



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

Formation Doctorale : Sciences et Technologies de l'Information et de la Communication (STIC)

Discipline : Informatique

Spécialité : Informatique

Laboratoire : Informatique, Imagerie et Analyse Numérique (LIAN)

THESE DE DOCTORAT

Présentée par

LAAROSSI Khadija

Nouvelles approches de suivi par modèle d'apparence.

Soutenue le 16/12/2017 devant le jury composé de :

Pr. Ismail BOUMHIDI	Faculté des Sciences Dhar El Mahraz - Fès	Président
Pr. El Hassan SBAI	École Supérieure de Technologie - Meknès	Rapporteur
Pr. Hassan SATORI	Faculté Polydisciplinaire - Nador	Rapporteur
Pr. Hamid TAIRI	Faculté des Sciences Dhar El Mahraz - Fès	Rapporteur
Pr. Farid ABDI	Faculté des Sciences et Technique - Fès	Examineur
Pr. Ali YAHYAOUY	Faculté des Sciences Dhar El Mahraz - Fès	Examineur
Pr. Mohammed OURIAGLI	Faculté Polydisciplinaire - Taza	Examineur
Pr. Abderrahim SAAIDI	Faculté Polydisciplinaire - Taza	Co-Directeur de thèse
Pr. Khalid SATORI	Faculté des Sciences Dhar El Mahraz - Fès	Directeur de thèse

*A mes parents,
A mon mari, A ma fille Arij
A mon grand frère Mohammed
A toute ma famille et,
A tous ceux qui me sont chers.*

Khadija

Remerciements

C'est avec un grand plaisir que je réserve cette page, en signe de gratitude et de reconnaissance à tous ceux qui m'ont aidé à la réalisation de ce travail.

Je tiens à exprimer ma reconnaissance au professeur **Khalid SATORI** en tant que directeur de thèse pour m'avoir accueilli au sein de son équipe, pour ses directives scientifiques, pédagogiques et même personnelles pour lesquelles je lui suis hautement redevable. De même, je lui suis extrêmement reconnaissante pour son soutien et son aide précieuse à la préparation et à la rédaction de ce mémoire.

Je tiens tout particulièrement à remercier Monsieur **Abderrahim SAAIDI**, professeur à la Faculté Polydisciplinaire de Taza, Université Sidi Mohammed Ben Abdallah, en tant que co-directeur de thèse, pour son aide, son soutien permanent, sa grande disponibilité, son enthousiasme et sa bienveillance dont il a fait preuve durant toutes ces années. Merci de m'avoir écouté, compris et encouragé tout au long de l'encadrement.

Mes plus sincères remerciements vont à tous les membres du jury d'avoir accepté d'évaluer ce travail, et de prendre part à cette thèse. Tout d'abord, Monsieur **Ismail BOUMHIDI**, professeur à la Faculté des Sciences Dhar El Mahraz de Fès, qui m'a fait l'honneur de présider mon jury de thèse. Je remercie Monsieur **El Hassan SBAI**, professeur à l'École Supérieure de Technologie (EST) de Meknès, Monsieur **Hassan SATORI**, professeur à la Faculté Polydisciplinaire de Nador, Université Mohammed Premier, Monsieur **Hamid TAIRI**, professeur à la Faculté des Sciences Dhar El Mahraz de Fès, d'avoir accepté d'être les rapporteurs de ce manuscrit. Leurs remarques et suggestions lors de la lecture de mon rapport m'ont permis d'apporter des améliorations à la qualité de ce dernier.

J'adresse mes plus vifs remerciements aux membres examinateurs, Monsieur **Farid ABDI**, professeur à la Faculté des Sciences et Technique (FST) de Fès, Monsieur **Ali YAHYAOUY**, professeur à la Faculté des Sciences Dhar El Mahraz de Fès, Monsieur **Mohammed OURIAGLI**, professeur à la Faculté Polydisciplinaire de Taza, Université Sidi Mohammed Ben Abdallah, pour le temps, l'effort et l'intérêt qu'ils ont porté à mon travail.

Je remercie du fond du cœur mes parents pour leur soutien et pour la confiance qu'ils m'ont témoignée pendant toutes ces années d'études. Je ne saurai jamais vous remercier pour

tout ce que vous m'avez apporté. Je n'oublie pas mon mari qui m'a beaucoup poussé à donner mon mieux, et ne cessait jamais à préparer un climat confortable pour travailler.

Je remercie chaleureusement mes collègues de travail au lycée Ibn Zaydoun, et particulièrement à Mme **Khadija EL BAH** pour son aide et son soutien permanent, qui m'a toujours encouragée au cours de la réalisation de ce mémoire.

Je ne saurais oublier dans mes remerciements tout le corps professoral du laboratoire LIIAN, leur persévérance garantit une réelle évolution de la recherche scientifique. Enfin, tous ceux et celles qui ont contribué de près ou de loin à l'accomplissement de ce travail trouvent l'expression de mes remerciements les plus chaleureux.

Table des matières

Remerciements	3
Table des figures.....	9
Liste des tableaux	12
Liste des abréviations.....	13
Résumé	14
Abstract.....	15
Introduction générale.....	16
Contexte de la thèse.....	16
Problématique.....	17
Contributions	19
Organisation du manuscrit.....	20
Chapitre 1 : Modèles et principes de suivi d’objets en vision par ordinateur	21
1.1 Introduction	21
1.2 Principe de suivi d’objets	22
1.3 Représentation des objets	23
1.4 Sélection des caractéristiques.....	27
1.5 Descripteurs statistiques de la couleur	29
1.5.1 Histogramme de couleurs.....	30
1.5.2 Spatiogramme.....	32
1.6 Suivi d’objets.....	33
1.6.1 Suivi par points.....	34
1.6.1.1 Méthodes déterministes	34
1.6.1.2 Méthodes statistiques	36
1.6.1.3 Discussion	36

1.6.2	Suivi par silhouette.....	37
1.6.2.1	Appariement des formes	37
1.6.2.2	Évolution de contour.....	38
1.6.2.3	Discussion	39
1.6.3	Suivi par noyaux.....	39
1.6.3.1	Approches statistiques	40
1.6.3.2	Approches multi-vues	42
1.6.3.3	Discussion	43
1.6.4	Suivi d'objets en estimant l'échelle et l'orientation.....	44
1.6.5	Suivi d'objets par des histogrammes.....	46
1.6.5.1	Représentation par un histogramme mono-caractéristique.....	46
1.6.5.2	Représentation par un histogramme multi-caractéristiques.....	47
1.7	Mesures de similarité	49
1.8	Évaluation des algorithmes de suivi d'objets.....	49
1.9	Conclusion.....	51
Chapitre 2 : Méthode de suivi basée sur les points d'intérêt couleurs		52
2.1	Introduction	52
2.2	Outils de suivi.....	52
2.2.1	Soustraction de l'arrière-plan.....	53
2.2.1.1	Modélisation gaussienne de l'arrière-plan	54
2.2.1.2	Modélisation avec un mélange de gaussiennes.....	54
2.2.2	Détection de la couleur de la peau	57
2.2.2.1	Modélisation de la couleur de la peau.....	58
2.2.2.2	Opérations morphologiques	59
2.2.2.3	Détermination des régions de la couleur de la peau	59
2.2.3	Détection et appariement des points d'intérêt.....	62
2.3	Approche proposée.....	63

2.4	Résultats expérimentaux	66
2.4.1	Données synthétiques	66
2.4.2	Données réelles	67
2.4.3	Discussion	73
2.5	Conclusion.....	74
Chapitre 3 : Nouvelles méthodes de suivi basées sur les histogrammes		75
3.1	Introduction	75
3.2	Méthode de suivi basée sur l’histogramme pondéré et corrigé de l’arrière-plan	76
3.2.1	Estimation de l’échelle et de l’orientation.....	77
3.2.1.1	Estimation de la région cible.....	78
3.2.1.2	Estimation de la largeur, la hauteur et l’orientation de la cible	79
3.2.2	Histogramme pondéré et corrigé de l’arrière-plan	81
3.2.3	Algorithme proposé.....	83
3.3	Méthode de suivi basée sur les caractéristiques couleur-texture.....	84
3.3.1	Description de la texture avec les motifs binaires locaux	85
3.3.2	Histogramme pondéré de l’arrière-plan	88
3.3.3	Algorithme proposé.....	89
3.4	Nouvelle approche de suivi basée sur l’histogramme pondéré du premier-plan	90
3.4.1	Histogramme combiné des caractéristiques couleur-texture.....	91
3.4.2	Histogramme pondéré du premier-plan.....	94
3.4.3	Mise à jour du premier-plan	97
3.4.4	Algorithme proposé.....	98
3.5	Conclusion.....	99
Chapitre 4 : Mise en œuvre pratique et expérimentations		101
4.1	Introduction	101
4.2	Expérimentation 1	101
4.2.1	Méthodes comparées	101

4.2.2 Base de données	102
4.2.3 Résultats expérimentaux	103
4.2.4 Discussion	111
4.3 Expérimentation 2	112
4.3.1 Méthodes comparées	112
4.3.2 Base de données	113
4.3.3 Méthodes d'évaluation.....	114
4.3.4 Résultats expérimentaux	115
4.3.5 Discussion	142
4.4 Conclusion.....	144
Conclusion générale	145
Bibliographie.....	147
Annexe A : Algorithmes de Harris Couleur et de RANSAC et la mesure de corrélation ZNCC.....	166
A.1 Détecteur de Harris Couleur.....	166
A.2 Mesure de corrélation ZNCC	167
A.3 Algorithme de RANSAC	168
A.4 Conclusion.....	169
Annexe B : Algorithme Mean Shift pour le suivi d'objets.....	170
B.1 Introduction	170
B.2 Procédure Mean Shift pour le suivi d'objets.....	170
B.2.1 Représentation de la cible	170
B.2.2 Localisation de la cible.....	172
B.3 Algorithme Mean Shift.....	173
B.4 Conclusion.....	175
Publications.....	176
Articles publiés dans des journaux internationaux.....	176
Communications.....	176

Table des figures

Figure 1.1 : Illustration d'un algorithme de suivi sans modèle à priori de l'objet.....	23
Figure 1.2 : Processus de suivi avec (a) l'étape de la représentation et (b) l'étape de la localisation.	24
Figure 1.3 : Illustration de quelques représentations utilisées pour le suivi d'objets. (a) le centroïde, (b) un ensemble de points caractéristiques, (c) une boîte englobante, (d) ellipse, (e) un contour, (f) une silhouette, (g) une forme articulée et (h) un squelette.....	25
Figure 1.4 : Construction d'un descripteur de couleur basée sur la distribution de l'intensité.	31
Figure 1.5 : Deux objets différents ayant le même histogramme.	31
Figure 1.6 : Histogramme avec des moments d'ordre supérieur : Spatiogramme. (a) Histogramme sans information spatiale et (b) Spatiogramme avec information spatiale.....	33
Figure 1.7 : Taxonomie des méthodes de suivi d'objets.....	34
Figure 1.8 : Évaluation des performances d'un algorithme de suivi d'objets.....	50
Figure 2.1 : Problèmes de détection de la peau. (a) changement d'illumination, (b) richesse ethnique avec des teintes variées et (c) décors complexes.....	57
Figure 2.2 : Exemple de détection de la peau. (a) image originale, (b) image binaire montrant la peau détectée dans l'espace YCbCr en utilisant le modèle gaussien et (c) image binaire montrant la peau détectée dans l'espace YCbCr en utilisant le modèle non-paramétrique.....	58
Figure 2.3 : Exemple de détection de la peau avec la méthode explicite. (a) image originale, (b) pixels de la peau dans l'espace rgb normalisé, (c) pixels de la peau dans l'espace YCbCr, (d) pixels de la peau dans l'espace HSV, (e) pixels de la peau dans l'espace combiné rgb normalisé-HSV-YCbCr.....	61
Figure 2.4 : Étapes de l'approche proposée pour le suivi des personnes.....	64
Figure 2.5 : Influence du bruit et du contraste sur la détection de points d'intérêt couleurs. (a) image originale, (b) résultat de détection des points d'intérêt couleurs sans modification du contraste et du bruit, (c) bruit gaussien avec $\sigma = 0.001$, (d) bruit gaussien avec $\sigma = 0.01$, (e) détection avec faible contraste et (f) détection avec fort contraste.....	67
Figure 2.6 : Résultats du suivi pour la séquence Classe.	69
Figure 2.7 : Résultats du suivi pour la séquence Sport.	71
Figure 3.1 : Processus de calcul d'un motif binaire local (LBP).	86
Figure 3.2 : Quelques exemples des motifs binaires locaux uniformes et invariants à la rotation $\text{LBP}_{8,1}^{\text{riu}2}$	87
Figure 3.3 : Ensemble de toutes les régions utilisées : la région de la cible, la région du premier-plan, la région de recherche et la région candidate.	91

Figure 3.4 : Calcul des histogrammes combinés des caractéristiques couleur-texture q' et f correspondants au modèle de la cible et celui du premier-plan.	92
Figure 3.5 : Étapes de l'approche proposée. Cette figure illustre un exemple pour trouver la nouvelle position de la cible à l'image $t + 1$ en utilisant l'approche proposée CTFWH.....	95
Figure 4.1 : Résultats du suivi pour la séquence Human9 et pour les images de gauche à droite 6, 72, 108, 154 et 287 respectivement.	105
Figure 4.2 : Erreur de la localisation du centre CLE (en pixels) par image pour la séquence Human9.	106
Figure 4.3 : Résultats du suivi pour la séquence Diving et pour les images de gauche à droite 2, 26, 120, 168 et 214 respectivement.....	107
Figure 4.4 : Erreur de la localisation du centre CLE (en pixels) par image pour la séquence Diving.....	108
Figure 4.5 : Résultats du suivi pour la séquence David et pour les images de gauche à droite 2, 9, 296, 439 et 471 respectivement.....	109
Figure 4.6 : Erreur de la localisation du centre CLE (en pixels) par image pour la séquence David.	110
Figure 4.7 : Résultats du suivi pour la séquence Woman et pour les images de gauche à droite 8, 135, 194, 369 et 587 respectivement.....	111
Figure 4.8 : Erreur de la localisation du centre CLE (en pixels) par image pour la séquence Woman.	112
Figure 4.9 : Annotation de la première image des séquences vidéo utilisées pour l'expérimentation. De gauche à droite, de haut en bas: Football1, Human7, Woman, Basketball, Jp1, Jp2, Wbook, Wdesk, David3, Jogging1, Jogging2, Boy, Gym, DragonBaby et Crossing.....	113
Figure 4.10 : Taux de chevauchement (OR) pour les séquences d'images Football1, Human7, Woman, Basketball, Jp1 et Jp2 respectivement.	117
Figure 4.11 : Taux de chevauchement (OR) pour les séquences d'images Wbook, Wdesk, David3, Jogging1, Jogging2 et Boy respectivement.	118
Figure 4.12 : Taux de chevauchement (OR) pour les séquences d'images Gym, DragonBaby et Crossing respectivement.....	119
Figure 4.13 : Erreur de la localisation du centre CLE (en pixels) pour les séquences d'images Football1, Human7, Woman, Basketball, Jp1 et Jp2 respectivement.	120
Figure 4.14 : Erreur de la localisation du centre CLE (en pixels) pour les séquences d'images Wbook, Wdesk, David3, Jogging1, Jogging2 et Boy respectivement.	121
Figure 4.15 : Erreur de la localisation du centre CLE (en pixels) pour les séquences d'images Gym, DragonBaby et Crossing respectivement.	122
Figure 4.16 : Résultats du suivi pour la séquence d'images Woman avec occultation.....	123
Figure 4.17 : Résultats du suivi pour la séquence d'images Wdesk avec occultation.	125
Figure 4.18 : Résultats du suivi pour la séquence d'images Jogging2 avec occultation.....	126

