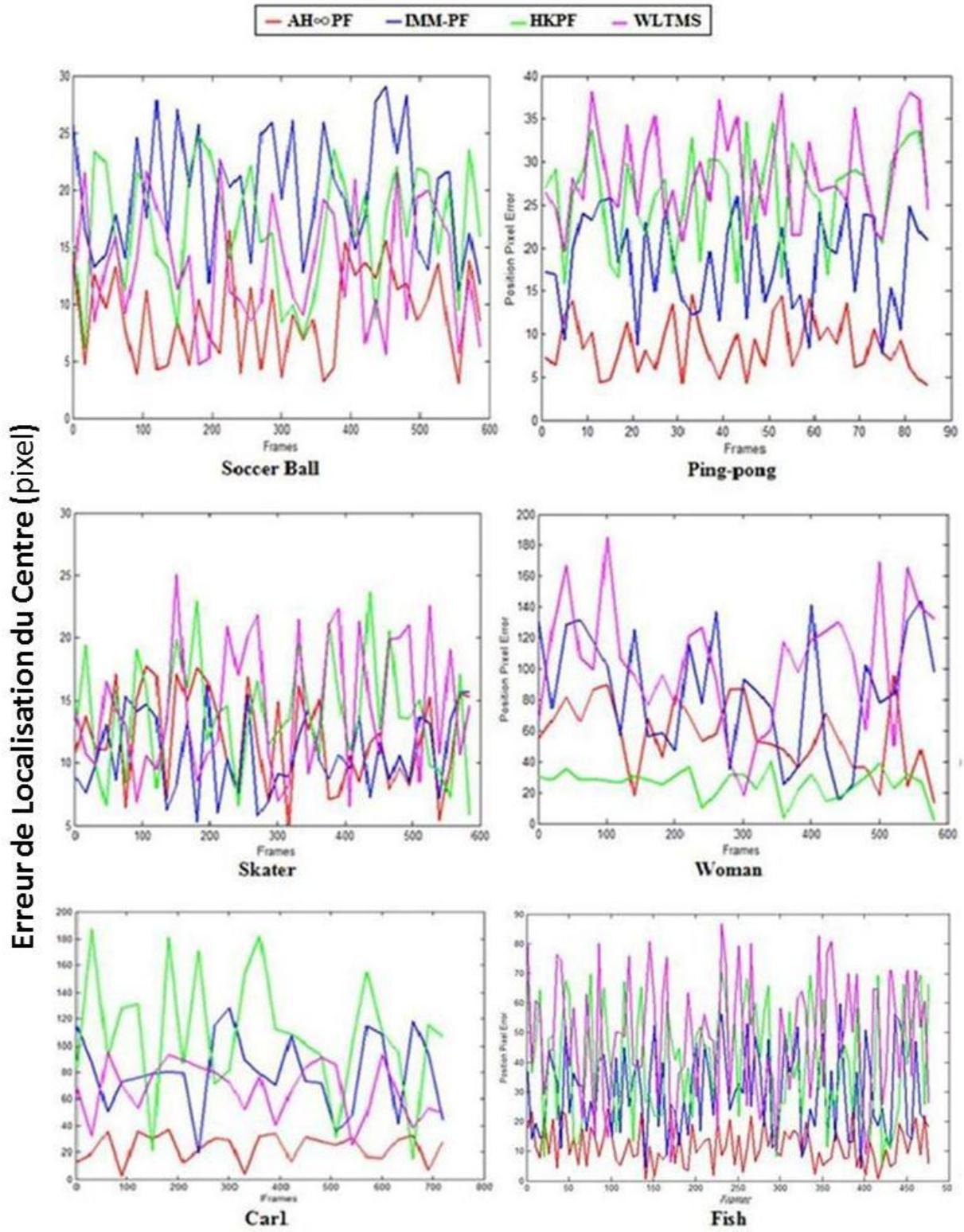


Figure IV-29. Performances de suivi par $AH_{\infty}PF$, de l'IMM-PF, du HKPF et du WLTS en termes d'erreur de position