Figure 4.19 : Résultats du suivi de visage dans une scène modérément chargée pour la séquence d'images Jp1 en présence d'éléments distracteurs.	127
Figure 4.20 : Résultats du suivi de visage dans une scène modérément chargée pour la séquence d'images Jp2 en présence d'éléments distracteurs.	129
Figure 4.21 : Résultats du suivi de visage dans une scène modérément chargée pour la séquence d'images Football1 en présence d'éléments distracteurs.	130
Figure 4.22 : Résultats du suivi pour la séquence d'images Basketball avec encombrement de l'arrière-plan.....	132
Figure 4.23 : Résultats du suivi pour la séquence d'images David3 avec encombrement de l'arrière-plan.....	134
Figure 4.24 : Résultats du suivi pour la séquence d'images DragonBaby avec la sortie du champ visuel.....	135
Figure 4.25 : Résultats du suivi pour la séquence d'images Human7 avec le flou du mouvement et le mouvement rapide.	137
Figure 4.26 : Résultats du suivi pour la séquence d'images Boy avec rotation hors et dans le plan de l'image, le flou du mouvement et le mouvement rapide.	138
Figure 4.27 : Résultats du suivi pour la séquence d'images Gym avec rotation hors ou dans le plan de l'image.	139
Figure B.1 : Itérations du suivi par l'algorithme Mean Shift. (B.1a) la position estimée de l'objet à l'instant $t - 1$, (B.1b) l'image à l'instant t avec une estimation initiale de position correspondant à l'estimation précédente, (B.1c), (B.1d), (B.1e) mise à jour de la position estimée avec le calcul itératif du vecteur Mean Shift, (B.1f) estimation finale de la position de l'objet à l'image t . Les images sont prises de la séquence vidéo David [Benchmark 17]....	174

Liste des tableaux

Tableau 2.1 : Règles de segmentation des régions de la peau.	60
Tableau 2.2 : Influence du contraste et du bruit sur la détection de points d'intérêt couleurs.	67
Tableau 2.3 : Précision de suivi (%) et le temps moyen d'exécution (s/image) pour chaque méthode considérée.	73
Tableau 4.1 : Difficultés principales qui caractérisent les séquences d'images utilisées. OCC: Occultation, IV: Variation d'illumination, SV: Variation d'échelle, DEF: Déformation, MB: Flou du mouvement, FM: Mouvement rapide, IPR: Rotation dans le plan de l'image, OPR: Rotation hors le plan de l'image, DS: Élément distracteur, CM: Mouvement de la caméra.	103
Tableau 4.2 : Erreur de la localisation du centre CLE (en pixels) pour notre approche SOBWH et les trois méthodes comparées.	104
Tableau 4.3 : Nombre moyen d'itérations pour notre approche SOBWH et les trois méthodes comparées.	104
Tableau 4.4 : Temps moyen d'exécution (s/image) pour notre approche SOBWH et les trois méthodes comparées.	104
Tableau 4.5 : Difficultés principales qui caractérisent les séquences d'images utilisées. OCC: Occultation, IV: Variation d'illumination, SV: Variation d'échelle, DEF: Déformation, MB: Flou du mouvement, FM: Mouvement rapide, OV: Sortie du champ visuel, IPR: Rotation dans le plan de l'image, OPR: Rotation hors le plan de l'image, BC: Encombrement de l'arrière-plan, DS: Élément distracteur, CM: Mouvement de la caméra.	114
Tableau 4.6 : Pourcentage des images correctement suivies (le taux de la réussite SR) pour nos approches AMT, CTFWH et les huit méthodes comparées.	115
Tableau 4.7 : Erreur de la localisation du centre CLE (en pixels) pour nos approches AMT, CTFWH et les huit méthodes comparées.	116
Tableau 4.8 : Comparaison en termes du temps moyen d'exécution (TPF), la moyenne des images suivies par seconde (FPS) et le nombre moyen d'itérations (NoI).	142

Liste des abréviations

ACI	Analyse en Composantes Indépendantes
ACP	Analyse en Composantes Principales
BWH	Background-Weighted Histogram
CAMSHIFT	Continuously Adaptive Mean Shift
CBWH	Corrected Background-Weighted Histogram
CLE	Center Location Error
CLTP	Completed Local Ternary Pattern
DWT	Discrete Wavelet Transform
EM-Shift	Expectation-Maximisation Shift
FAST	Features from Accelerated Segment Test
FDP	Fonction de Densité de Probabilité
FWH	Foreground-Weighted Histogram
GLCM	Gray-Level Cooccurrence Matrices
GMM	Gaussian Mixture Model
HCI	Human Computer Interaction
HSV	Hue Saturation Value
IPR	In-Plane Rotation
LBP	Local Binary Patterns
LBWH	Likelihood Background Weighted Histogram
OCC	Occultation
OPR	Out-of-Plane Rotation
OR	Overlap Ratio
PTZ	Pan-Tilt-Zoom
RANSAC	RANdom SAmple Consensus
RGB	Red Green Blue
SIFT	Scale-Invariant Feature Transform
SVM	Machines à Vecteurs de Support
SURF	Speeded Up Robust Features
TLD	Tracking Learning Detection
ZNCC	Zero Normalized Cross-Correlation

Résumé

Le suivi des personnes, dans le domaine de la vision par ordinateur, a retenu une attention considérable, cette dernière décennie. En effet, les avancées importantes sont réalisées dans la modélisation du contenu visuel et les gammes d'applications qui utilisent les algorithmes de suivi des personnes particulièrement en : vidéosurveillance, interaction homme-machine, robotique, etc. Bien que, d'autres alternatives sont offertes dans la littérature, le suivi des personnes demeure une difficulté à cause du nombre élevé de facteurs environnementaux.

Les travaux de cette thèse s'inscrivent dans le cadre d'utilisation des représentations par modèle d'apparence pour le suivi des personnes dans une séquence d'images. De nouvelles méthodes pour suivre les personnes dans des scénarios du monde réel sont présentées, avec la résolution des difficultés en relation avec les environnements non contraints où l'unique information disponible sur la cible est sa position sur la première image de la séquence.

Ainsi, les différents problèmes : l'occultation, les changements d'échelle et d'orientation, l'apparition d'éléments distrayeurs, le mouvement de la caméra, la rotation, la similarité d'apparence entre la cible et l'arrière-plan, le changement d'illumination, etc., sont traités.

Les principales contributions proposées dans cette thèse sont contractées sur quatre axes :

- Premièrement, nous avons proposé une méthode qui s'appuie sur les points d'intérêt couleurs et la couleur de la peau. L'utilisation de l'ensemble de ces points d'intérêt détectés sur les régions de la peau, nous permet de suivre ces régions en faisant des correspondances de ces points d'une image à l'autre ;
- Deuxièmement, nous avons proposée une approche basée sur l'algorithme Mean Shift en combinant les descripteurs de moments du poids de l'image et l'histogramme pondéré et corrigé de l'arrière-plan afin d'estimer les changements de la position, de l'échelle et de l'orientation de la cible ;
- Troisièmement, nous avons proposé une approche basée sur la combinaison des caractéristiques couleur-texture et l'information de l'arrière-plan pour représenter la cible ;
- Quatrièmement, nous avons proposée une nouvelle méthode qui vise à renforcer notre précédente contribution en améliorant considérablement la précision et le temps d'exécution. Cette méthode utilise la combinaison des caractéristiques couleur-texture qui est une partie commune avec la précédente et l'histogramme pondéré du premier-plan pour une représentation plus distinctive de la cible.

Nos différentes approches de suivi des personnes sont testées et évaluées par des expérimentations et des comparaisons détaillées dans des scénarios du monde réel. De même, nous avons montré que l'exploitation efficace des caractéristiques locales (points d'intérêt couleurs) et globales (histogrammes) permet de prédire avec précision la position des personnes suivies, malgré les facteurs perturbateurs de la scène. Les avantages de ces approches dans le domaine de la vision par ordinateur sont multiples et bien dégagés. Ainsi, les résultats obtenus sont très satisfaisants.

Mots clés: Suivi des personnes, algorithme Mean Shift, estimation d'échelle et d'orientation, descripteurs de moments, combinaison des caractéristiques couleur-texture, histogramme pondéré de l'arrière-plan, histogramme pondéré du premier-plan.

Abstract

People tracking in computer vision domain have received considerable attention in the last decade. In fact, the significant advances made in visual content modeling and the wide range of applications using people tracking algorithms, especially in: video-surveillance, human-computer interaction, robotic, etc. Although, other alternatives are offered in the literature, tracking people remains an unsolved problem due to the large number of environmental factors.

The work presented in this thesis is part of using representations by appearance model for people tracking in a sequence of images. New methods for people tracking are presented, with resolution of difficulties in relation to unconstrained environments, where the only information available on the target is its position in the first frame of the video sequence.

Thus, the various problems: occlusion, scale and orientation changes, presence of distractors, background clutter, camera motion, illumination variation, rotation, appearance similarity between the target and its background, etc., are processed.

Our main contributions proposed in this thesis are concentrated on four axes:

- Firstly, we have proposed a method based on color interest points and skin color. The use of color interest points set detected on skin regions allows us to track these regions by matching these ones from image to image.
- Secondly, we have proposed an approach based on Mean Shift algorithm which combines moment features of the weight image and the background-weighted histogram in order to estimate the position, scale and orientation changes of the target.
- Third, we have proposed an approach that relies on joint color-texture features and the background-weighted histogram in order to represent the target.
- Finally, we have proposed a new method that aims to enhance our previous contribution by improving greatly the accuracy and the computational cost. This method uses the joint color-texture features which is a common part with the previous contribution and the foreground-weighted histogram for a more effective and distinctive target representation.

Our different approaches of people tracking are tested and evaluated by experiments and detailed comparisons with real world scenarios. Through, the performed experiments, we demonstrate that our efficient exploitation of local features (color interest points) and global features (histograms) ensures a high tracking precision, under several environmental perturbation factors. The advantages of these approaches in the field of computer vision are numerous and clear. Moreover, the results are very satisfactory.

Keywords: People tracking, Mean Shift algorithm, scale and orientation estimation, moment features, joint color-texture features, background-weighted histogram, foreground-weighted histogram.

Introduction générale

Contexte de la thèse

Le travail présenté dans cette thèse s'inscrit dans le cadre de suivi des personnes dans une séquence d'images, et plus particulièrement, sur l'utilisation des représentations par modèle d'apparence pour le suivi. Depuis plusieurs années, la sécurité des personnes est devenue une thématique omniprésente dans l'actualité et porteuse d'enjeux politiques majeurs. Par ailleurs, les progrès réalisés par les fabricants des composants électroniques et informatiques ont fait que la mise en place d'importants réseaux des caméras de vidéosurveillance est devenue abordable pour les compagnies industrielles. Les chercheurs en vision par ordinateur ont pour objectif de permettre aux systèmes informatisés d'extraire automatiquement des informations de haut niveau, d'un point de vue sémantique, à partir du contenu d'une vidéo. Cependant, la compréhension sémantique des événements se produisant dans les vidéos s'appuie sur certaines tâches de bas niveau, telles que la détection des mouvements inhabituels, l'établissement de la trajectoire des personnes visibles, la compréhension des interactions entre les personnes, la reconnaissance et l'interprétation des gestes, etc. Chacune de ces tâches nécessite le suivi de ces personnes au cours du temps. Dans ce travail, nous nous sommes intéressés au problème de suivi des personnes dans une séquence d'images. Ce sujet est aujourd'hui au cœur de nombreuses applications, parmi lesquelles nous pouvons citer :

Systèmes de vidéo-surveillance : Plusieurs applications de vidéo-surveillance liées au suivi des personnes existent [Kettnaker 99], comme le comptage des personnes empruntant un lieu de passage identifié [Hou 11], où le suivi qui aide les fonctionnaires à recueillir des informations concernant le nombre, les comportements, les habitudes et les événements anormaux de ces personnes, par exemple une agression ou un mouvement de foule. Pourvu qu'il existe une description permettant de caractériser la normalité d'une situation.

Systèmes d'interaction homme-machine (HCI) : Dans les systèmes d'interaction homme-machine [Yinghong 11], l'utilisateur est observé par les ordinateurs et interagit avec eux. Donc, il est nécessaire de suivre les comportements des personnes afin de rendre ces systèmes plus intuitifs et intelligents. La plupart des approches existantes sont basées sur des instructions explicites, qui sont communiquées via un périphérique dédié par exemple un joystick pour les jeux vidéo

[Mora 03], ou explicitement par écrit. D'autres approches s'intéressent à la communication homme-machine implicite, soit par la parole ou à travers les gestes [Cadoz 94].

Systèmes de surveillance des conducteurs : Le suivi des conducteurs permet d'améliorer la sécurité de la conduite automobile. Cela est également utile pour avertir le conducteur d'un risque de collision [Ji 04], de surveiller la vigilance des conducteurs pour les empêcher de s'endormir au volant [Bugeau 11], ou pour détecter la présence des piétons se trouvant sur la trajectoire des véhicules [Checchin 08].

Systèmes d'assistance aux personnes âgées : De nombreux travaux s'intéressent à cette problématique où le suivi des personnes peut apporter énormément de solutions, à savoir la détection automatique des chutes, la vérification de la prise des médicaments, ou encore l'étude de certains symptômes comme les troubles du sommeil [Huynh 10]. Ce type d'applications permet de réduire les risques de la santé de ces personnes et rassurer les membres de la famille.

Vision robotique : Plusieurs travaux en vision robotique s'intéressent, en plus des problématiques classiques de reconnaissance d'objets, à des problématiques de suivi des personnes comme la reconnaissance d'actions. Un exemple récent est la plateforme VOIR (Vision and Observation In Robotics) [VOIR 17].

Problématique

Malgré les progrès considérables réalisés dans le suivi des personnes, la conception d'un algorithme de suivi robuste, efficace et sans connaissance à priori du modèle d'apparence de la cible demeure un défi de taille à cause du nombre élevé de facteurs environnementaux. Parmi les principales difficultés rencontrées dans le suivi des personnes, nous pouvons citer:

Occultation partielle ou totale [Bouachir 15, Comaniciu 03c, Wu 15]: Dans le suivi des personnes, la gestion d'occultation partielle ou totale reste encore un problème important causant la contamination du modèle d'apparence de la cible. La difficulté est de continuer à suivre la cible occultée quand certaines de ses parties deviennent invisibles à certains moments de la séquence, soit parce qu'elle est cachée par d'autres objets en mouvement, soit par une structure de l'arrière-plan ou par une autre personne traversant le champ de vision de la caméra. En effet, dans toutes ces situations le modèle d'apparence risque d'évoluer d'une manière imprévue en incluant des caractéristiques de l'objet occultant la cible et l'algorithme de suivi peut dévier pour suivre cet objet ou l'arrière-plan. Ce comportement non souhaitable

est connu sous le nom de dérive de l'algorithme de suivi.

Changement d'illumination [Ross 08]: Les images enregistrées par une caméra dépendent du contenu physique de la scène et des conditions d'éclairage. Considérons l'exemple d'un objet en mouvement dans une pièce éclairée par une lampe à incandescence. Un algorithme de suivi, utilisant des caractéristiques de couleur, risque d'être perturbé par le déplacement de la cible vers une fenêtre l'exposant à la lumière du jour. Le changement d'illumination peut s'effectuer d'une manière abrupte ou graduelle. En effet, ce changement influence considérablement sur l'algorithme de suivi.

Apparition d'éléments distracteurs [Bouachir 15, Wu 15]: Ce problème se présente lorsque la scène surveillée comprend des objets similaires à la cible du point de vue de l'apparence. Ces objets peuvent être confondus avec la cible et risquent de détourner l'algorithme de suivi, d'où leur qualification d'éléments distracteurs. Cependant, la distraction de l'algorithme de suivi peut aussi être causée par un arrière-plan dont la couleur ou la texture est similaire à celle de la cible. Ce scénario représente un défi majeur causant la dérive de l'algorithme de suivi et il diminue la précision de la localisation de l'objet.

Rotation dans et/ou hors le plan de l'image [Bouachir 15]: Dans le suivi sans modèle à priori, une seule vue de la cible est connue par l'algorithme de suivi. Cependant, plusieurs approches dans ce domaine supposent que la pose de la cible ne change pas durant le suivi [Baumberg 94, Wren 97]. Ainsi que, le corps humain est un objet articulé et son apparence est très variable et les différentes parties du corps peuvent être très mobiles et changent rapidement d'orientation. Dans le monde réel, la rotation dans et/ou hors le plan de l'image peut engendrer un changement de pose considérable, ce qui affecte la qualité de suivi. En général, l'efficacité et la flexibilité d'un algorithme de suivi dépendent en partie de sa capacité à tenir compte des changements de rotation.

Encombrement de l'arrière-plan [Bouachir 15, Wu 15]: Une autre difficulté principale causant la dérive de l'algorithme de suivi est l'encombrement de l'arrière-plan. Ce problème se présente lorsque l'arrière-plan de la cible dans la scène contient de nombreux autres objets, il est appelé un environnement encombré pour la cible, dont son apparence peut perturber l'algorithme de suivi.

Mouvement de la caméra [Bousetouane 13, Comaniciu 03c, Wang 16c]: Ce problème est primordial pour la majorité des algorithmes de suivi d'objets. Par ailleurs, plusieurs situations ont été analysées dans la littérature. La caméra peut être fixe et la scène observée en mouvement [Haritaoglu 00a]. C'est le contexte simplifié et le mouvement observé est le fruit

uniquement de la cible que l'on suit, car la caméra n'a aucun mouvement propre, ou un mouvement simple que l'on peut facilement compenser dans le cas des caméras PTZ (Pan Tilt-Zoom) [Murray 94]. D'autre cas, le plus souvent de suivi d'objets sont ceux où la caméra est en mouvement et l'environnement observé est aussi dynamique [Wang 16c].

Changements d'échelle et d'orientation [Bousetouane 13, Ning 12c] : Plusieurs algorithmes de suivi fixent la taille et l'orientation de la fenêtre de suivi, ce qui limite leurs performances lorsque la cible subit un changement d'échelle et/ou d'orientation. En effet, une taille de fenêtre qui fonctionne à une échelle ne convient pas à une autre lorsque la cible se déplace vers ou loin de la caméra. Si la taille de la fenêtre choisie est plus grande, la fenêtre de suivi contient de nombreux pixels de l'arrière-plan avec les pixels de la cible, ce qui mène l'algorithme de suivi à converger vers une zone située entre plusieurs modes, au lieu de converger vers l'un des modes. Si la taille de la fenêtre choisie est très petite, la localisation de la cible est faible, car les pixels de la cible ne sont pas inclus dans la boîte englobante, ce qui dérive l'algorithme de suivi.

Contributions

Pour résoudre les problèmes cités ci-avant, nous avons proposé quatre contributions pour le suivi des personnes dans une séquence d'images et qui peuvent être résumées par :

Système de vidéo-surveillance pour le suivi des personnes en mouvement en utilisant les points d'intérêt couleurs [Laaroussi 14d, Laaroussi 14e]. Ce premier travail présente une approche pour le suivi des personnes basé sur les points d'intérêt couleurs. Pour détecter les régions d'intérêt qui représentent les personnes suivies, nous avons utilisé le modèle de mélange gaussien (GMM) et la couleur de la peau, puis les points d'intérêt couleurs ont été identifiés sur les régions de la peau trouvées. L'utilisation de l'ensemble de ces points d'intérêt nous permet de suivre les régions d'intérêt en faisant des correspondances entre ces points d'une image à l'autre.

Suivi d'une personne basé sur l'histogramme pondéré de l'arrière-plan en estimant l'échelle et l'orientation [Laaroussi 16c]. Cette approche vise à surmonter les principales difficultés de suivi des personnes liées aux changements d'échelle et d'orientation basé sur l'algorithme Mean Shift [Comaniciu 03c]. Nous avons fusionné les descripteurs de moments du poids de l'image [Bradski 98] et l'histogramme pondéré et corrigé de l'arrière-plan afin de concevoir une approche capable de suivre la position, la taille et l'orientation de la cible.

Suivi d'une personne basé sur un modèle d'apparence [Laaroussi 16a]. Cette approche s'appuie sur les caractéristiques de couleur-texture qui sont intégrées dans l'histogramme pondéré de l'arrière-plan afin de représenter la cible suivie.

Suivi d'une personne en utilisant la jointure des caractéristiques couleur-texture et l'histogramme pondéré du premier-plan [Laaroussi 17b]. Dans cette contribution, nous avons présenté une nouvelle approche qui combine les caractéristiques de texture et de couleur pour représenter la cible, puis nous avons proposé un modèle de représentation du premier-plan nommé l'histogramme pondéré du premier-plan afin d'exploiter les caractéristiques principales de celui-ci et de les utiliser pour sélectionner uniquement les caractéristiques importantes de la représentation du modèle de la cible.

Organisation du manuscrit

Ce mémoire est organisé en quatre chapitres comme suit :

Le premier chapitre présente les modèles et les principes de suivi d'objets en vision par ordinateur. Nous définissons le principe de suivi d'objets, puis nous exposons les représentations des objets basées sur leurs formes et leurs apparences. Les différentes méthodes de suivi ont été présentées selon plusieurs catégories. Leurs avantages et défauts sont discutés.

Le deuxième chapitre est consacré à notre première contribution proposée pour le suivi des personnes en utilisant les points d'intérêt couleurs.

Dans le troisième chapitre, nous avons proposé trois approches pour le suivi des personnes dans une séquence d'images basées sur les histogrammes et sous le cadre de l'algorithme Mean Shift.

Le quatrième et le dernier chapitre est focalisé sur la mise en œuvre pratique et les expérimentations pour les trois approches proposées dans le chapitre 3.

La conclusion générale présente un bilan faisant à la fois la synthèse de nos travaux de recherche qui ont porté sur le suivi des personnes en utilisant les modèles d'apparences, ensuite ouvre des perspectives de recherche applicatives sur les méthodes proposées.

Les annexes présentent, d'une part, les algorithmes de Harris couleur et celui de *RANSAC*, la méthode *ZNCC* et d'autre part l'algorithme Mean Shift pour le suivi d'objets.

Chapitre 1

Modèles et principes de suivi d'objets en vision par ordinateur

1.1 Introduction

Deux aspects indissociables et duaux coexistent dans le problème de suivi d'objets, à savoir la représentation de l'apparence de l'objet connaissant sa localisation et la localisation de l'objet connaissant son apparence. Le premier cas concerne la représentation informatique des objets, tandis que le second cas concerne l'utilisation de cette représentation dans un objectif de localisation. Les approches de suivi fondées sur l'apparence des objets diffèrent principalement entre elles par la manière d'aborder les questions suivantes: Quelle représentation de l'objet convient pour le suivi ? Quelles caractéristiques doivent être utilisées ? Comment modéliser le mouvement, l'apparence et la forme de l'objet ? Quelles mesures de similarité doivent être utilisées pour reconnaître l'objet au cours de la séquence ? Les réponses possibles à ces questions dépendent du contexte dans lequel le suivi est réalisé.

Le choix d'un modèle d'apparence est un des problèmes centraux pour le suivi d'objets. Il consiste à choisir une représentation pertinente de l'objet, donnant naissance à des caractéristiques significatives et fiables qui décrivent le contenu visuel de cet objet tout en permettant de le reconnaître au cours du temps, malgré les changements des conditions environnementales et des variations de l'apparence le long de la séquence d'images. L'un des principaux facteurs qui limite les performances des algorithmes de suivi est la capacité du modèle d'apparence à rester valide lorsque l'apparence de l'objet change rapidement. L'absence de cette adaptation peut en effet entraîner une localisation erronée de l'objet.

Ce chapitre est organisé comme suit. La section 1.2 définit le principe de suivi d'objets. Les représentations des objets basées sur leurs formes et leurs apparences ont été présentées dans la section 1.3. La section 1.4 présente les caractéristiques les plus utilisées pour le suivi d'objets. La section 1.5 expose les descripteurs statistiques de la couleur. Les méthodes existantes pour le suivi d'objets ont été présentées dans la section 1.6. Les mesures de similarité ont été abordées dans la section 1.7. La Section 1.8 présente quelques techniques pour l'évaluation des algorithmes de suivi. Finalement, une conclusion est présentée dans la

section 1.9.

1.2 Principe de suivi d'objets

Le suivi d'objets dans une séquence d'images est depuis ces dernières décennies, un thème de recherche très actif en vision par ordinateur. Généralement, le suivi consiste à localiser l'objet d'intérêt sur chaque image de la séquence. Le procédé de localisation se fonde sur la reconnaissance de l'objet à partir d'un ensemble de caractéristiques visuelles comme la couleur, la forme, la vitesse, la texture ou une combinaison de ces caractéristiques. À l'initialisation, l'algorithme de suivi construit un modèle de référence de l'objet suivi. Le modèle d'apparence peut être basé sur des connaissances à priori comme des images prises hors ligne de différents points de vue, ou seulement sur l'annotation de la région de l'objet sur la première image de la séquence, par exemple une boîte englobante. Plus formellement, l'algorithme de suivi estime dans chaque image de la séquence, les paramètres des objets présents dans le champ de vision de la caméra. Les paramètres à estimer peuvent être divers, mais comprennent principalement une composante géométrique, indiquant la position du centre de l'objet suivi dans l'image [Comaniciu 03c, Lucas 81] à laquelle peuvent s'ajouter d'autres informations selon le domaine d'application. L'ensemble de ces caractéristiques constitue l'état de l'objet et il est défini pour chaque image de la séquence. La figure 1.1 ci-après illustre le fonctionnement général d'un algorithme de suivi sans modèle à priori de l'objet.

Dans la littérature de la vision par ordinateur, il existe une multitude des méthodes de suivi d'objets. Les méthodes proposées se distinguent principalement par le modèle d'apparence décrivant la cible et la stratégie de recherche utilisée pour en prédire l'état. Ces deux composantes représentent les deux aspects majeurs d'un algorithme de suivi et doivent être conçues en tenant compte de l'environnement et de la finalité du suivi. Dans ce travail, nous faisons la distinction entre les notions de suivi par mise en correspondance et suivi par mise à jour [Mikram 08]. Le suivi par mise en correspondance est classiquement utilisé pour la construction des trajectoires des caractéristiques de l'objet telles que les points d'intérêt [Chetverikov 00, Serby 04] détectés préalablement. Cette approche s'attache à détecter un objet dans chaque image, puis à agglomérer temporellement les objets détectés de façon à obtenir des trajectoires cohérentes au cours du temps. Le suivi par mise à jour détecte ou localise l'objet au sein d'une image en fonction de l'état de l'objet à l'instant précédent. Dans ce dernier cas, le vecteur d'état est mis à jour sur chaque nouvelle image, à partir d'une

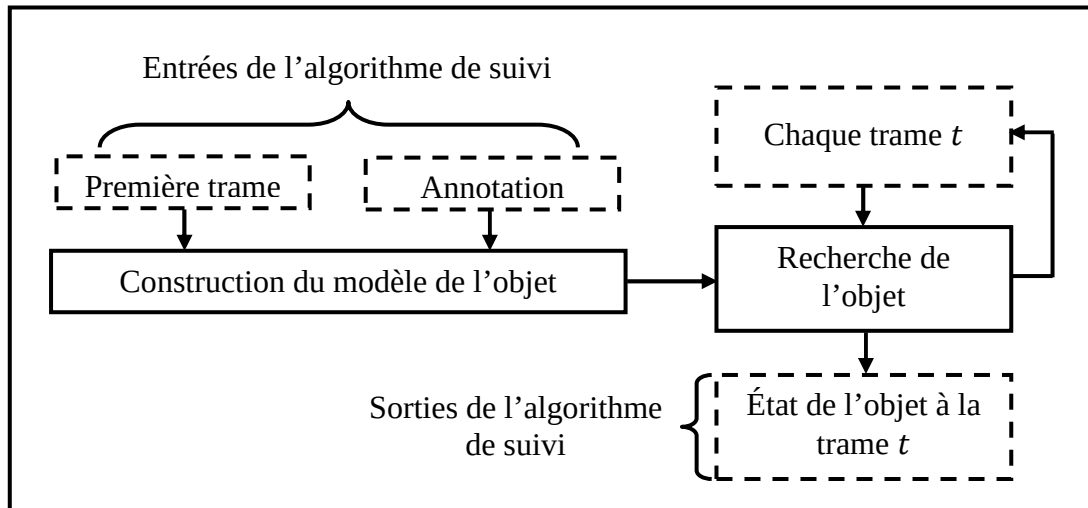


Figure 1.1 : Illustration d'un algorithme de suivi sans modèle à priori de l'objet.

initialisation définie manuellement ou automatiquement sur l'une des images de la séquence. D'un point de vue fonctionnel, le suivi par mise à jour se décompose en deux parties principales : la représentation et la localisation de l'objet.

- **Étape de représentation** : Elle consiste à associer à l'objet suivi des caractéristiques de forme et/ou d'apparence permettant de le reconnaître dans des images successives. Les caractéristiques sélectionnées sont destinées à former le modèle de référence, il s'agit d'une description numérique ou symbolique qui permet de caractériser l'apparence visuelle de l'objet dans l'image ou la séquence. Ceci peut être accompli de façon classique par l'initialisation du modèle sur une image de référence (Figure 1.2 (a)) bien que certaines méthodes prennent en compte plusieurs images de référence afin de transcrire une plus grande variabilité dans l'apparence de l'objet [Black 98c].
- **Étape de localisation** : C'est le procédé numérique réalisant le suivi qui a pour but de fournir la position de l'objet dans l'image courante de la séquence, sur la base de son modèle de représentation et de son état dans l'image précédente (Figure 1.2 (b)).