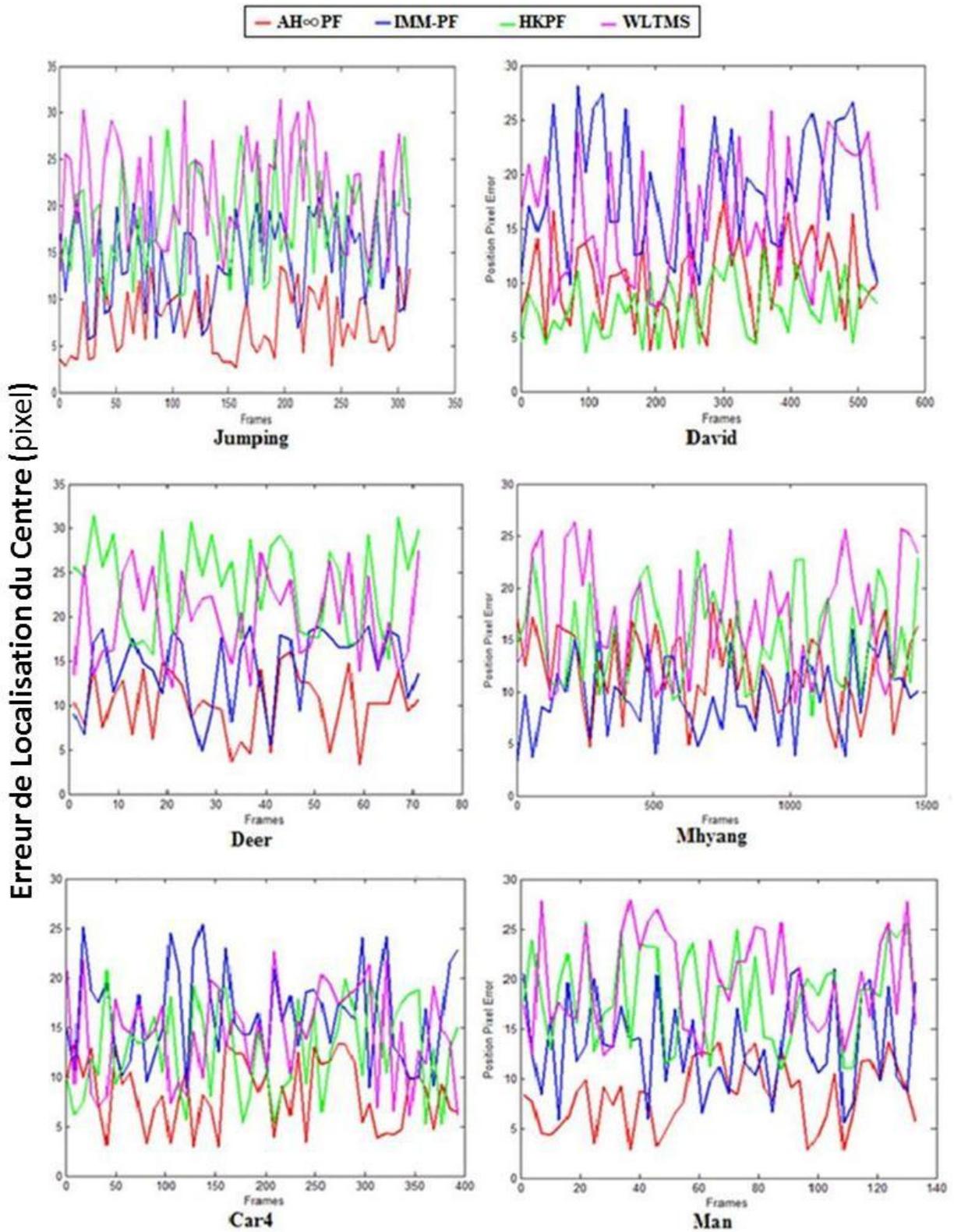


Figure IV-30. Performances de suivi par $AH_{\infty}PF$, de l'IMM-PF, du HKPF et du WLTS en termes d'erreur de position

En effet la couverture moyenne de la zone estimée dépasse 89% en termes de F1-score, en outre, notre méthode atteint un plus grand nombre d'images favorables avec un TR supérieur à 0,9. Dans le cas où la cible a un mouvement complexe, comme pour les séquences, Soccer Ball, Ping-Pong, Skater et Jumping, notre algorithme de suivi réalise une faible erreur de suivi en termes d'ELMN. Dans de nombreuses images, la zone estimée chevauche bien avec la zone vérité terrain, ainsi $AH_{\infty}PF$ produit les meilleurs résultats en termes de PSM qui est inférieur à 1,8 et un F1-score supérieur à 87%. Malgré les conditions de suivi difficiles, notre méthode suit avec succès les cibles le long des séquences utilisées dans les expérimentations, obtenant ainsi le meilleur nombre d'images favorables exprimées par un TR supérieur à 0,91. Ces résultats favorables sont bien justifiés, puisque notre méthode est basée sur un système dynamique adaptatif composé de deux sous modèles de Single paramétrés, dont chacun couvre un mode de mouvement différent.

En plus la stratégie d'ajustement automatique des paramètres du système dynamique permet de couvrir une large gamme de mouvements, ce qui donne à notre algorithme la possibilité de suivre la cible avec précision dans différentes conditions difficiles, à savoir le mouvement uniforme, rapide et complexe. De plus la modélisation de l'apparence de l'objet basé sur l'utilisation de plusieurs caractéristique permet de construire un modèle d'apparence, assez robuste et stable vis-à-vis des changements de la luminosité et de confusion de fond, ainsi dans les séquences Fish, Man, Car4 et Deer, dans lesquelles l'objet se déplace dans un environnement dont la luminosité change et la zone autour de la cible est confuse, notre algorithme marque de bonnes performances, produisant ainsi, la plus petite erreur en termes de ELMN. Bien qu'il y ait des cas où $AH_{\infty}PF$ a montré une certaine faiblesse sans en aucun cas perdre le suivi, comme dans les séquences Woman, David et Mhyang. Dans le cas des séquences David et Mhyang l'objet en mouvement à une forme non rigide dont les positions relatives des pixels changent, en plus la cible tourne hors plan, tandis que le cas de la séquence Woman, la cible est exposée à des occlusions. Les résultats numériques montrent que notre méthode n'est pas la meilleure, elle réalise une erreur en terme d'ELMN plus grande que HKPT pour les séquences Woman, David et plus grande que IMM-PF pour la séquence Mhyang, par exemple, entre les images 100-200 dans le cas de la séquence Woman, entre 130-160 pour la séquence de David et entre 800-1200 pour la séquence de Mhyang,

elle montre quelques perturbations et commence à dériver loin de la position vérité de terrain sans en aucun cas perdre le suivi, par conséquent, la surface de chevauchement entre la zone estimée et celle de la vérité terrain commence à diminuer, ce qui justifie les résultats numériques en termes de PSM et de score F1. Le nombre d'images favorables obtenues par $AH_{\infty}PF$ dans le cas des séquences Woman et David est inférieur à celui réalisé par le HKPF, et dans le cas de la séquence de Mhyang est inférieur à celui obtenu par l'IMM-PF, ce qui justifie le résultat numérique obtenu en termes de TR.

Nous terminons cette partie des expérimentations en présentant les temps moyens d'exécution par image pour les méthodes de comparaison, les résultats sont exposés dans le Tableau IV-7. Le temps de calcul enregistré par l'algorithme proposé est pour chacune des séquences proposées est encourageant, même si dans l'ensemble un peu plus élevé que WLTS et HKPF, en fait notre méthode utilise deux modèles sous-dynamique ajustés automatiquement et combinés dans le cadre de l'algorithme IMM, permettant ainsi de minimiser le nombre de modèles dynamiques adoptés généralement pour représenter le mouvement de la cible aussi l'application du filtre H_{∞} qui est plus robuste et plus rapide que le filtre de Kalman, réduit le nombre d'itérations, par conséquent $AH_{\infty}PF$ converge dans un délai raisonnable. En outre, le modèle d'apparence adopté basé sur la couleur et la texture préserve la précision et la stabilité de poursuite sans augmenter la complexité de calcul, vu que le nombre de composants α -GMM est 6 qui est très petit, en plus l'histogramme distinctif ULBP est calculé sur 59 bins au lieu de 256 pour l'histogramme LBP. Les deux caractéristiques sont fusionnées dans le cadre du PF avec seulement un très petit nombre de particules qui égal à 20.

Tableau IV-4. Précision de suivi. Erreur de localisation moyenne normalisée (ELMN) entre le centre cible suivi et le centre d'objet de vérité terrain est présenté pour les méthodes de comparaison.

Séquence	N_{Images}	Aspects difficiles	WLTMS	HKPF	IMM-PF	Ours
Soccer Ball	600	Changement brusque orientation, mouvement variable	0.2218	0.2613	0.4297	0.1437
Ping-Pong	86	Mouvement rapide, Changement brusque orientation	0.9675	0.7940	0.4862	0.3368
Skater	435	Déformation, mouvement complexe	0.7387	0.2364	0.4267	0.1378
Woman	597	Encombrement de fond, changement de luminosité, occlusion	0.6406	0.2069	0.5023	0.4505
Car1	988	Mouvement de camera, mouvement rapide	0.3477	0.5864	0.4302	0.1401
Fish	476	Changement de la luminosité	0.4469	0.2974	0.3362	0.2208
Jumping	313	Changement brusque d'orientation, flou, mouvement rapide	0.7865	0.5083	0.4192	0.2486
David	537	Encombrement, changement de luminosité	0.6443	0.2425	0.5176	0.4593
Deer	71	Mouvement rapide, confusion	0.4785	0.7211	0.3310	0.1977
Mhyang	1490	Changement de la luminosité, confusion	0.8327	0.6377	0.2233	0.3214
Car4	397	Mouvement de caméra, mouvement rapide, changement de luminosité	0.7198	0.4201	0.9419	0.2307
Man	134	Changement de luminosité, changement brusque d'orientation	0.6998	0.6401	0.5025	0.2142

Tableau IV-5. Précision du suivi. La précision de suivi moyenne (PSM) entre la fenêtre englobante estimée et celle de la vérité de terrain est présentée pour les méthodes de comparaison

Séquence	N_{Images}	Aspects difficiles	WLTMS	HKPF	IMM-PF	Ours
Soccer Ball	600	Changement brusque orientation, mouvement variable	1.4262	1.5824	2.3062	1.1626
Ping-Pong	86	Mouvement rapide, Changement brusque orientation	5.1478	3.7077	2.1947	1.7560
Skater	435	Déformation, mouvement complexe	3.2092	1.5800	2.1938	1.1035
Woman	597	Encombrement de fond, changement de luminosité, occlusion	2.8153	1.2857	2.8139	1.9187
Car1	988	Mouvement de camera, mouvement rapide	2.4788	2.2045	1.7599	1.1044
Fish	476	Changement de la luminosité	2.3978	1.8760	2.1042	1.5595
Jumping	313	Changement brusque d'orientation, flou, mouvement rapide	3.1060	2.8782	1.9634	1.6588
David	537	Encombrement, changement de luminosité	3.8279	1.5995	2.7128	1.7504
Deer	71	Mouvement rapide, confusion	2.1690	3.9651	1.6099	1.4672
Mhyang	1490	Changement de la luminosité, confusion	4.7756	2.6063	1.4771	1.6078
Car4	397	Mouvement de caméra, mouvement rapide, changement de luminosité	1.9423	1.6198	2.1194	1.5671
Man	134	Changement de luminosité, changement brusque d'orientation	4.8059	3.4318	2.2481	1.7099

Tableau IV-6. Précision de suivi. F1 score est présenté pour les méthodes de comparaison

Séquence	N _{Images}	Aspects difficiles	WLTMS	HKPF	IMM-PF	Ours
Soccer Ball	600	Changement brusque orientation, mouvement variable	84.12	72.51	67.59	87.38
Ping-Pong	86	Mouvement rapide, Changement brusque orientation	63.15	71.99	74.67	92.37
Skater	435	Déformation, mouvement complexe	69.17	75.40	74.98	89.89
Woman	597	Encombrement de fond, changement de luminosité, occlusion	65.42	90.37	81.39	72.52
Car1	988	Mouvement de camera, mouvement rapide	67.70	59.40	72.46	89.79
Fish	476	Changement de la luminosité	62.58	80.63	77.60	94.69
Jumping	313	Changement brusque d'orientation, flou, mouvement rapide	71.65	80.05	85.10	95.25
David	537	Encombrement, changement de luminosité	68.73	86.41	70.39	78.16
Deer	71	Mouvement rapide, confusion	58.41	69.28	81.52	93.25
Mhyang	1490	Changement de la luminosité, confusion	60.17	63.81	88.64	70.03
Car4	397	Mouvement de caméra, mouvement rapide, changement de luminosité	73.42	76.13	61.92	91.68
Man	134	Changement de luminosité, changement brusque d'orientation	67.51	70.83	77.42	86.04

Tableau IV-7. Précision du suivi. Taux de réussite (TR) pour les méthodes de comparaison

Séquence	N _{Images}	Challenging aspects	WLTMS	HKPF	IMM-PF	Ours
Soccer Ball	600	Changement brusque orientation, mouvement variable	0.85	0.77	0.76	0.95
Ping-Pong	86	Mouvement rapide, Changement brusque orientation	0.51	0.63	0.81	0.91
Skater	435	Déformation, mouvement complexe	0.38	0.84	0.80	0.94
Woman	597	Encombrement de fond, changement de luminosité, occlusion	0.39	0.92	0.56	0.82
Car1	988	Mouvement de camera, mouvement rapide	0.73	0.58	0.59	0.91
Fish	476	Illumination variation	0.36	0.55	0.41	0.92
Jumping	313	Changement brusque d'orientation, flou, mouvement rapide	0.32	0.67	0.84	0.95
David	537	Encombrement, changement de luminosité	0.42	0.95	0.48	0.74
Deer	71	Mouvement rapide, confusion	0.63	0.53	0.72	0.95
Mhyang	1490	Changement de la luminosité, confusion	0.40	0.62	0.87	0.71
Car4	397	Mouvement de caméra, mouvement rapide, changement de luminosité	0.86	0.93	0.73	0.96
Man	134	Changement de luminosité, changement brusque d'orientation	0.48	0.76	0.78	0.81

Tableau IV-1. Temps de calcul. Temps d'exécution moyens pour les méthodes de comparaison (s/image)

Séquence	N_{Frames}	WLTMS	HKPF	IMM-PF	Ours
Soccer Ball	600	0.0596	0.0461	0.0883	0.0602
Ping-Pong	86	0.0253	0.0422	0.0672	0.0361
Skater	435	0.1047	0.0975	0.1394	0.1378
Woman	597	0.0635	0.0804	0.1054	0.0743
Carl	988	0.1971	0.1802	0.4302	0.1910
Fish	476	0.1262	0.1154	0.1573	0.1323
Jumping	313	0.0566	0.0735	0.0985	0.0674
David	537	0.0498	0.0667	0.0985	0.0606
Deer	71	0.0580	0.0411	0.0832	0.0519
Mhyang	1490	0.0884	0.0776	0.1195	0.0945
Car4	397	0.2109	0.2001	0.2620	0.2170
Man	134	0.0420	0.0589	0.0839	0.0528

IV.6. Conclusion

Dans ce chapitre nous avons présenté une nouvelle approche de suivi visuel qui permet de surmonter les deux difficultés auxquelles on est confronté lors du développement d'un suivi visuel à savoir le mouvement rapide et complexe de la cible et le changement de l'apparence de l'objet tout au long de la séquence. L'idée de base est de faire coopérer deux sous-systèmes pouvant traiter aussi bien le comportement dynamique linéaire de la cible que le comportement local de la cible lorsqu'il se trouve non linéaire. Les deux sous systèmes sont modélisé par le modèle de S paramétré et adaptatif, capable

Pour chaque sous système, l'état de la cible à suivre est estimé par le filtre adaptatif H ∞ appliqué à un système dynamique. Alors que le module local gère le comportement local de la cible lorsqu'il se trouve non linéaire. Dans ce cas, l'état du système est estimé par le filtre à particules avec un modèle d'apparence faisant impliquer la fusion des caractéristiques de couleur et de texture.