1.3 Représentation des objets

Dans un scénario de suivi, un objet peut être défini comme tout ce qui est d'intérêt pour une analyse ultérieure. Par exemple, les bateaux sur la mer, les poissons dans un aquarium, les véhicules sur une route, les avions dans l'air, les personnes marchant sur une route, ou les bulles dans l'eau sont tous un ensemble d'objets qui peuvent être suivis dans un domaine spécifique. Les objets peuvent être représentés par leurs formes et leurs apparences. Dans cette section, nous allons décrire les représentations de la forme des objets généralement

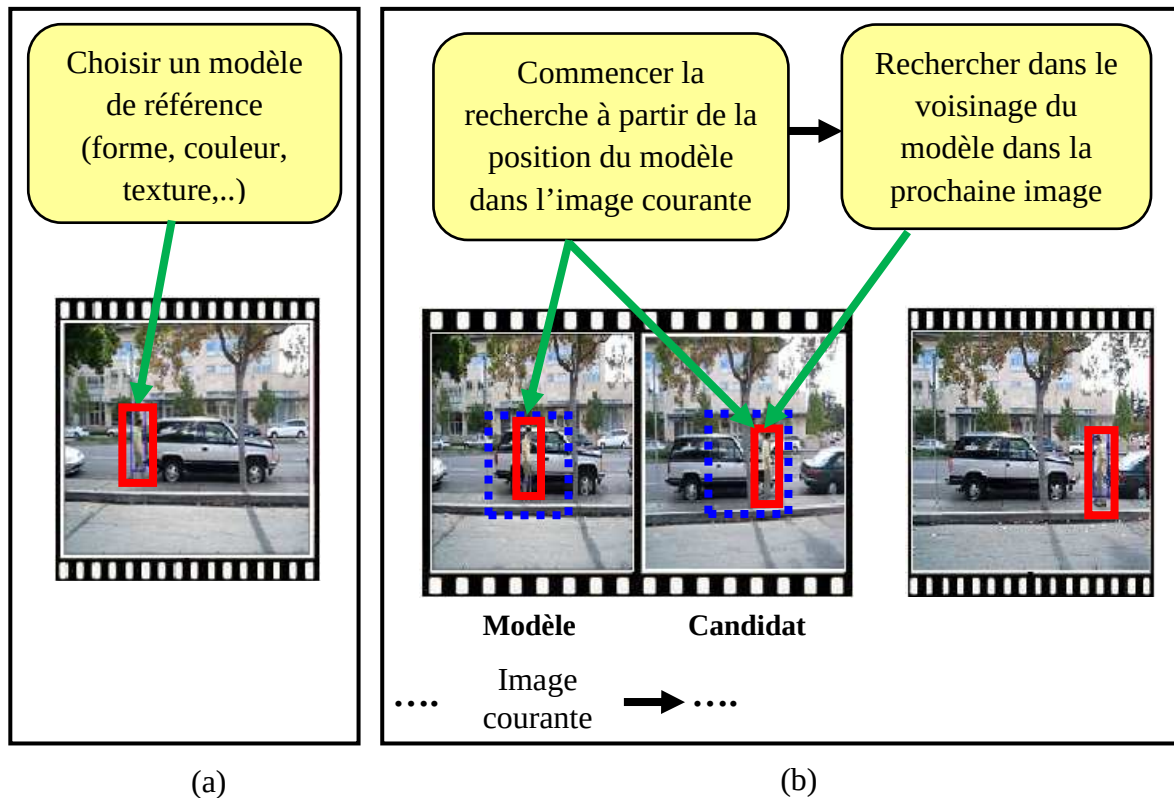


Figure 1.2 : Processus de suivi avec (a) l'étape de la représentation et (b) l'étape de la localisation.

utilisées pour le suivi et ensuite aborder les représentations combinées de la forme et de l'apparence. Selon une étude faite par Yilmaz et al. [Yilmaz 06a] les représentations de la forme sont classées comme suit :

- **Points** : Un objet est représenté par un point qui est son centroïde (Figure 1.3 (a)) [Veenman 01] ou par un ensemble de points (Figure 1.3 (b)) [Serby 04]. La représentation par points est appropriée pour le suivi d'objets qui occupent de petites régions dans une image ou pour suivre les objets non rigides.
- **Formes géométriques primitives** : La forme de l'objet est représentée par un rectangle (Figure 1.3 (c)), une ellipse (Figure 1.3 (d)) ou un cercle, afin de délimiter la région d'intérêt [Bernardin 07, Comaniciu 03c, Cucchiara 03]. Le mouvement de l'objet pour une telle représentation est généralement modélisé par une transformation de translation, affine ou projective (homographie). Cependant, les formes géométriques primitives sont plus appropriées pour représenter des objets rigides simples et elles sont également utilisées pour le suivi d'objets non rigides comme les personnes.
- **Silhouette ou contour de l'objet** : Une représentation par contour définit les frontières de l'objet (Figure 1.3 (e)). La région interne du contour est appelée la silhouette de l'objet (Figure 1.3 (f)) est peut être utilisée conjointement avec l'information du contour pour le

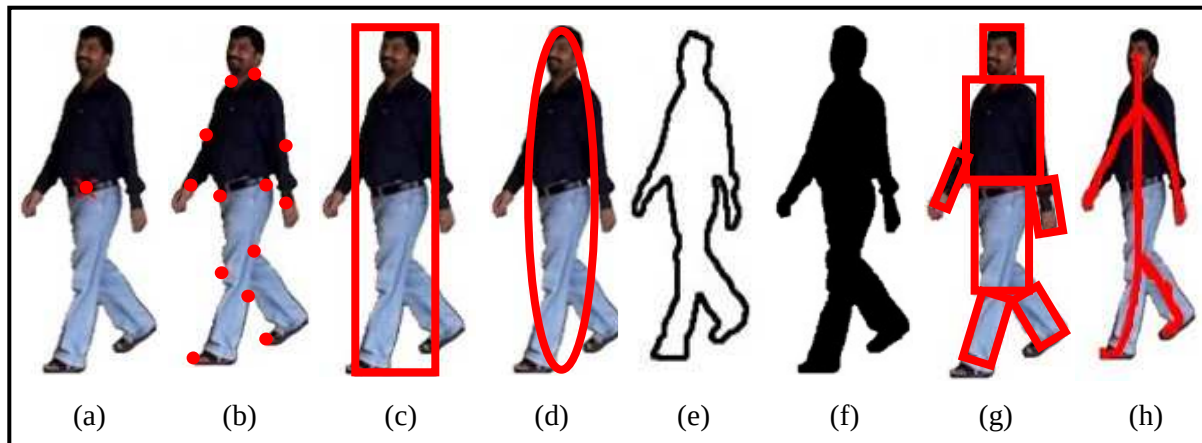


Figure 1.3 : Illustration de quelques représentations utilisées pour le suivi d'objets. (a) le centroïde, (b) un ensemble de points caractéristiques, (c) une boîte englobante, (d) ellipse, (e) un contour, (f) une silhouette, (g) une forme articulée et (h) un squelette.

suivi d'objets. Les représentations par silhouette ou contour sont appropriées pour le suivi des formes non rigides complexes [Yilmaz 04b].

- **Modèles de forme articulée :** Les objets articulés sont composés des parties du corps qui sont maintenues ensemble avec des articulations [Zhao 06a]. Par exemple, le corps humain est un objet articulé avec le torse, les jambes, les mains, la tête et les pieds reliés par des articulations. La relation entre les différentes parties du corps peut être régie par des modèles de mouvement cinématique comme l'angle d'articulation. Pour représenter un objet articulé, on peut modéliser les parties constitutives en utilisant des cylindres ou des ellipses (Figure 1.3 (g)).
- **Modèles de squelette :** Le squelette d'un objet peut être extrait en appliquant la transformation de l'axe médian à la silhouette de l'objet [Ballard 82]. Ce modèle est couramment utilisé comme une représentation de la forme pour la reconnaissance des objets [Ali 01]. La représentation par squelette peut être utilisée pour modéliser les objets articulés et rigides (Figure 1.3 (h)).

Il existe plusieurs façons pour représenter les caractéristiques d'apparence des objets. Notez que, les représentations de la forme peuvent être également combinées avec les représentations de l'apparence pour le suivi afin de compléter la représentation de l'objet [Cootes 01, Cutler 00]. Certaines représentations de l'apparence communes dans le cadre de suivi d'objets sont :

- **Densités de probabilité de l'apparence :** Les estimations de la densité de probabilité de l'apparence peuvent être paramétriques, comme une gaussienne [Zhu 96] ou un mélange de gaussiennes [Paragios 02c], ou non paramétriques, comme les fenêtres de Parzen

[Elgammal 02c] et les histogrammes [Comaniciu 03c]. Les densités de probabilité des caractéristiques de l'apparence d'un objet peuvent être calculées à partir des régions d'une image spécifiées par les modèles de la forme (région intérieure d'une ellipse ou d'un contour).

- **Modèles** : Les modèles ou les templates sont formés en utilisant des formes géométriques simples ou des silhouettes [Fieguth 97]. Un avantage de l'utilisation d'un modèle est qu'il porte à la fois l'information spatiale et celle de l'apparence. Cependant, les modèles ne codent que l'apparence de l'objet générée à partir d'une seule vue, alors ils ne conviennent que pour le suivi d'objets dont leurs poses ne varient pas considérablement au cours de suivi.
- **Modèles actifs de l'apparence** : Les modèles actifs de l'apparence sont générés en modélisant simultanément la forme et l'apparence de l'objet [Edwards 98]. En général, la forme de l'objet est définie par un ensemble de repères. De façon similaire à la représentation basée sur le contour, les repères peuvent résider sur la frontière de l'objet ou, alternativement, ils peuvent résider à l'intérieur de la région de l'objet. Pour chaque repère, un vecteur d'apparence est enregistré et qui prend la forme de la couleur, la texture ou l'amplitude du gradient. Les modèles actifs de l'apparence nécessitent une phase d'apprentissage permettant d'apprendre à partir d'un ensemble d'échantillons, la forme et son apparence associée à l'objet en utilisant, par exemple l'analyse en composantes principales (ACP).
- **Modèles d'apparence multi-vues** : Ces méthodes encodent les différentes vues d'un objet. Une approche pour représenter ces différentes vues est de générer un sous-espace à partir de vues données. Les approches sous-espaces, par exemple l'analyse en composantes principales (ACP) et l'analyse en composantes indépendantes (ACI), ont été utilisées pour la représentation de la forme et de l'apparence [Mughadam 97, Black 98c]. Une autre approche pour apprendre les différentes vues d'un objet est de former un ensemble de classifieurs comme les machines à vecteurs de support (SVM) [Avidan 04] ou les réseaux Bayésiens [Park 04]. Une des limites des modèles d'apparence multi-vues est que les apparences dans toutes les vues sont nécessaires à l'avance.

Discussion : Généralement, il y a une forte relation entre la représentation de l'objet et l'algorithme de suivi. Cette représentation est choisie en fonction du domaine d'application. En effet, pour le suivi d'objets qui apparaissent très petits dans une image, la représentation par point est généralement appropriée. Par exemple, les auteurs Veenman et al. [Veenman 01]

utilisent la représentation par point pour suivre les graines dans une séquence de plat mobile. De même, les auteurs Shafique et al. [Shafique 05] utilisent une représentation par point pour suivre les oiseaux éloignés. Pour les objets dont les formes peuvent être approximées par des rectangles ou des ellipses, les représentations par des formes géométriques primitives sont plus appropriées. Les auteurs Comaniciu et al. [Comaniciu 03c] utilisent une représentation par forme elliptique et un histogramme de couleurs calculé à partir de la région elliptique afin de modéliser l'apparence. Black and Jepson [Black 98c] utilisent les vecteurs propres pour représenter l'apparence de l'objet. Les vecteurs propres ont été générés à partir des modèles de l'objet rectangulaire. Pour le suivi d'objets non rigides ou de formes complexes comme les personnes, les représentations par contour ou silhouette [Haritaoglu 00a] ou par des formes géométriques primitives [Bousetouane 13] sont les plus appropriées.

1.4 Sélection des caractéristiques

La sélection des caractéristiques est primordiale pour assurer un bon suivi, car cette étape est fortement liée à la représentation de l'objet. En effet, la propriété la plus souhaitable d'une caractéristique visuelle est son unicité de sorte que les objets peuvent être facilement distingués dans l'espace des caractéristiques. Cependant, certains algorithmes de suivi utilisent une seule caractéristique pour la représentation de la cible, néanmoins d'autres algorithmes utilisent une combinaison de plusieurs caractéristiques afin d'améliorer les performances des algorithmes de suivi. Il existe une littérature abondante sur les caractéristiques destinées à la représentation des objets [Yang 11e]. Nous présentons ci-après les caractéristiques visuelles les plus utilisées pour le suivi d'objets.

- **Couleur** : La couleur, de par son pouvoir discriminant, est l'une des caractéristiques les plus utilisées pour la représentation des objets. Cet intérêt provient de son invariance par rapport à l'échelle, l'orientation et la translation. La représentation des objets basée sur la couleur repose sur plusieurs critères : le choix de l'espace colorimétrique [Allen 04], le mode de représentation de la couleur dans cet espace [Stricker 95], l'étude des problèmes d'invariance aux conditions d'illumination [Lee 01b] et la combinaison avec des attributs complémentaires comme la texture [Takala 07, Bousetouane 13], la forme [Wang 16c] ou l'information spatiale [Zhao 05c]. Plusieurs travaux ont été réalisés dont l'objectif est d'identifier les espaces colorimétriques les plus discriminants [Stricker 95]. L'abondance de la littérature sur cette thématique a montré qu'il n'existe pas d'espace couleur idéal. L'espace RGB (Rouge, Vert, Bleu) est l'espace standard le plus utilisé [Collins 05c,

Comaniciu 03c, Georgescu 04]. Néanmoins, cet espace est sensible aux variations d'illumination [Gros 97]. L'espace HSV par sa composante teinte permet de résister aux conditions d'éclairage et de point de vue [Allen 04, Geusebroek 01]. D'autres espaces définissent comme perceptuellement uniformes, tels que CIELab et CIELuv servent également à la caractérisation de la couleur [Liu 06]. Certains travaux se singularisent par l'utilisation des composantes isolées des espaces précédents et par l'utilisation combinée de celles-ci [Collins 05c].

- **Forme** : Les caractéristiques de la forme sont extraites à partir des régions qui ont été segmentées au préalable. Une fois un masque de segmentation est obtenu sur la région d'intérêt, la forme de ce masque peut être utilisée afin d'associer un descripteur à la région qui a donné lieu à ce masque. Les descripteurs fondés sur l'intérieur du masque de segmentation caractérisent l'intégralité de la forme d'une région, par exemple à travers les moments invariants [Hu 62a]. Les descripteurs fondés sur le contour du masque font référence aux descripteurs de Fourier [Persoon 77]. Une propriété importante des caractéristiques de la forme est qu'elles sont moins sensibles aux changements d'illumination par rapport aux caractéristiques de la couleur, ainsi qu'elles sont robustes aux transformations géométriques comme la translation, la rotation et les changements d'échelle. En raison de sa simplicité et de sa précision, l'approche la plus populaire pour la détection de la forme est le détecteur de Canny Edge [Canny 86]. Une évaluation des algorithmes de détection des formes est fournie par Bowyer et al. [Bowyer 01].
- **Texture** : La texture est une mesure pour la variation de l'intensité d'une surface qui quantifie certaines propriétés comme la douceur et la régularité. L'information véhiculée par les caractéristiques de texture concerne la manière dont les couleurs sont organisées sur la surface de l'objet en répétition d'éléments [Haralick 73]. Généralement, une texture peut avoir un aspect périodique ou aléatoire. Selon Gagalowicz et al. [Gagalowicz 83] une texture est modélisée comme une structure spatiale constituée de l'organisation des primitives ayant chacune un aspect aléatoire. Plusieurs travaux sont penchés pour étudier et analyser les caractéristiques de texture. Les auteurs Haralick et al. [Haralick 73] proposent l'utilisation des matrices de co-occurrences pour la représentation de la texture (Gray-Level Cooccurrence Matrice GLCM). Gotlieb et Kreyszig [Gotlieb 90] se sont succédés afin d'améliorer cette approche. Ils ont montré expérimentalement que les paramètres du contraste, d'homogénéité et d'entropie sont fortement discriminatifs. D'autres travaux utilisent des études psycho-visuelles [Tamura 78] ou mathématiques, telles que les

ondelettes ou les filtres de Gabor pour décrire la texture. Dans de nombreuses applications de suivi d'objets [Takala 07, Bousetouane 13], l'utilisation de la texture est fortement choisie pour la représentation de la cible. Les caractéristiques de texture sont moins sensibles aux changements d'illumination par rapport à la couleur.

- **Flot optique** : Le flot optique est un champ vectoriel dense de déplacement visuel où à chaque pixel de l'image est associé un vecteur de déplacement qui permet d'exprimer la vitesse de déplacement d'un pixel dans le repère de l'image et la direction vers laquelle le pixel se déplace en faisant l'hypothèse que l'intensité (ou la couleur) d'un pixel est conservée au cours du déplacement [Horn 81b]. Le calcul du flot optique sert à suivre les objets qui se déplacent dans une scène filmée avec une caméra fixe ou mobile. Le flot optique est connu pour sa robustesse vis à vis des variations de luminosité et des déformations de l'objet suivi. Cependant, le flot optique est couramment utilisé comme caractéristique dans les applications de segmentation de mouvement et le suivi d'objets. Les techniques populaires pour calculer le flux optique dense incluent les méthodes de Horn et Schunck [Horn 81b], Lucas et Kanade [Lucas 81], Black et Anandan [Black 96a] et Szeliski et Coughlan [Szeliski 97].
- **Caractéristiques hybrides** : Pour améliorer les performances des algorithmes de suivi, certains auteurs combinent plusieurs caractéristiques. Par exemple, les auteurs Takala et al. [Takala 07] utilisent un modèle hybride pour réaliser le suivi en temps réel. Ces auteurs utilisent le mouvement et la texture afin d'améliorer la représentation des objets et assurer un bon suivi. Les auteurs Wang et al., Yilmaz et al. [Wang 08d, Yilmaz 04b] ajoutent la forme de l'objet aux caractéristiques couleur-texture pour récupérer les parties occultées de la cible lors de l'occultation. Ces auteurs ont été démontrés que la représentation multi-caractéristiques des objets peut se traduire par une meilleure performance de suivi contre les distractions. Toutefois, lorsqu'il n'y a pas d'occultation, la forme de l'objet peut avoir des inconvénients, tels que des objets similaires avec différentes formes ou un changement brusque de la taille de l'objet.

1.5 Descripteurs statistiques de la couleur

Nous nous focaliserons dans cette section sur les modèles d'apparences basés sur la représentation statistique de la couleur, pour lesquels nous proposerons des améliorations, dans la suite du mémoire, en visant à améliorer la qualité de suivi des personnes et le temps d'exécution en intégrant d'autres caractéristiques dans la représentation de la cible.

1.5.1 Histogramme de couleurs

L'histogramme de couleurs d'un objet est l'une des représentations décrivant son apparence dans un processus de suivi. Il permet de représenter statistiquement la distribution des couleurs des pixels, c'est-à-dire la proportion des pixels répartis sur un ensemble de classes de couleurs, comme illustré sur la figure 1.4 ci-après. Pour un histogramme normalisé, la valeur de chaque classe s'interprète comme la probabilité d'occurrence de la classe dans l'image. Cependant, l'histogramme de couleurs est un outil fondamental du traitement d'images, avec de très nombreuses applications à savoir la reconnaissance d'objets [Swain 91], la recherche d'images et des séquences [Bach 96] et le suivi d'objets [Bradski 98, Comaniciu 03c]. Ces approches utilisent l'information de la couleur afin de décrire l'objet d'intérêt. Le succès des approches par histogramme de couleurs provient de leur faible complexité calculatoire associée à une bonne robustesse vis-à-vis des bruits et de leur invariance aux rotations, aux translations et aux changements d'échelle.

Dans de nombreuses applications, l'histogramme est calculé sur une région d'intérêt définie par une forme géométrique simple englobant cet objet. Soit $R = (x_1, x_2, \dots, x_n)^T$ cette région, $\{x_i^*\}_{i=1 \dots n}$ l'ensemble des coordonnées normalisées des n pixels dans la région R centré à 0, $I(x_i)$ la couleur du pixel i et U un ensemble de m classes de l'espace de couleur utilisé (RGB, HSV, LUV, etc.), $\text{card}(U) = m$, la $u^{\text{ème}}$ classe est caractérisée par sa fonction indicatrice $\varphi_u(c)$ qui vaut 1 pour une couleur c associée au bin u et 0 sinon [Comaniciu 03c]. Avec toutes ces définitions, l'histogramme pour un bin u est défini par :

$$q_u = C_1 \sum_{i=1}^n \varphi_u(I(x_i^*)) \quad (1.1)$$

avec C_1 est un facteur de normalisation où $\sum_{u=1}^m q_u = 1$ [Comaniciu 03c]. Ainsi, nous pouvons représenter un m-histogramme de couleurs d'un objet par la concaténation des valeurs des bins comme suit $q \equiv \langle q_u | u = 1, \dots, m \rangle$. Dans certains cas la région d'intérêt peut englober l'objet suivi et des éléments de l'arrière-plan. Pour limiter l'influence de l'arrière-plan sur la localisation de la cible et privilégier l'information pertinente, il est possible de calculer un histogramme pondéré [Comanciu 03c]. Plus les pixels sont loin du centre de l'objet, plus le poids pris en compte dans l'histogramme final est faible. D'où, l'histogramme pondéré pour un bin u est défini par :

$$q_u = C_2 \sum_{i=1}^n K(x_i^*) \varphi_u(I(x_i^*)) \quad (1.2)$$

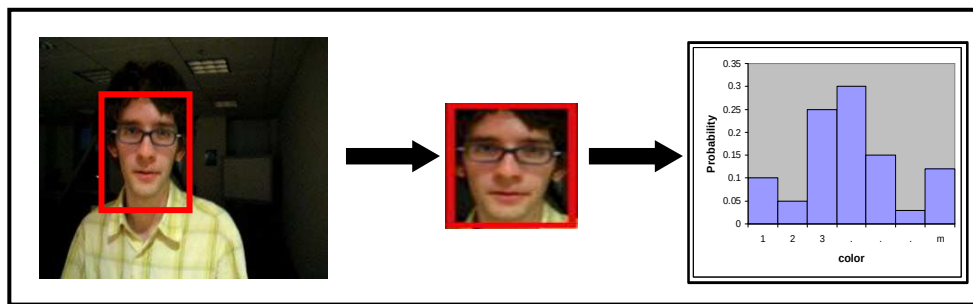


Figure 1.4 : Construction d'un descripteur de couleur basée sur la distribution de l'intensité.

avec K est un noyau spatial associant un poids à chaque position spatiale x_i et C_2 est un facteur de normalisation. Notez que, plusieurs types de noyaux sont couramment utilisés : uniforme, triangle, Epanechnikov, quadratique, cubique, gaussien, et circulaire.

Les histogrammes de couleurs sont principalement utilisés pour suivre la position et parfois les changements d'échelle. Cependant, ils ne codent aucune information spatiale sur la région d'intérêt de l'objet. Ceci présente l'avantage de donner une représentation invariante aux transformations spatiales affines et relativement peu sensible aux transformations non rigides qui permet de modifier les surfaces apparentes relatives aux différentes parties d'un objet.

L'histogramme de couleurs tend alors à donner des résultats insuffisants pour le processus du suivi lorsque les couleurs de l'objet ne sont pas suffisamment contrastés par rapport à l'arrière-plan. Deux objets visuellement très différents peuvent avoir la même distribution c'est-à-dire ayant des histogrammes semblables. En effet, ils peuvent n'être différenciés que par les arrangements différents des pixels dans les fenêtres (Figure 1.5).

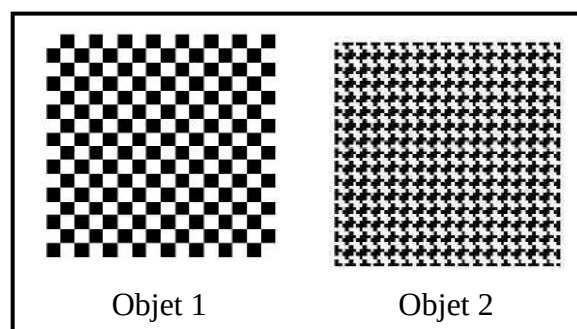


Figure 1.5 : Deux objets différents ayant le même histogramme.

Un autre inconvénient majeur de l'utilisation des histogrammes de couleurs est que le choix du nombre de classes (bins) d'un histogramme est un paramètre important. La discrétisation en un nombre de classes très important ne permet pas d'être robuste aux changements d'illumination de la scène ou aux changements d'apparence de l'objet, et la plupart du temps l'algorithme de suivi dérive au fur et à mesure de la séquence. À l'inverse,

un nombre de classes très faible ne permet pas de discriminer efficacement l'objet du reste de la scène, et le suivi échoue. Dans la plupart des approches existantes, le nombre de classes peut être choisi arbitrairement au début de la séquence et reste fixé durant le processus de suivi. Rien n'indique qu'une telle partition est optimale étant donné la densité que l'on veut estimer. S'il était possible de trouver la partition optimale, l'algorithme serait plus robuste. Cette problématique de la détermination d'un nombre optimal de classes pour un histogramme a été étudiée dans les travaux de [Birgé 02].

1.5.2 Spatiogramme

Même si les histogrammes de couleurs ont montré leur intérêt pour représenter les couleurs de l'objet, ils ne prennent pas en compte les informations spatiales qui peuvent fortement influencer sur la discrimination entre l'objet à suivre et le reste de la scène. Pour contourner cette limitation, Birchfield et al. [Birchfield 05b] ont introduit le concept de spatiogramme, qui ajoute des informations spatiales, à savoir la moyenne spatiale et la matrice de covariance des pixels aux histogrammes comme montré sur la figure 1.6 ci-après. Cette information spatiale permet encore des transformations assez générales, comme dans un histogramme, mais capture une description plus riche de la cible pour augmenter la robustesse du suivi. Le spatiogramme h d'une image ou d'une région de l'image est représentée par :

$$h^{(2)} = \langle p_u, \mu_u, \Sigma_u \rangle_{u \in U} \quad (1.3)$$

avec p_u est proportionnel au nombre de pixels qui appartiennent au $u^{\text{ème}}$ bin, μ_u et Σ_u sont respectivement les vecteurs "moyennes des coordonnées" des pixels participant à ce bin et leur matrice de covariance. Ces paramètres sont calculés par :

$$\mu_u = \frac{1}{\sum_{j=1}^n \varphi_u(I(x_j))} \sum_{i=1}^n x_i \varphi_u(I(x_i)) \quad (1.4)$$

$$p_u = C_3 \sum_{i=1}^n \varphi_u(I(x_i)) \quad (1.5)$$

$$\Sigma_u = \frac{1}{\sum_{j=1}^n \varphi_u(I(x_j))} \sum_{i=1}^n (x_i - \mu_u)(x_i - \mu_u)^T \varphi_u(I(x_i)) \quad (1.6)$$

où C_3 est un facteur de normalisation tel que $\sum_{u=1}^m p_u = 1$ [Comaniciu 03c], $\{x_i\}_{i=1 \dots n}$ l'ensemble des coordonnées normalisées des n pixels, $I(x_i)$ la couleur du pixel i et U un ensemble de m classes de l'espace de couleur utilisé, la $u^{\text{ème}}$ classe est caractérisée par sa

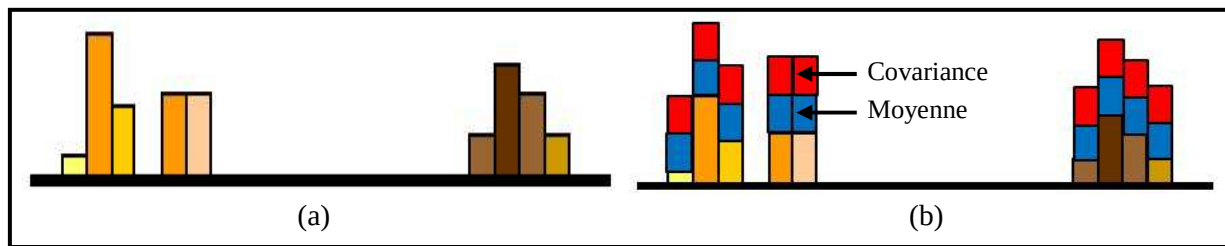


Figure 1.6 : Histogramme avec des moments d'ordre supérieur : Spatiogramme. (a) Histogramme sans information spatiale et (b) Spatiogramme avec information spatiale.

fonction indicatrice $\varphi_u(c)$ qui vaut 1 pour une couleur c associée au bin u et 0 sinon [Comaniciu 03c]. En effet, les méthodes de suivi basées sur les spatiogrammes sont très discriminantes et parfois rendent le suivi très sensible à l'orientation et à la déformation. De ce fait, ces méthodes sont moins adaptées pour le suivi des personnes qui présentent de fortes déformations. Cependant, le spatiogramme est plus vulnérable à la grande variabilité spatiale de la distribution des couleurs. Il est aussi très sensible aux changements d'illumination, surtout avec l'utilisation d'une mesure de similarité non adéquate [Conaire 07].

1.6 Suivi d'objets

Le suivi d'objets, dans le domaine de la vision par ordinateur, a retenu une attention particulière cette dernière décennie. Une multitude d'états de l'art ont été proposés dans la littérature [Hu 04b, Kim 10, Li 13c, Pan 16, Yilmaz 06a]. Chacun de ces travaux est focalisé sur une ou plusieurs classes de problèmes rencontrés dans la conception d'un algorithme de suivi. L'objectif de l'ensemble de ces travaux est de classer et d'évaluer les méthodes utilisées pour le suivi d'objets. Les méthodes existantes se distinguent par la représentation de la forme et de l'apparence de l'objet suivi.

Il existe plus d'une catégorisation possible des algorithmes de suivi dans la littérature. Les auteurs Yilmaz et al. [Yilmaz 06a] ont publié une revue de la littérature couvrant les travaux majeurs de suivi d'objets, en catégorisant les algorithmes de suivi selon la représentation de l'objet. Leur classification comprend trois catégories de méthodes : le suivi par points [Veenman 01, Grabner 07a], le suivi par silhouette ou contour [Sato 04, Xu 02] et le suivi par noyau [Tao 02, Leichter 10]. Dans cette section, nous présentons les principaux travaux inspirants de cette recherche en adoptant une catégorisation des méthodes basées sur les modèles d'apparences. La catégorisation proposée tient compte à la fois des classifications antérieures et des derniers travaux sur la modélisation visuelle des objets suivis. La figure 1.7 ci-après présente la taxinomie proposée en illustrant les sous-catégories de chaque catégorie.