L'étude expérimentale aussi bien qualitative que quantitative ainsi que l'étude comparative effectuée avec une série de séquences vidéo montrent les bonnes performances en termes de robustesse et de stabilité de notre approche dans différentes conditions de suivi.

Conclusion et perspectives

Dans ce rapport de thèse nous avons présenté une approche permettant un suivi stable et robuste des objets et des personnes dans des conditions difficiles. Les contributions proposées concernent la modélisation et l'estimation du système qui constituent les principales fonctions dans toute approche de suivi visuel.

La modélisation est établie par l'élaboration de deux sous modèles de Single paramétrés, le premier permet de décrire l'ensemble des mouvements de manœuvre et le second sous modèle permet de décrire l'ensemble des mouvements uniformes. L'estimation de l'état de l'objet en mouvement sous différentes contraintes est réalisée par une méthode de prédiction qui implique la combinaison séquentielle de deux filtres Bayésiens.

Le filtre H_∞ a été appliqué, sans aucune hypothèse sur les perturbations, pour l'estimation du mouvement global, supposé linéaire. Son utilisation permet d'atténuer la dégénération des particules surtout dans le cas où la cible a un mouvement complexe ou de manœuvre et permet ainsi de minimiser l'erreur d'estimation du pire cas. Tandis que le comportement local de la cible supposé non linéaire est estimé par le filtre particulaire qui est plus précis et gère efficacement les cas non linéaire et non gaussien. Le filtre à particules permet d'exploiter l'état dynamique global pour générer un faible nombre d'échantillons et ainsi de réduire le temps de calcul sans aucune perte dans la précision de suivi. Une stratégie de mise à jour adaptative, qui consiste à ajuster automatiquement les paramètres du système dynamique suivant des règles logiques, a permis de couvrir différents modes de mouvement (uniformes, complexes ou de manœuvres).

Un nouveau modèle d'apparence est élaboré, basé sur les caractéristiques visuelles de la couleur et de la texture. Le modèle proposé exploite à la fois les caractéristiques statistiques globales et locales de l'apparence de la cible de façon

complémentaire. L'histogramme pondéré des composantes H et V du mode HSV, a été modélisé par le mélange α -gaussien. L'information spatiale est prise en compte par l'utilisation, d'un masque avec un profil de noyau isotrope. L'histogramme distinctif basé sur le motif binaire local uniforme (ULBP) est construit sur la composante V du mode HSV, la zone objet est subdivisée en blocs et chaque bloc est décrit par un histogramme distinctif qui est définie comme la différence entre l'histogramme basé sur l'opérateur ULBP du bloc centrale est celui du bloc voisinage. Par la suite tous les histogrammes distinctifs sont concaténés pour former un seul descripteur de texture. Finalement les eux descripteurs sont fusionnés dans le cadre du filtre particulière. Les coefficients de fusion sont ajustés automatiquement suivant une analyse de la distribution spatiale des particules, ainsi le coefficient de variation en tant que paramètre de dispersion est utilisé pour évaluer la contribution de chaque caractéristique dans la précision du suivi.

Des expérimentations qualitatives et quantitatives ont été menées sur des bases de données publiques pour tester et évaluer les performances de la méthode proposée. Les séquences retenues dans les tests sont caractérisées par des conditions de suivi variées et complexe, telle que la variation d'éclairage, encombrements de fond, flou, mouvement rapide et de manœuvres. Les résultats des expérimentations ainsi que les comparaisons réalisées avec des méthodes récentes de la littérature ont validé la précision et la stabilité de notre algorithme de suivi visuel avec une complexité de calcul satisfaisante.

Notre approche de suivi basée sur un modèle dynamique paramétré s'avère une stratégie adaptative de prédiction robuste et précise de suivi de mouvement en présence de conditions difficiles notamment, dans les séquences vidéo où la cible à un mouvement complexe ou de manœuvre. En effet les résultats de simulation ont montré la puissance de cette stratégie à couvrir les différents les différents modes de mouvement, et l'efficacité à saisir les mouvements uniformes et de manœuvres. Cela est dû au fait que le modèle d'apparence proposé basé sur la fusion de la couleur et la texture dans le cadre du filtre à particule, assure bonne discrimination spécialement dans les conditions de changement de la luminosité, d'encombrement de l'arrière-plan ou de flou. Cela a contribué d'une part à l'amélioration de la robustesse et a garanti une bonne performance en termes de précision de suivi et une complexité de calcul raisonnable.

En perspective, nous envisageons de développer des approches qui intègrent:

- Des caractéristiques profondes et discriminantes pour la description de l'apparence de l'objet avec un apprentissage profond, et une stratégie de mise à jour en ligne du modèle.
- d'autres représentations de la cible telle que la structure élastique de parcelles locales pour gérer les occlusions partielles et les objets articulés et non-rigides.
- des stratégies basées sur l'apprentissage profond et en ligne nécessaires dans l'analyse et la reconnaissance du mouvement pour une description de haut niveau des actions et des interactions.
-

Revues internationals

Y. Dhassi, A. Aarab : Visual Tracking Based on Adaptive Mean Shift Multiple Appearance Models. Pattern Recognition and Image Analysis, Volume 8, Issue 3, pp 439–449, July 2018.

Y. Dhassi, A. Aarab : Visual tracking based on adaptive interacting multiple model particle filter by fusing multiples cues. Multimedia Tools and Applications, Volume 77, Issue 20, pp 26259–26292, October 2018.

Y Dhassi, A Aarab, M Alfid : Robust visual tracking based on H^∞ Particle filter by adaptively integrating multiple cues. International Journal of Imaging and Robotics, Volume 17, Issue Number 1, 2017.

Conferences internationals

Y Dhassi, A Aarab : Gradient descent optimization for visual tracking with geometrics transformation adaptation. The second International Conference on Intelligent Computing in Data Sciences (ICDS), Fes 2018.