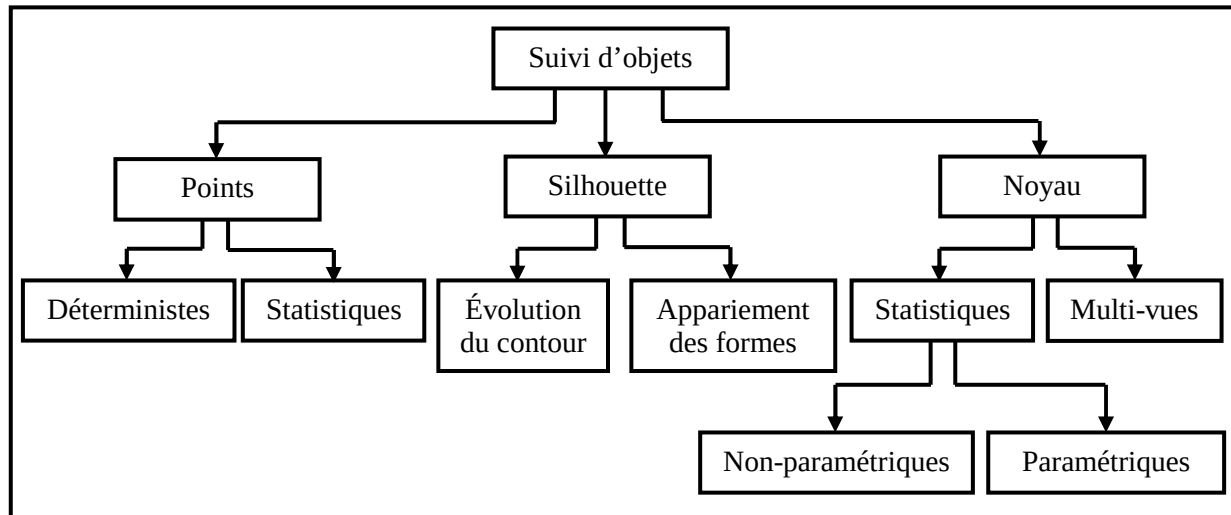


Figure 1.7 : Taxonomie des méthodes de suivi d'objets.

1.6.1 Suivi par points

Considérons qu'un objet est représenté par un ensemble de points caractéristiques appelés aussi les points d'intérêt [Serby 04]. Ces méthodes représentent la cible par des descripteurs locaux calculés sur une région autour du point considéré. Les points d'intérêt sont déterminés sur l'objet par les méthodes de détection. Parmi les détecteurs de points d'intérêt nous citons Harris [Harris 88], SIFT (Scale-Invariant Feature Transform) [Lowe 04], FAST (Features from Accelerated Segment Test) [Rosten 10] et SURF (Speeded Up Robust Features) [Bay 06]. Le suivi des points d'intérêt repose généralement sur une mesure de similarité entre les descripteurs locaux dans le but d'apparier les points de l'image courante avec les points de la cible détectés précédemment [Yang 13d, Guo 14, Hare 12]. Dans ce cas, le problème de suivi est défini comme un problème de mise en correspondance des points entre des images successives. Selon Yilmaz et al. [Yilmaz 06a], les approches de mise en correspondance des points peuvent être déterministes ou statistiques. Nous présentons par la suite un échantillon des différentes méthodes proposées dans la littérature pour chaque catégorie.

1.6.1.1 Méthodes déterministes

Le suivi d'objets par des méthodes déterministes s'effectue en minimisant une distance calculée sur certaines caractéristiques de l'objet. Les modèles des objets basés sur l'apparence peuvent être une densité de probabilité, une carte de contours ou une combinaison de ces modèles [Haritaoglu 00a, Koller 93]. Sethi et Jain [Sethi 87] résolvent le problème de mise en correspondance en utilisant une approche gloutonne (en anglais greedy) fondée sur l'ajout des contraintes de proximité et de rigidité des objets. L'initialisation de l'algorithme est faite en associant les paires de points les plus proches. Ensuite un processus heuristique consistant à

privilégier les associations qui favorise la cohérence des vitesses d'une image à l'autre est utilisé pour chercher des correspondances entre les paires de points. L'algorithme a l'avantage de simuler une prise en compte globale du problème de suivi. Cependant, il ne permet pas de prendre en compte les occultations. Les travaux de [Sethi 87, Veenman 01] introduisent une contrainte de cohérence du mouvement dans leur méthode de mise en correspondance. Cette contrainte est parfaitement adaptée au suivi des points appartenant au même objet et améliore la qualité de suivi. À contrario, cette contrainte n'est pas adaptée si les points à suivre appartiennent à des objets ayant des mouvements différents. Le suivi est dans un premier temps réalisé sans contrainte sur les deux premières images de la vidéo. La contrainte du mouvement est utilisée dès la troisième image. Cette approche permet de gérer les occultations et les erreurs de détection des points.

Yang et al. [Yang 13d] proposent un algorithme basé sur les points caractéristiques SIFT et les patches aléatoires pour modéliser la cible. Les auteurs utilisent conjointement deux modèles d'apparence à savoir un modèle de couleur-texture et un modèle de structure. Le modèle de couleur-texture inclut les patches aléatoires décrits par les histogrammes de couleurs locaux RGB (Rouge, Vert, Bleu) et les descripteurs LBP (Local Binary Patterns). Le modèle de structure consiste à un histogramme spatial global encodant la disposition des points SIFT de l'objet. Afin d'évaluer une région candidate sur l'image courante, les points caractéristiques sont détectés sur cette région, puis appariés avec ceux de l'objet à l'image précédente. L'ensemble des points appariés est utilisé pour construire l'histogramme spatial des points de l'objet. En effet, l'algorithme de suivi utilise trois types différents de caractéristiques, les distributions locales de la couleur, les caractéristiques locales de texture et les propriétés géométriques globales. Cependant le modèle de structure ne retient que les propriétés structurelles récentes vu que l'histogramme spatial ne considère que les points appariés avec ceux de l'objet sur l'image précédente.

Schmid et al. [Schmid 96a, Schmid 97b] introduisent une contrainte géométrique au processus d'appariement des points. Deux points voisins doivent être appariés à deux points également voisins dans l'image suivante. Les appariements qui ne respectent pas cette contrainte sont simplement ignorés dans la suite du processus de suivi. Cette contrainte géométrique a pour but de diminuer le nombre d'appariements incorrects. L'inconvénient majeur de cette méthode est le fait que des appariements corrects sont également éliminés. De manière similaire aux travaux de Schmid et al. [Schmid 96a, Schmid 97b], Trichet et Merialdo [Trichet 07] ajoutent une contrainte géométrique appelée triangulation de Delaunay

au processus d'appariement. Les auteurs réalisent la détection des points dans le voisinage des objets détectés à l'image précédente. La triangulation de Delaunay [Trichet 07] de chaque ensemble de points doit être respectée entre deux images consécutives. Cette approche permet de diminuer le nombre d'appariements incorrects des points. Cependant, la contrainte étant plus forte que pour les travaux de Schmid et al. [Schmid 96a, Schmid 97b] où un nombre important d'appariements corrects est éliminé.

1.6.1.2 Méthodes statistiques

Les mesures obtenues à partir des capteurs vidéo contiennent toujours du bruit, ainsi que les mouvements de l'objet peuvent subir des perturbations aléatoires comme des manœuvres des véhicules. La mise en correspondance par des méthodes statistiques résout ces types de problèmes en tenant compte de la mesure de similarité et les incertitudes du modèle lors de l'estimation de l'état de l'objet. En effet, le filtre de Kalman [Comaniciu 03c] est utilisé pour estimer l'état d'un système linéaire où l'état est supposé être distribué par une gaussienne. Le filtrage de Kalman est composé de deux étapes, la prédiction et la correction. L'étape de prédiction utilise le modèle d'état pour prédire le nouvel état des variables, tandis que l'étape de correction utilise les observations actuelles pour mettre à jour l'état de l'objet. Le filtre de Kalman a été largement utilisé dans la communauté de la vision par ordinateur pour le suivi d'objets. Broida et Chellappa [Broida 86] ont utilisé le filtre de Kalman pour suivre les points dans des images bruitées. Pour le suivi d'objets par une caméra stéréo, Beymer et Konolige [Beymer 99] utilisent le filtre de Kalman pour prédire la position et la vitesse de l'objet. Rosales et Sclaroff [Rosales 99] utilisent le filtre de Kalman étendu pour estimer la trajectoire 3D d'un objet à partir du mouvement 2D. Une limitation du filtre de Kalman est l'hypothèse que les variables d'état sont généralement distribuées par des gaussiennes. Ainsi, le filtre de Kalman donne de mauvaises estimations pour les variables d'état qui ne suivent pas une distribution gaussienne. Cette limitation peut être surmontée en utilisant le filtrage de particules [Kitagawa 87, Smith 05b].

1.6.1.3 Discussion

Parmi les problèmes rencontrés dans les méthodes de suivi par points nous citons la présence du bruit, l'occultation partielle ou totale et les erreurs lors de détection des points d'intérêt. Pour résoudre ces problèmes, les approches déterministes utilisent souvent une combinaison des contraintes basées sur les mouvements, c'est-à-dire un mouvement commun [Veenman 01] ou l'uniformité proximale [Rangarajan 91]. Cependant, les méthodes statistiques traitent explicitement le bruit en tenant compte les incertitudes du modèle. Ces

incertitudes sont habituellement sous la forme d'un bruit gaussien. Or, l'hypothèse selon laquelle les mesures sont normalement distribuées autour de leur position prédite peut ne pas les contenir et dans de nombreux cas, les paramètres du bruit ne sont pas connus. Dans le cas des hypothèses valables sur les distributions et le bruit, les filtres de Kalman [Bar-Shalom 90] offrent une solution optimale. Les méthodes de suivi par un seul point sont adaptées pour suivre des objets qui sont très petits. En outre, la représentation par un ensemble de points est nécessaire pour suivre des objets plus grands. Dans le contexte du suivi d'objets en utilisant un ensemble de points, le regroupement automatique des points qui appartiennent au même objet est un problème important. Ceci est dû à la nécessité de distinguer entre les objets multiples et l'arrière-plan. Les approches des mouvements basées sur le regroupement ou la segmentation [Vidal 04, Black 96a] supposent généralement que les points recherchés reposent sur des objets rigides afin de simplifier le problème de segmentation.

1.6.2 Suivi par silhouette

Les objets peuvent avoir des formes complexes, par exemple, les mains, la tête et les épaules qui ne peuvent pas être bien décrits par des formes géométriques simples. Les méthodes basées sur la silhouette fournissent une description précise de la forme pour ces objets [Yilmaz 04b, Xiang 08]. L'objectif d'un algorithme de suivi basé sur la silhouette est de trouver la région de l'objet dans chaque image à partir du modèle de l'objet généré à l'aide des images précédentes. Ce modèle peut être sous la forme d'un histogramme de couleurs calculé à l'intérieur des silhouettes [Kang 04], des arêtes de l'objet [Huttenlocher 93] ou du contour de l'objet. Nous divisons les algorithmes de suivi basé sur la silhouette en deux catégories : l'appariement des formes [Kang 04, Li 01b, Sato 04] et l'évolution du contour [Yilmaz 04b, Chen 01, Cremers 02]. Les approches d'appariement des formes cherchent la silhouette de l'objet dans l'image actuelle. Les approches de suivi par contour, d'autre part, développent un contour initial vers sa nouvelle position dans l'image actuelle, soit en utilisant les modèles de l'espace d'état, soit en minimisant directement certaines fonctions énergétiques.

1.6.2.1 Appariement des formes

L'appariement des formes peut être effectué d'une manière similaire au suivi basé sur l'appariement des modèles, où la silhouette de l'objet et son modèle associé sont recherchés dans l'image actuelle. La recherche est effectuée en calculant la similarité de l'objet avec le modèle généré à partir de la silhouette hypothétique de cet objet dans l'image précédente. Le modèle de l'objet est réinitialisé pour gérer les changements d'apparence dans chaque image après l'apparition de l'objet. Cette mise à jour est nécessaire pour surmonter les problèmes de

suivi liés aux points de vue et aux changements des conditions d'éclairage ainsi que le mouvement des objets non-rigides. Huttenlocher et al. [Huttenlocher 93] ont effectué une correspondance des formes en utilisant une représentation basée sur les bords de l'objet. Les auteurs utilisent la distance de Hausdorff [Hausdorff 62] afin de construire une surface de corrélation dans laquelle le minimum est sélectionné comme la nouvelle position de l'objet.

Dans [Kang 04], les silhouettes sont modélisées par les histogrammes d'arêtes et de couleurs calculées à partir des cercles couvrant la silhouette de l'objet, ayant différentes longueurs du rayon. D'autre part, un histogramme de couleurs est construit pour la silhouette en utilisant la même configuration des cercles concentriques. Les similarités entre les histogrammes d'arêtes et les histogrammes de couleurs provenant de deux images consécutives sont basées sur le coefficient de Bhattacharya [Kailath 67]. Ce modèle est invariant à la rotation, car une rotation de la silhouette est équivalente à la permutation des points de contrôle. D'autre part, l'invariance à l'échelle est également garantie par la normalisation dans l'équation de l'histogramme. Pour faire l'appariement des silhouettes dans des images consécutives, Haritaoglu et al. [Haritaoglu 00a] modélisent l'apparence de l'objet par l'information de l'arête obtenue dans la silhouette de l'objet. En particulier, le modèle d'arête est utilisé pour affiner la translation de l'objet en utilisant l'hypothèse de la vitesse constante. Ce raffinement est effectué en utilisant une corrélation binaire entre les arêtes de l'objet dans les images consécutives.

1.6.2.2 Évolution de contour

Les méthodes par évolution de contour résolvent le problème de suivi en évoluant les contours sur l'image courante à partir des contours de l'image précédente. L'évolution de contour peut être effectuée par les modèles de l'espace des états qui modélisent l'évolution de la forme et des mouvements du contour. Ces méthodes cherchent à représenter le contour par un ensemble de paramètres et de le suivre par une méthode de filtrage appropriée. Les paramètres d'un contour jouent le rôle du modèle d'état de l'algorithme de filtrage. Par exemple, dans [Terzopoulos 93], l'objet à suivre est représenté par un ensemble de points de contrôle placés le long du contour. Les auteurs ont utilisé un filtre de Kalman [Weng 06] pour prédire ces points dans l'image courante. Ceci permet d'obtenir la nouvelle position du contour. Le résultat de l'étape de prédiction est raffiné en utilisant le gradient de l'image. Les auteurs Isard et al. [Isard 98], définissent l'état de l'objet par un ensemble de splines et de paramètres du mouvement et la mise à jour de l'état de l'objet est faite par un filtrage particulière. Cette méthode a été étendue au cas du suivi multi-objets et à la gestion des

occultations [MacCormick 00]. L'avantage de ces approches est qu'elles permettent de suivre un objet avec un coût de calcul assez faible mais elles ne s'adaptent plus aux changements de la topologie des objets.

D'autres méthodes évoluent le contour par la minimisation de son énergie en utilisant un modèle de courbes déformables connu par les contours actifs "snake" [Paragios 00b, Peterfreund 00]. Ces méthodes tentent à faire évoluer une courbe placée autour de l'objet d'intérêt pour la rapprocher à des zones du fort gradient (les frontières de l'objet). L'évolution de cette courbe est le résultat de minimisation de la fonction d'énergie. Les algorithmes qui utilisent les contours actifs décrivent les objets plus efficacement et réduisent la complexité du calcul, mais ils sont très sensibles à l'étape de l'initialisation. D'autres méthodes fondées sur la minimisation de la fonction d'énergie permettent de suivre un contour ou une région malgré les changements de la topologie. La région occupée par la cible peut être définie par un masque binaire [Paragios 99a]. La fonction d'énergie est définie en utilisant l'information sous la forme du gradient temporel (flot optique) [Bertalmio 00, Mansouri 02], ou en se basant sur l'apparence de l'objet [Ronfard 94, Yilmaz 04b].

1.6.2.3 Discussion

Le suivi par silhouette est le choix approprié lorsque l'algorithme de suivi requiert la région complète de l'objet. Cette représentation est également adaptée aux objets déformables ou des formes complexes. Cependant, le suivi par silhouette présente des limites dans les environnements non contraints. D'une part, le suivi par appariement des formes utilise des mécanismes externes pour la détection des changements, par exemple, la soustraction de l'arrière-plan qui n'est pas toujours adapté aux arrière-plans dynamiques. D'autre part, le suivi par évolution du contour nécessite qu'une partie de l'objet sur l'image courante doit être en recouvrement avec une région dans l'image précédente. Cette hypothèse n'est souvent pas valide à cause des déplacements importants de la cible. En effet, ces méthodes permettent de suivre tous les types des objets de formes diverses et s'adaptent relativement bien aux objets déformables, ainsi ils sont robustes aux changements de topologie des objets mais la minimisation des fonctions d'énergie est généralement plus coûteuse et la convergence vers un minimum global n'est pas nécessairement assurée.

1.6.3 Suivi par noyaux

Le suivi par noyaux consiste à suivre un objet représenté par des formes géométriques simples comme les rectangles [Zhao 10d], les ellipses [Varcheie 11] et les cercles [Bernardin 07] afin de délimiter la région d'intérêt. Nous divisons ces méthodes de suivi en deux

catégories selon la représentation d'apparence utilisée : les approches statistiques et les approches multi-vues. Dans ce qui suit, nous présentons les méthodes les plus utilisées dans la littérature pour chaque catégorie.

1.6.3.1 Approches statistiques

Les méthodes statistiques ont été largement utilisées pour le suivi d'objets en raison de leur simplicité et de leur faible coût calculatoire. Nous distinguons deux types des méthodes dans cette catégorie à savoir, les modèles ou les templates et les approches statistiques qui utilisent les modèles d'apparence basés sur les densités des probabilités. L'approche la plus évidente pour suivre un objet en utilisant les modèles est l'appariement des templates [Birchfield 98a, Schweitzer 02]. En effet, si l'objet à suivre est de forme connue par exemple une voiture, il est relativement simple de trouver la partie de l'image la plus similaire au template considéré. Pour ce faire, la recherche est réalisée d'une manière exhaustive sur tout ou partie de l'image et la position du template dans l'image actuelle est calculée par une mesure de similarité, par exemple une corrélation croisée [Islam 09]. Les caractéristiques les plus utilisées pour former les templates sont l'intensité de l'image ou de la couleur. Le gradient de l'image est également utilisé pour sa robustesse aux variations d'illumination [Birchfield 98a]. Une limitation des approches par appariement des templates est que son coût du calcul est très élevé en raison de la recherche exhaustive du template. Pour résoudre ce problème, certains chercheurs limitent la recherche de l'objet au voisinage de leur position précédente. Cependant, parmi les avantages des templates est qu'ils incorporent l'information spatiale et celle de l'apparence de l'objet. Certaines variantes utilisent des représentations dans un domaine transformé déduites de la représentation template permettant de conserver ces propriétés, tout en diminuant la taille de descripteur. Ainsi, les auteurs de [Château 02] se basent sur les coefficients des ondelettes de Haar. En résumé, les approches basées sur les templates considèrent l'apparence de l'objet à partir d'un seul point de vue. Ainsi, ils sont seulement appropriés pour suivre des objets dont leur pose ne change pas considérablement pendant le processus de suivi.

Les approches statistiques représentent l'apparence de l'objet par des méthodes statistiques où cet apparence est modélisée par une variable aléatoire dans un espace des attributs (couleur, texture, information spatiale, etc.) avec une fonction de densité de probabilité associée. Les estimations de la densité de probabilité de l'apparence de l'objet peuvent être paramétriques, par exemple une gaussienne [Wren 97] ou un mélange de gaussiennes [Paragios 02c], ou non paramétriques, telles que les fenêtres de Parzen [Elgammal 00a] et les

histogrammes [Comaniciu 03c]. La recherche de la cible à l'instant courant se base sur des distributions, par exemple les histogrammes ou les mélanges des gaussiennes dans une forme géométrique par appariement direct entre un modèle de référence et des régions candidates. La procédure de suivi repose généralement sur l'évaluation de la ressemblance entre le modèle de référence de la cible et les régions hypothétiques selon une mesure de similarité. D'une manière générale, le calcul de la similarité est effectué en utilisant des propriétés statistiques des régions telles que la distribution des couleurs ou des intensités. Concernant les mesures de similarité les plus utilisées nous citons la corrélation croisée [Islam 09], la somme de différence au carré [Hager 04], ou encore le coefficient de Bhattacharyya [Kailath 67]. D'autres méthodes statistiques utilisent le filtre de Kalman [Broida 86] et le filtre de particules [Kitagawa 87]. Grâce à sa capacité à modéliser les changements aléatoires des états, le filtre de particules est une méthode largement utilisée, non seulement pour le suivi par forme géométrique [Islam 09, Nummiaro 03], mais aussi pour le suivi par contours [Isard 98].

Une autre méthode très populaire pour le suivi par approche statistique est la procédure Mean Shift [Comaniciu 99a, Comaniciu 03c, Comaniciu 00d] (Annexe B). L'objet à suivre est représenté par un histogramme de couleurs pondéré et calculé sur une région elliptique contenant l'objet d'intérêt. Plutôt que d'utiliser une approche exhaustive, les auteurs utilisent la méthode Mean Shift [Fukunaga 75]. L'algorithme consiste alors à déplacer une fenêtre d'analyse (noyau spatial) de manière à déterminer la fenêtre dont l'histogramme coïncide le mieux avec l'histogramme de référence. La similarité entre l'histogramme de référence et l'histogramme candidat est mesurée par le coefficient de Bhattacharyya [Kailath 67]. Le déplacement du noyau est contrôlé par une montée du gradient itérative. La pondération de l'histogramme donne plus d'importance aux pixels proches du centre et inversement. Comaniciu et al. [Comaniciu 02b] a étendu l'approche Mean Shift pour le suivi qui utilise un histogramme spatiale-couleur par l'utilisation d'un histogramme de couleurs uniquement. Un avantage évident de suivi par la technique Mean Shift par rapport à l'appariement des templates est l'élimination d'une recherche exhaustive et le calcul de la translation de l'objet dans un nombre très petit d'itérations. Cependant, le suivi par Mean Shift exige qu'une partie de l'objet doit être à l'intérieur de la région circulaire lors de l'initialisation.

Elgammal et al. [Elgammal 03b] utilisent des histogrammes contenant les informations de la couleur et de la géométrie. La similarité entre les histogrammes est évaluée en utilisant la divergence de Kullback-Leibler [Kullback 59a, Kullback 51b]. Les auteurs ont montré que le résultat du suivi est précis et robuste aux variations géométriques. L'inconvénient de la

méthode réside dans l'estimation de la divergence utilisée. En effet, l'estimation des mesures statistiques est complexe. Dans ce contexte, Boltz et al. [Boltz 09] améliorent l'approche de Elgammal et al. [Elgammal 03b] en proposant d'utiliser la recherche des plus proches voisins [Wolsztynski 06] pour l'estimation de la divergence de Kullback-Leibler. La méthode proposée permet de suivre les distributions sans calculer explicitement ces distributions. Elle consiste à ajouter, en plus de la couleur, une information sur la géométrie de l'objet. L'estimation des densités de probabilité en grande dimension n'étant pas triviale, les auteurs proposent une méthode permettant de calculer une distance de Kullback-Leibler entre deux densités de probabilités sans estimer explicitement ces densités. Une autre approche commune de suivi par approche statistique consiste à appliquer les méthodes de détection exhaustivement sur toute l'image ou sur une partie de l'image. Cette approche, souvent appelée suivi par détection (tracking-by-detection) en appliquant un classifieur binaire qui a pour but de distinguer l'objet suivi et son arrière-plan. L'approche comprend plusieurs variantes qui se distinguent par la manière d'utiliser le résultat de classification. Dans [Babenko 09] la classification est utilisée seule pour déterminer l'état final de la cible. Dans [Williams 05] le classifieur valide la prédiction de l'algorithme de suivi principal et exécute une recherche exhaustive si la cible est perdue et dans la méthode TLD (Tracking-Learning-Detection) [Kalal 12] la prédiction finale combine le résultat d'un algorithme de suivi avec celui du classifieur.

Jepson et al. [Jepson 03] propose un algorithme qui suit un objet en tant que mélange de trois composants, à savoir les caractéristiques stables de l'apparence, les caractéristiques transitoires et le processus du bruit. Le composant stable identifie l'aspect le plus fiable de l'apparence pour l'estimation du mouvement, c'est-à-dire les régions de l'objet dont l'apparence ne change pas rapidement avec le temps. Le composant transitoire identifie les pixels qui changent rapidement. Le composant du bruit gère les valeurs aberrantes dans l'apparence de l'objet qui surviennent en raison du bruit.

1.6.3.2 Approches multi-vues

Dans les approches statistiques, les modèles d'apparence utilisés sont habituellement générés en ligne. Ainsi, ces modèles représentent les informations recueillies sur l'objet à partir des observations les plus récentes. Cependant, les objets peuvent apparaître différemment et avoir de différentes vues. En effet, si la vue de l'objet change d'une manière spectaculaire pendant le suivi, le modèle d'apparence peut ne plus être valide et l'objet suivi peut être perdu. Pour surmonter ces problèmes, les différentes vues de l'objet peuvent être

appprises hors ligne et utilisées pour le suivi. Black et al. [Black 98c] ont proposé une approche basée sur le sous-espace, c'est-à-dire l'espace propre, pour calculer la transformation affine dans l'image actuelle de l'objet à l'image reconstruite à l'aide des vecteurs propres. Tout d'abord, une représentation sous-espace de l'apparence d'un objet est construite à l'aide de l'analyse en composantes principales (ACP), puis la transformation de l'image à l'espace propre est calculée en minimisant l'équation de constance sous-espace qui évalue la différence entre l'image reconstruite en utilisant les vecteurs propres et l'image d'entrée. La minimisation se fait en deux étapes: trouver les coefficients du sous-espace et calculer les paramètres affins.

Les auteurs de [Mughadam 97, Black 96a] utilisent d'autres approches par sous-espaces comme l'analyse en composantes indépendantes (ACP) pour la représentation de la forme et de l'apparence de l'objet. Une autre approche pour prendre en considération les différentes vues d'un objet est l'apprentissage d'un ensemble des classifieurs fondés, par exemple les machines à vecteur de support SVM [Avidan 04] ou les réseaux bayésiens [Park 04]. Toutefois, ces modèles d'apparence multi-vues nécessitent un nombre important de vues de référence afin de modéliser correctement l'ensemble des variations d'apparence possible du template.

1.6.3.3 Discussion

L'utilisation des caractéristiques globales calculées à l'intérieur d'une forme géométrique simple (rectangle, ellipse, cercle, etc) est une technique commune offrant une représentation souvent efficace pour modéliser l'apparence de l'objet d'intérêt avec une faible complexité. L'un des inconvénients de cette représentation est que certaines parties de l'objet peuvent rester à l'extérieur de la forme géométrique, tandis que, les pixels de l'arrière-plan interviennent dans le calcul des caractéristiques. Cet effet est dû principalement aux formes complexes des objets ne pouvant pas être décrites par des formes géométriques primitives et aux déformations que peuvent subir les objets articulés durant le suivi. Étant donnée une approche par boîte englobante, la qualité d'un système de suivi est fortement conditionnée par le choix d'un modèle d'apparence adapté. Il ne s'agit pas de coder toute l'information portée par l'objet mais de se concentrer sur l'information qui permet de définir un critère objectif efficace pour déterminer la position réelle de l'objet suivi dans l'image. Sans dissocier l'extraction des attributs de la structuration de l'information sous forme des descripteurs et de l'utilisation de cette représentation pour la localisation de l'objet, une des clés de la représentation efficace est l'identification des caractéristiques primaires en accord avec le type et le but de suivi visé par le système.

1.6.4 Suivi d'objets en estimant l'échelle et l'orientation

Malgré les avancées importantes réalisées par de nombreux algorithmes de suivi d'objets, leur performance souffre de l'utilisation d'une fenêtre de taille fixe. En effet, une taille de fenêtre qui fonctionne à une échelle ne convient pas à une autre lorsque la cible se déplace vers et loin de la caméra. Dans la suite de cette section, nous présentons les méthodes de suivi fondées sur l'apparence de l'objet, et plus particulièrement celles qui sont basées sur l'adaptation de l'échelle et le changement d'orientation de la cible. Cependant, les recherches pour répondre à ces problèmes peuvent être classées en trois catégories [Yu 15].

La première catégorie est les méthodes heuristiques basées sur l'incrémementation. Dans l'approche de Comaniciu et al. [Comaniciu 03c] l'algorithme Mean Shift est effectué à trois niveaux différents d'échelles : l'échelle précédente, l'échelle légèrement plus grande et l'échelle légèrement plus petite. Pour chaque taille de la fenêtre, la mesure de similarité basée sur le coefficient Bhattacharyya [Kailath 67] est calculée pour la comparaison. La taille du noyau qui donne la plus grande mesure est choisie comme meilleure échelle. Ce principe a été utilisé par de nombreux auteurs [Bousetouane 13, Parameswaran 06]. Cependant, cette approche ne fonctionne pas bien dans le cas des changements brusques de la taille de l'objet, ainsi que les petites fenêtres ont généralement une similarité plus élevée et par conséquent l'échelle est souvent sous-estimée. Collins [Collins 03a] a étendu l'algorithme Mean Shift en adaptant la théorie de Lindeberg pour la sélection de l'échelle en se basant sur les maximaux locaux des filtres différentiels espace-échelle, et l'algorithme Mean Shift a été appliqué dans les dimensions spatiales et échelles des poids de l'image. Cette approche utilise le suivi par boîte englobante et un noyau de l'échelle pour saisir avec précision la variation de l'échelle de la cible. Cependant, cette méthode nécessite de recalculer les noyaux gaussiens à chaque itération de l'algorithme Mean Shift, ce qui demande un calcul très coûteux; également elle ne peut pas gérer les changements de rotation de la cible. Un autre travail similaire est le suivi par Mean Shift en utilisant une pyramide adaptative a été proposé par Li et al. [Li 10a]. Différents seuils pour l'incrémementation de l'échelle ont été appliqués. Cette méthode de sélection d'échelle permet de suivre des objets de différentes tailles mais avec une complexité stable et le processus de suivi peut être interrompu par un changement brusque d'illumination.