Y Dhassi, A Aarab : Combined Mean Shift and Interactive Multiple Model for Visual Tracking by Fusing Multiple Cues. The 2nd International Conference of Mediterranean Symposium on Smart City Applications (SCAMS), Tanger 2017.

Y Dhassi, A Aarab, M Alfid : IMM-PF visual tracking based on multi-feature fusion. The 11th International Conference on Intelligent Systems: Theories and Applications (SITA), El Mohammadia 2016.

Y Dhassi, A Aarab, M Alfid : Object tracking using particle filter based on Lyapunov stability. The first International Conference on Intelligent Systems and Computer Vision (ISCV). Fès 2015.

Bibliographie

- [Arn, 2014] W. Arnold , Dung M. Chu, Rita Cucchiara. Visual Tracking: An Experimental Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume: 36, Issue: 7, July 2014.
- [Ash, 2017] C.S.Asha, A.V.Narasimhadhan, Robust infrared target tracking using discriminative and generative approaches, *Infrared Physics & Technology* Volume 85, Pages 114-127, September 2017.
- [Bah, 2011] Bahadir Karasulu, Serdar Korukoglu. A Software for Performance Evaluation and Comparison of People Detection and Tracking Methods in Video Processing. *Multimedia Tools and Applications*, Volume 55, Issue 3, pp 677–723, December 2011.
- [Bha, 1943] A. Bhattacharyya, On a measure of divergence between two statistical populations defined by their probability distributions, *Bull. Calcutta Math. Soc.* 99–109,1943.
- [Bin, 2014] Bineng Zhong, Yan Chen, Yingju Shen, Yewang Chen, Zhen Cui, Rongrong Ji, Xiaotong Yuan, Duansheng Chen, Weibin Chen. Robust tracking via patch-based appearance model and local background estimation. *Neurocomputing*, Volume 123, Pages 344-353, 10 January 2014.
- [Che, 2017] Chenwei Deng, BaoxianWang, WeisiLin, Guang-BinHuang, BaojunZhao. Effective visual tracking by pairwise metric learning. *Neurocomputing*, Volume 261, Pages 266-275, 25 October 2017.
- [Dav, 2008] A. David, Ross, Jongwoo Lim, Ruei-Sung Lin, and Ming-Hsuan Yang, “Incremental learning for robust visual tracking”, *International Journal of Computer Vision*, vol. 77, no. 1-3, pp. 125-141, 2008.
- [Dee, 2017] Deepak Ghimire, Joonwhoan Lee, Ze-Nian Li, Sunghwan Jeong, Recognition of facial expressions based on salient geometric features and support vector machines, *Multimedia Tools and Applications*, Volume 76, Issue 6, pp 7921–7946, March 2017.
- [Din, 2018] Dinesh Prasanna, M. Prabhakar, An efficient human tracking system using Haar-like and hog feature extraction, *Cluster Computing*, pp 1–8, January 2018.
- [Dou, 2014] Dou J, Li J, Robust visual tracking based on interactive multiple model particle filter by integrating multiple cues. *Neurocomputing* 135:118–129, 2014.
- [Fen, 2018] Feng Wu, Shao wu, Peng Jingkai Zhou, Qiong Liu, Xiaojia Xie. Object tracking via Online Multiple Instance Learning with reliable components. *Computer Vision and Image Understanding*, 11 April 2018.
- [Fou, 2013] Fouad Bousetouane, Lynda Dib, Hichem Snoussi, Improved mean shift integrating texture and color features for robust real time object tracking. *The Visual Computer*, Volume 29, Issue 3, pp 155–170, March 2013.
- [Fra, 2007] L. Frank, X. Iihua and P. Dan, *Optimal and Robust Estimation*, CRC Press. 2007.
- [H. Liu, 2011] H. Liu, F. Sun, M. Gao. Visual tracking using iterative sparse approximation, in: *Proceedings of the Eighth International Symposium on Neural Networks*, pp. 207–214, 2011.

-
- [Hak, 2012] Hakima Asaidi, Abdellah Aarab, Mohamed Bellouki, Shadow Detection Approach Combining Spectral and Geometrical Properties in Highway Video-Surveillance. *International Journal of Computer Applications* (0975 – 8887) Volume 53– No.17, September 2012.
- [Han, 2016] Hanxi Li, Yi Li, Fatih Porikli. DeepTrack: Learning Discriminative Feature Representations Online for Robust Visual Tracking. *IEEE Transactions on Image Processing*, Volume: 25, Issue: 4, April 2016.
- [Han, 2018] Y. Hang, C. Derong, G Jiulu, Object tracking using both a kernel and a non-parametric active contour model, *Neurocomputing*, March 2018.
- [Ho, 2004] J. Ho, K. Lee, M. Yang, and D. Kriegman, “Visual Tracking Using Learned Linear Subspaces,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 1, pp. 782-789, 2004.
- [Hon, 2018] Honghong Yang, Shiru Qu, Zunxin Zheng. Visual tracking via online discriminative multiple instance metric learning. *Multimedia Tools and Applications*, Volume 77, Issue 4, pp 4113–4131, February 2018.
- [Hua, 2016] Huanqiang Zeng, Jing Chen, Xiaolin Cui, Canhui Cai, Kai-Kuang Ma, Quad binary pattern and its application in mean-shift tracking, *Neurocomputing*, Volume 217, Pages 3-10, December 2016.
- [Hua, 2017] Huawei Lin, Yan Tong, Yu Cao, Youjie Zhou and Song Wang. Visual-Attention Based Background Modeling for Detecting Infrequently Moving Objects. *IEEE Transactions on Circuits and Systems for Video Technology*, Volume: 27, Issue: 6, June 2017.
- [Hui, 2016] Huilan Jiang, Jianhua Li, Dong Wang. Multi-feature tracking via adaptive weights. *Neurocomputing*, Volume 207, Pages 189-201, 26 September 2016.
- [Hwa, 2014] Hwan Kim, Hyo-Kak Kim, Seung-Jun Lee, Won-Jae Park, and Sung-Jea Ko, Kernel-Based Structural Binary Pattern Tracking, *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY*, VOL. 24, NO. 8, AUGUST 2014.
- [Ire, 2017] Irene Iswanto, Bin Li, Visual Object Tracking Based on Mean-shift and Particle-Kalman Filter, 2nd International Conference on Computer Science and Computational Intelligence, ICCSCI 2017, 13-14 October 2017.
- [Jas, 2015] W. Jasper, L.Agilandeeswari, K.Muralibabu. An Adaptive Approach for Validation in Visual Object Tracking. *Procedia Computer Science*, Volume 58, Pages 478-485, 2015.
- [Jia, 2012] X. Jia, H. Lu, M. Yang, Visual tracking via adaptive structural local sparse appearance model, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1822–1829, 2012.
- [Jia, 2017] Jianfang Dou, Qin Qin, Zimei Tu, Robust visual tracking based on generative and discriminative model collaboration, *Multimedia Tools and Applications*, Volume 76, Issue 14, pp 15839–15866, July 2017.
- [Jin, 2018] Jinhang Liu, Xian Zhong, An object tracking method based on Mean Shift algorithm with HSV color space and texture features, *Cluster Computing*, pp 1–12, February 2018.
- [Jun, 2014] Jungho Kim, Zhe Lin, In SoKweon. Rao-Blackwellized particle filtering with Gaussian mixture models for robust visual tracking. *Computer Vision and Image Understanding*, Volume 125, Pages 128-137, August 2014.
- [Jun, 2015] Jun Wang, Hanzi Wang, Yan Yan. Robust visual tracking by metric learning with weighted histogram representations. *Neurocomputing*, Volume 153, Pages 77-88, 4 April 2015.
- [Jun, 2016] Jun Wang, Yuehuan Wang, Multi-period visual tracking via online DeepBoost learning, *Neurocomputing*, Volume 200, Pages 55-69, , 5 August 2016.
-

-
- [Jun, 2018] Jun Wang, Weibin Liu, Weiwei Xing, Shunli Zhang, Visual object tracking with multi-scale superpixels and color-feature guided Kernelized correlation filters, *Signal Processing: Image Communication*, Volume 63, Pages 44-62, April 2018.
- [Jung, 2014] Junghyun Kwon, Hee Seok Lee, Frank C. Park, and Kyoung Mu Lee, A Geometric Particle Filter for Template-Based Visual Tracking, *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, VOL. 36, NO. 4, APRIL 2014.
- [Kai, 2013] Kaihua Zhang, Lei Zhang and Ming-Hsuan Yang. Real-Time Object Tracking via Online Discriminative Feature Selection. *IEEE Transactions on Image Processing*, Volume: 22, Issue: 12, Page(s): 4664 – 4677, Dec. 2013.
- [Kan, 2018] Kang Li, Fa-Zhi He, and Hai-Ping Yu, Robust Visual Tracking Based on Convolutional Features with Illumination and Occlusion Handling, *J. Comput. Sci. & Technol*, Vol.33, No.1, Jan. 2018.
- [Kana, 2018] S. Kanagamalliga, S. Vasuki, Contour-based object tracking in video scenes through optical flow and gabor features, *Optik* Volume 157, Pages 787-797, March 2018.
- [Ke, 2013] Ke Lu, Zhengming Ding, Sam Ge. Locally connected graph for visual tracking. *Neurocomputing*, Volume 120, Pages 45-53, 23 November 2013.
- [Kum, 2016] S.Kumar, Jigyendra SenYadav.Video object extraction and its tracking using background subtraction in complex environments. *Perspectives in Science*, Volume 8, Pages 317-322, September 2016.
- [Lei, 2013] Lei Qin, Hichem Snoussi, Fahed Abdallah, Cascaded Generative and Discriminative Learning for Visual Tracking, *International Conference Image Analysis and Recognition ICIAR: Image Analysis and Recognition* pp 397-406, 2013.
- [Li, 2004] Y. Li, “On Incremental and Robust Subspace Learning” *Pattern Recognition*, vol. 37, no. 7, pp. 1509-1518, 2004.
- [Li, 2008] X. Li, W.M. Hu, Z.F. Zhang, X.Q. Zhang, M.L. Zhu, and J. Cheng, “Visual Tracking via Incremental Log-Euclidean Riemannian Subspace Learning,” *Proc. IEEE Int’l Conf. Computer Vision and Pattern Recognition*, pp. 1-8, June 2008.
- [Li, 2011] H. Li, C. Shen, Q. Shi, Real-time visual tracking using compressive sensing, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1305–1312, 2011.
- [Lin, 2010] H. Ling, L. Bai, E. Blasch, X. Mei, Robust infrared vehicle tracking across target pose change using L1 regularization, in: *Proceedings of the 13th Conference on Information Fusion*, pp. 1–8, 2010.
- [Liu, 2011] B. Liu, J. Huang, L. Yang, C. Kulikowsk, Robust tracking using local sparse appearance model and K-selection, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1313–1320, 2011.
- [Liu, 2015] Liujun Zhao, Qingjie Zhao, Object Tracking via Combining Discriminative Global and Generative Local Models, *Pacific Rim Conference on Multimedia Advances in Multimedia Information Processing -- PCM* pp 570-579, 2015.
- [Liu, 2010] B. Liu, L. Yang, J. Huang, P. Meer, L. Gong, C. Kulikowski, Robust and fast collaborative tracking with two stage sparse optimization, in: *Proceedings of the 11th European Conference on Computer Vision*, pp. 624–637, 2010.
- [Luk, 2013] Luka Cehovin, Matej Kristan, Ales Leonardis. Robust Visual Tracking Using an Adaptive Coupled-Layer Visual Model. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume: 35, Issue: 4, Page(s): 941 – 953, April 2013.

-
- [Man, 2018] Manna Dai, Shuying Cheng, Xiangjian He, Hybrid generative–discriminative hash tracking with spatio-temporal contextual cues, *Neural Computing and Applications*, Volume 29, Issue 2, pp 389–399, January 2018.
- [Mar, 2017] Marjan Abdechiri, KarimFaez, Hamidreza Amindavar. Visual object tracking with online weighted chaotic multiple instance learning. *Neurocomputing* Volume 247, Pages 16-30, 19 July 2017.
- [Mat, 2010] Matej Kristan, Stanislav Kovacic, Ales Leonardis and Janez Pers, A Two-Stage Dynamic Model for Visual Tracking, *IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS*, VOL. 40, NO. 6, DECEMBER 2010.
- [Mee, 2017] Meenakshi Gupta, Swagat Kumar, Laxmidhar Behera, K.S. Venkatesh, A novel fusion framework for robust human tracking by a service robot, *Robotics and Autonomous Systems* Volume 94, Pages 134-147, August 2017.
- [Mei, 2009] X. Mei, H. Ling, Robust visual tracking using L1 minimization, *Proceedings of the 12th International Conference on Computer Vision*, pp. 1436–1443, 2009.
- [Min, 2016] Ming Zhao, Han-Ming Qian, Rong Ying-Jiao, Guo Chen, Robust object tracking via sparse representation based on compressive collaborative Haar-like feature, in: *Proceedings of the IEEE Conference on Audio, Language and Image Processing (ICALIP)*, 2016.
- [Min, 2017] Minjie Wan, Guohua Gu, Weixian Qian, Kan Ren, Qian Chen, Hue preservation based color space transformation for brightness-robust tracking, *Optik - International Journal for Light and Electron Optics*, Volume 144, Pages 26-33, September 2017.
- [Moh, 2017] D. Mohanapriya, Mahesh K, A novel foreground region analysis using NCP-DBP texture pattern for robust visual tracking, *Multimedia Tools and Applications*, Volume 76, Issue 24, pp 25731–25748, December 2017.
- [Moh, 2018] Mohammad Alkhatib, Adel Hafiane, Omar Tahri, Pierre Vieyres and Alain Delbos, Adaptive Median Binary Patterns for fully automatic Nerves Tracking in Ultrasound Images, *Computer Methods and Programs in Biomedicine*, March 2018.
- [Nag, 1991] Nagpal, K. M. and P. P. Khargonekar, “Filtering and Smoothing in an H_∞ Setting,” *IEEE Trans. Autom. Control*, AC-36(2), 152–166, 1991.
- [Oja, 2002] T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, *IEEE Trans. Pattern Anal. Mach. Intell.* 972–987, 2002.
- [Pen, 2016] Peng Zhang, Tao Zhuo, Yanning Zhang, Hanqiao Huang, Kangli Chen. Bayesian tracking fusion framework with online classifier ensemble for immersive visual applications. *Multimedia Tools and Applications*, Volume 75, Issue 9, pp 5075–5092, May 2016.
- [Pen, 2017] Peng Lv, Qingjie Zhao, Yanming Chen and LiuJun Zhao, Multiple cues-based active contours for target contour tracking under sophisticated background, *The Visual Computer*, Volume 33, Issue 9, pp 1103–1119, September 2017.
- [Pra, 2017] Prajna Parimita, Dipti Patra, Efficient visual tracking using multi-feature regularized robust sparse coding and quantum particle filter based localization, *Journal of Ambient Intelligence and Humanized Computing*, pp 1–14, December 2017.
- [Qin, 2017] Qingshan Liu, Jing Yang, Kaihua Zhang and Yi Wu. Adaptive Compressive Tracking via Online Vector Boosting Feature Selection. *IEEE Transactions on Cybernetics*, Volume: 47, Issue: 12, Page(s): 4289 – 4301, Dec. 2017.
-

-
- [Ron, 2004] X. Rong Li, V.P. Jilkov.: Survey of maneuvering target tracking. Part I. Dynamic models. *IEEE Transactions on Aerospace and Electronic Systems*, Volume: 39, pp. 1333-1364, January 2004.
- [Ros, 2008] D.A. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, "Incremental Learning for Robust Visual Tracking," *Int'l J. Computer Vision*, vol. 77, no. 2, pp. 125-141, May 2008.
- [Saa, 2018] Saadia Medouakh, Mohamed Boumechraz, Nadjiba Terki, Improved object tracking via joint color-LPQ texture histogram based mean shift algorithm, *Signal, Image and Video Processing*, Volume 12, Issue 3, pp 583-590, March 2018.
- [Sab, 2018] My Abdelouahed Sabri, Siham Aqel, Abdellah Aarab. A multiscale based approach for automatic shadow detection and removal in natural images. *Multimedia Tools and Applications*, pp 1-13, September 2018.
- [San, 2002] M. Sanjeev, Simon Maskell, Neil Gordon, and Tim Clapp. A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking. *IEEE TRANSACTIONS ON SIGNAL PROCESSING*, VOL. 50, NO. 2, FEBRUARY 2002.
- [Seo, 2017] Seon Heo, Hyung Il Koo and Nam Ik Cho, Open contour tracking using a new state space model and nonrigid motion training, *IEEE Transactions on Circuits and Systems for Video Technology*, Volume: 27, Issue: 11, Nov. 2017.
- [Sha, 2010] SM Shahed, Jeffrey Ho, and Ming-Hsuan Yang, "Online visual tracking with histograms and articulating blocks," *Computer Vision and Image Understanding*, vol. 114, no. 8, pp. 901-914, 2010.
- [She, 2017] Sheheryar Khan, Guoxia Xu, Raymond Chan, Hong Yan, An online spatio-temporal tensor learning model for visual tracking and its applications to facial expression recognition, *Expert Systems with Applications* Volume 90, Pages 427-438, December 2017.
- [Sho, 2008] J. Shotton, J. Winn, C. Rother, A. Criminisi, TextonBoost for image understanding: multi-class object recognition and segmentation by jointly modeling texture, layout, and context, *Int. J. Comput. Vision* 81 (1), 2008.
- [Sko, 2003] D. Skocaj and A. Leonardis, "Weighted and Robust Incremental Method for Subspace Learning," *Proc. Ninth IEEE Int'l Conf. Computer Vision*, vol. 2, pp. 1494-1501, Oct. 2003.
- [Ste, 2012] Stephan Liwicki, Stefanos Zafeiriou, Georgios Tzimiropoulos, Maja Pantic, Efficient Online Subspace Learning With an Indefinite Kernel for Visual Tracking and Recognition, *IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS*, VOL. 23, NO. 10, OCTOBER 2012.
- [Tha, 2014] Thang BaDinh, QianYu, GérardMedioni, Co-trained generative and discriminative trackers with cascade particle filter, *Computer Vision and Image Understanding*, Volume 119, Pages 41-56, February 2014.
- [Tzi, 2011] G. Tzimiropoulos, S. Zafeiriou, M. Pantic, Sparse representations of image gradient orientations for visual recognition and tracking, in: *Proceedings of the CVPR Workshop for Human Behaviour Analysis*, pp. 26, 2011.
- [Vas, 2011] Vasileios Karavasilis, Christophoros Nikou, Aristidis Likas. Visual tracking using the Earth Mover's Distance between Gaussian mixtures and Kalman filtering. *Image and Vision Computing*, Volume 29, Issue 5, Pages 295-305, April 2011.
- [Vas, 2015] Vasileios Karavasilis, Christophoros Nikou, Aristidis Likas. Visual tracking using spatially weighted likelihood of Gaussian mixtures. *Computer Vision and Image Understanding*, Volume 140, Pages 43-57, November 2015.

-
- [Vas, 2017] Vasileios Karavasilis, Christophoros Nikou, Aristidis, Likas. Real time visual tracking using a spatially weighted von Mises mixture model. *Pattern Recognition Letters*, Volume 90, Pages 50-57, 15 April 2017.
- [Ven, 2015] R. Venkatesh, Priti Parate, K. Aniruddha Achary, Robust Tracking with Interest Points: A Sparse Representation Approach, *Image and Vision Computing*, Volume 33, Pages 44-56, January 2015.
- [Wan, 2012] Q. Wang, F. Chen, W. Xu, M. Yang, Online discriminative object tracking with local sparse representation, in: *Proceedings of the IEEE Workshop on the Applications of Computer Vision*, pp. 425–432, 2012.
- [Wei, 2012] Weiming Hu, Xi Li, Wenhan Luo, Xiaoqin Zhang, Stephen Maybank, and Zhongfei Zhang. Single and Multiple Object Tracking Using Log-Euclidean Riemannian Subspace and Block-Division Appearance Model *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, VOL. 34, NO. 12, DECEMBER 2012
- [Wri, 2009] J. Wright, A. Yang, A. Ganesh, S. Sastry, Y. Ma, Robust face recognition via sparse representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 210–227, 2009.
- [Wu, 2009] Y. Wu, Cheng, J., Wang, J., Lu, H.: ‘Real-time visual tracking via incremental covariance tensor learning’. *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, Kyoto, Japan, pp. 1631–1638, September 2009.
- [Wu, 2011] Y. Wu, E. Blasch, G. Chen, L. Bai, H. Ling, Multiple source data fusion via sparse representation for robust visual tracking, in: *Proceedings of the 13th Conference on Information Fusion*, pp. 1–8, 2011.
- [Wu, 2013] Lu, K., Ding, Z.M., Ge, S.: ‘Locally connected graph for visual tracking’. *Neurocomputing*, 120, pp. 45–53, 2013.
- [Xi, 2013] Xi Li, Weiming Hu, Chunhua Shen, Zhongfei Zhang, Anthony Dick, Anton van den Hengel, A Survey of Appearance Models in Visual Object Tracking. *ACM Transactions on Intelligent Systems and Technology*, Volume 4 Issue 4, September 2013.
- [Xia, 2015] Xiaoyuan Yu, Jianchao Yang, Tianjiang Wang and Thomas Huang, Key Point Detection by Max Pooling for Tracking, *IEEE TRANSACTIONS ON CYBERNETICS*, VOL. 45, NO. 3, MARCH 2015.
- [Xue, 2012] Xueliang Zhan and Bo Ma. Gaussian Mixture Model on Tensor Field for Visual Tracking. *IEEE SIGNAL PROCESSING LETTERS*, VOL. 19, NO. 11, NOVEMBER 2012.
- [Yad, 2016] Dileep Yadav, Karan Singh. A combined approach of Kullback–Leibler divergence and background subtraction for moving object detection in thermal video. *Infrared Physics & Technology*, Volume 76, Pages 21-31, May 2016.
- [Yan, 2015] Yang He, Mingtao Pei, Min Yang, Yuwei Wu, Yunde Jia, Online visual tracking by integrating spatio-temporal cues, *IET Computer Vision*, Volume: 9, Issue: 1, 2015.
- [Yaw, 2017] Yawen Wang, Hongchang Chen, Shaomei Li, Jianpeng Zhang, Chao Gao, Object tracking by color distribution fields with adaptive hierarchical structure, *The Visual Computer*, Volume 33, Issue 2, pp 235–247, February 2017.
- [Yin, 2011] Yin S, Na JH, Choi JY, Oh S, Hierarchical Kalman-particle filter with adaptation to motion changes for object tracking. *Comput Vis Image Underst* 115:885–900, 2011.
- [Yon, 2017] Yongxing Yang, Jie Yang, Liyuan Liu and Nanjian Wu, High-Speed Target Tracking System Based on a Hierarchical Parallel Vision Processor and Gray-
-

-
- Level LBP Algorithm, IEEE Transactions on Systems, Man, and Cybernetics: Systems, Volume: 47, Issue: 6, June 2017.
- [Yon, 2016] Yongxing Yang, Jie Yang, Zhongxing Zhang, Liyuan Liu, Nanjian Wu, High-speed visual tracking with mixed rotation invariant description, Electronics Letters, Volume: 52, Issue: 7, 2016.
- [Yong, 2017] Yongliang Ma, An object tracking algorithm based on optical flow and temporal-spatial context, Cluster Computing, pp 1-9, December 2017.
- [Yua, 2012] Yuan Xie, Yanyun Qu, Cuihua Li, Wensheng Zhang, Online multiple instance gradient feature selection for robust visual tracking, Pattern Recognition Letters, Volume 33, Issue 9, Pages 1075-1082, July 2012.
- [Yun, 2008] Yun LeI, Xiaoqing Ding, Shengjin Wang, Visual Tracker Using Sequential Bayesian Learning: Discriminative, Generative, and Hybrid, IEEE Transactions on Systems, Man, and Cybernetics, Volume: 38, Issue: 6, Dec. 2008.
- [Yun, 2016] Yunyoung Nam, Real-time abandoned and stolen object detection based on spatio-temporal features in crowded scenes, Multimedia Tools and Applications, Volume 75, Issue 12, pp 7003-7028, June 2016.
- [Yur, 2012] Yuru Wang, Xianglong Tang, Qing Cui. Dynamic appearance model for particle filter based visual tracking. Pattern Recognition, Volume 45, Issue 12, Pages 4510-4523, December 2012.
- [Zeb, 2016] Zebin Cai, Zhenghui Gu, Zhu Liang Yu, Hao Liu, Ke Zhang, A real-time visual object tracking system based on Kalman filter and MB-LBP feature matching, Multimedia Tools and Applications, Volume 75, Issue 4, pp 2393-2409, February 2016.
- [Zha, 2009] S. Zhang, H. Yao, X. Sun, S. Liu, Robust object tracking based on sparse representation, in: Proceedings of the International Conference on Visual Communications and Image Processing, pp. 77441N-1-8, 2010.
- [Zha, 2010] S. Zhang, H. Yao, S. Liu, Robust visual tracking using feature-based visual attention, in: Proceedings of the IEEE International Conference on Acoustics, Speech, Signal Processing, pp. 1150-1153, 2010.
- [Zhe, 2016] Zhenhai Wang, Bo Xu, An effective object tracking based on spatio-temporal context learning and Hog, orally against others tracking approach based on kernel method in many complex conditions. Natural Computation (ICNC), 2015 11th International Conference on, January 2016.
- [Zhi, 2015] Zhiyong Li, Shuang He, Mervat Hashem, Robust object tracking via multi-feature adaptive fusion based on stability: contrast analysis, The Visual Computer, Volume 31, Issue 10, pp 1319-1337, October 2015.
- [Zho, 2012] W. Zhong, H. Lu, M. Yang, Robust object tracking via sparsity-based collaborative model, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1838-1845, 2012.