La deuxième catégorie est les méthodes basées sur les moments spatiales. Bradski [Bradski 98] modifie la procédure Mean Shift et propose l'algorithme CAMSHIFT (Continuously Adaptive Mean Shift) pour le suivi du visage. Cet algorithme ne considère que les descripteurs de moments du poids de l'image déterminé par le modèle de la cible afin

d'estimer l'échelle et l'orientation de l'objet. Zivkovic et al. [Zivkovic 04a] ont proposé un algorithme de suivi similaire qui estime les changements de l'échelle et de l'orientation nommé Expectation-Maximisation Shift (EM-Shift). Ces auteurs ont simultanément estimé la position et la matrice de covariance qui décrit la forme de la cible en calculant le moment spatial de poids de l'image. Yang et al. [Yang 05b] ont introduit une nouvelle mesure de similarité qui estime l'échelle en comparant les moments d'ordre 2 du modèle de la cible et de la cible candidate. Ning et al. [Ning 12c] ont proposé un algorithme de suivi pour l'échelle et l'orientation adaptative par l'algorithme Mean Shift (SOAMST), ces auteurs utilisent les descripteurs de moments pour déterminer les changements de l'échelle et l'orientation de la cible. En résumé, toutes ces méthodes sont basées sur le poids de l'image pour estimer l'échelle et l'orientation de la cible qui est sensible à l'illumination et à l'encombrement de l'arrière-plan, ainsi qu'ils échouent dans le cas des objets ayant une couleur similaire avec l'arrière-plan.

La troisième catégorie est les méthodes basées sur la mise en correspondance des caractéristiques où les caractéristiques locales entre deux images consécutives contiennent des informations sur les changements de l'échelle et de l'orientation. Peng et al. [Peng 05] ont proposé une sélection automatique de bande passante pour exploiter une stratégie de mise en correspondance des caractéristiques afin d'estimer la variation de l'échelle basée sur le suivi en arrière et l'enregistrement du centre de l'objet. Les auteurs de [Zhou 09c] utilisent les caractéristiques SIFT (Scale Invariant Feature Transform) et la procédure Mean Shift pour le suivi d'objets afin d'estimer la position et l'échelle de l'objet. De même, [Liang 07, Zhao 08b] reposent sur les caractéristiques de support et l'algorithme Mean Shift pour l'estimation de l'échelle. Liang et al. [Liang 07] cherchent les limites de la cible en corrélant l'image avec quatre modèles. Les positions des frontières déterminent directement l'échelle de la cible. Zhao et al. [Zhao 08b] exploitent la structure affine pour récupérer l'échelle relative de la cible à partir des points caractéristiques correspondants entre des images consécutives. En résumé, les méthodes basées sur la mise en correspondance des caractéristiques sont capable d'estimer l'échelle, mais elles ne peuvent pas être parfaitement intégrées dans l'algorithme Mean Shift. En outre, l'estimation de l'échelle à partir de la mise en correspondance des caractéristiques prend du temps, ce qui nécessite la présence des caractéristiques bien localisées qui peuvent être détectées avec une répétabilité élevée. Finalement, toutes ces méthodes citées au-dessus ne considèrent pas l'influence de l'arrière-plan sur la localisation de la cible, d'où, elles ne sont pas assez discriminantes lorsque l'objet et son arrière-plan ont des couleurs similaires.

1.6.5 Suivi d'objets par des histogrammes

Les méthodes de suivi basées sur les histogrammes sont très populaires dans le suivi d'objets en raison de leur efficacité dans la capture de la distribution des caractéristiques visuelles à l'intérieur de la région d'objet. Selon Li et al. [Li 13c], les représentations par histogramme contiennent deux catégories : mono-caractéristique et multi-caractéristiques.

1.6.5.1 Représentation par un histogramme mono-caractéristique

Ces méthodes utilisent souvent un histogramme pour prendre les informations de la distribution dans la région de l'objet. Actuellement, une caractéristique largement utilisée pour la représentation de la cible est l'histogramme de couleurs. Le succès des approches par cet histogramme provient de leur faible complexité calculatoire associée à une bonne robustesse vis-à-vis du bruit et de leur invariance aux rotations et aux changements de l'échelle. Le premier travail de Swain et Ballard [Swain 91] ont utilisé l'histogramme de couleurs comme une caractéristique visuelle globale, en démontrant que la couleur peut être exploitée comme une caractéristique utile pour une détection rapide. Par la suite, Bradski [Bradski 98] utilise un histogramme de couleurs dans l'espace HSV (Hue Saturation Value) pour la représentation de l'objet, puis incorpore l'histogramme de couleurs dans l'algorithme CAMSHIFT (Continuously Adaptive Mean Shift) pour le suivi d'objets. Comaniciu et al. [Comaniciu 03c] utilisent un histogramme de couleurs spatialement pondéré dans l'espace RGB pour la représentation visuelle de l'objet et ensuite incorporent cet histogramme dans le cadre de suivi par Mean Shift afin de localiser l'objet. Zhao et al. [Zhao 10d] transforment le problème de suivi d'objet dans celui de correspondance des distributions de couleurs RGB entre les images. Par conséquent, la localisation de l'objet est prise en utilisant un différentiel rapide EMD (Earth Mover's Distance) pour calculer le coefficient de similarité entre la cible apprise et la région candidate.

En résumé, la représentation par histogramme avec une seule caractéristique est capable de coder efficacement l'information de la distribution statistique des caractéristiques visuelles dans la région de l'objet. Cependant, l'utilisation directe de l'histogramme de couleurs pour caractériser la cible peut entraîner la perte de l'information spatiale, ainsi qu'elle est très sensible quand l'apparence de l'objet et son arrière-plan sont similaires, aux changements d'illumination et à l'influence de l'arrière-plan. D'où, pour mettre en évidence les défauts de la description de l'apparence en utilisant uniquement les caractéristiques de couleur, de nombreux chercheurs présentent un certain nombre d'améliorations en combinant plusieurs

caractéristiques afin de trouver les plus robustes pour le suivi d'objets.

1.6.5.2 Représentation par un histogramme multi-caractéristiques

Ces méthodes visent à incorporer plus d'informations dans la représentation de la cible afin d'améliorer la robustesse des algorithmes de suivi. Les auteurs de [Georgescu 04, Yang 05b] utilisent les caractéristiques (x, y, R, G, B) comme un espace de couleur spatial. La représentation par un histogramme de couleur spatial encode habituellement l'information spatiale en modélisant l'apparence de l'objet dans un espace des caractéristiques spatiales ou en utilisant une technique de division par patch. Birchfield et al. [Birchfield 05b] utilisent un spatiogramme qui contient la moyenne et la covariance pour chaque classe d'histogramme. Cependant, la stratégie de division par patch consiste à coder l'information spatiale dans les modèles d'apparence en divisant la région de suivi en un ensemble de patches [Adam 06]. En considérant la relation géométrique entre les patches, elle est capable de capturer l'information spatiale.

Pour exploiter l'information de texture dans la représentation de l'objet, de nombreux chercheurs ont proposé un certain nombre d'améliorations. Par exemple, Bousetouane et al. [Bousetouane 13] utilisent une combinaison des caractéristiques couleur-texture et l'algorithme Mean Shift [Comaniciu 03c] pour le suivi d'objets. Ces auteurs utilisent les caractéristiques de Haralick [Haralick 73] pour représenter la texture. La représentation proposée est invariante aux rotations, aux changements de l'illumination et de l'échelle. Zhang et al. [Zhang 10a] utilisent la transformée par ondelettes discrète DWT (Discrete Wavelet Transform) pour extraire les caractéristiques de texture, ensuite, l'histogramme a été calculé à partir des coefficients d'ondelettes. Ning et al. [Ning 09b] utilisent la technique des motifs binaires locaux LBP (Local Binary Patterns) pour extraire les caractéristiques de texture. Cependant, les deux méthodes DWT et LBP extraient uniquement les caractéristiques des coins et des frontières dans la région de la cible, ce qui est très utile pour accélérer la convergence de l'algorithme Mean Shift [Comaniciu 03c]. Les auteurs de [Xiaorong 12] utilisent un nouveau descripteur de texture tel que les motifs complets ternaires locaux CLTP (Completed Local Ternary Pattern) pour représenter l'information structurelle de la cible, cette technique est plus discriminante et moins sensible au bruit. Par conséquent, ces méthodes imposent des contraintes de texture sur l'apparence de l'objet ou de la scène ce qui n'est pas assez discriminante dans certaines conditions. En outre, les caractéristiques de texture deviennent inefficaces, lorsque les objets ayant de grandes régions lisses.

Une autre direction de recherche [Wang 16c, Yilmaz 04b] c'est d'ajouter la forme de l'objet aux caractéristiques couleur-texture pour récupérer les parties occultées de l'objet lors de l'occultation. Ces auteurs ont démontrés que l'utilisation multi-caractéristiques peut se traduire par une meilleure performance du suivi contre les distractions. Toutefois, lorsqu'il n'y a pas d'occultation, la forme de l'objet peut avoir des inconvénients, particulièrement dans le cas des objets similaires avec différentes formes ou lorsque la taille de l'objet change rapidement. Polat et Ozden [Polat 06] combinent les caractéristiques de couleur et de mouvement déterminées par la technique d'estimation de la densité du noyau. La méthode proposée à l'avantage de gérer les défaillances de suivi causées par des distributions des couleurs similaires entre l'objet et son arrière-plan. Haritaoglu et Flickner [Haritaoglu 01b] incorporent les informations du gradient ou de la forme dans la représentation visuelle basée sur l'histogramme de couleurs.

Malgré les avantages réalisés par les méthodes mentionnées précédemment pour le suivi d'objets, elles sont souvent affectées par les distractions de l'arrière-plan avec des couleurs similaires à l'objet suivi. Cependant, la cible ne peut pas être correctement représentée sans aucune information de l'arrière-plan et si la corrélation entre la cible et son arrière-plan est élevée, la précision de localisation de l'objet sera diminuée. Pour résoudre ces problèmes, des techniques étendues ont été présentées dans la littérature en incorporant des caractéristiques de l'arrière-plan dans la représentation de la cible. Comaniciu et al. [Comaniciu 03c] suggèrent un histogramme pondéré de l'arrière-plan BWH (Background Weighted Histogram). L'objectif de cette stratégie est de dériver une représentation simple des caractéristiques de l'arrière-plan et de l'utiliser pour sélectionner les caractéristiques principales du modèle de la cible et la cible candidate. Jeyakar et al. [Jeyakar 08] ont proposé une approche basée sur un fragment-pondéré dans lequel, le modèle de la cible de chaque fragment est amélioré par l'histogramme BWH. Les coefficients de pondération sont dérivés de la différence entre les couleurs du fragment et l'arrière-plan. Par la suite, Ning et al. [Ning 10a] ont prouvé que les poids attribués aux pixels dans la région candidate de la cible par l'histogramme BWH sont proportionnels à ceux sans information de l'arrière-plan. Ensuite, ils ont proposé un nouveau modèle intitulé l'histogramme pondéré et corrigé de l'arrière-plan CBWH (Corrected Background-Weighted Histogram) en transformant seulement le modèle de la cible. Une autre amélioration d'histogramme pondéré a été présentée par Wang et al. [Wang 14a], l'approche proposée combine l'histogramme pondéré de l'arrière-plan basé sur la vraisemblance LBWH (Likelihood Background-Weighted Histogram) et l'histogramme

CBWH. La combinaison CB-LBWH est fondée sur la notion que la représentation de la cible utilise à la fois les caractéristiques saillantes et de confiances pour former un critère d'histogramme composé et pondéré. Wang et al. [Wang 16c] introduisent les caractéristiques saillantes du premier-plan dans la modélisation de l'arrière-plan afin de dériver un nouveau histogramme de l'arrière-plan FSBWH (Foreground Saliency-based Background-Weighted Histogram) pour la représentation de la cible. Toutes ces méthodes présentent des bons résultats en présence d'occultation et une grande variation de l'arrière-plan. Cependant, elles ne sont pas assez discriminantes lorsque l'objet et son arrière-plan ont des couleurs similaires.

1.7 Mesures de similarité

La définition d'une similarité est la deuxième face de la définition d'un modèle d'apparence. En effet, une fois l'apparence de l'objet représentée numériquement sous la forme d'un descripteur, il faut pouvoir comparer un tel descripteur à un modèle de référence afin de quantifier s'il s'agit bien de l'objet cible ou pas. Cette comparaison doit être robuste aux altérations dues au bruit, de distorsions spatiales, ou d'autres phénomènes physiques. Le but est de définir une mesure de similarité entre l'objet cible et les objets candidats et qui permet au système d'effectuer une telle évaluation à la fois de façon fiable et rapide. Les mesures de similarité ont été largement analysées, étudiées et critiquées par différentes théories de la similarité [Patella 99]. Un modèle de similarité doit être suffisamment objectif (dans le sens où il limite efficacement le comportement perceptuel humain) et efficace (dans le sens où il permet de calculer rapidement une mesure de similarité). Parmi les mesures de similarité existantes dans la littérature nous distinguerons plusieurs catégories à savoir les mesures dans un espace vectoriel générique, adaptées à des représentations vectorielles directes, telles que les templates, les mesures adaptées à des représentations de la distribution statistique des attributs, parmi lesquelles nous distinguerons les approches bin-à-bin pour lesquelles les fonctions de densité de distribution sont comparées pour des valeurs d'attribut correspondant exactement et les approches inter-bins pour lesquelles une certaine flexibilité est autorisée dans la mise en correspondance de valeurs d'attribut voisines [Mikram 08].

1.8 Évaluation des algorithmes de suivi d'objets

La diversité des algorithmes de suivi existants a fait naître le besoin d'effectuer une évaluation approfondie. Se posent alors plusieurs questions au sujet de ces algorithmes, notamment au niveau de leur efficacité à savoir la précision de la localisation et la complexité

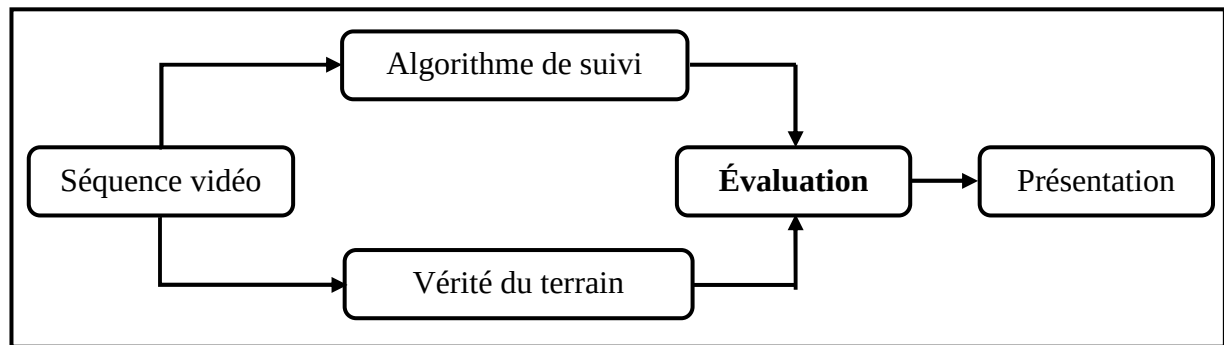


Figure 1.8 : Évaluation des performances d'un algorithme de suivi d'objets.

calculatoire. L'approche classique pour répondre à cette difficulté consiste à considérer l'algorithme de suivi comme une boîte noire, c'est-à-dire prenant en entrée une vidéo et une initialisation éventuelle, et fournissant en sortie les positions et les trajectoires estimées des objets détectés dans la vidéo. Différentes méthodes pour la mesure des performances des algorithmes de suivi ont déjà été proposées [Black 03b, Schneiders 05, Brown 05, Bashir 06]. Chacune de ces méthodes évalue les performances grâce à un certain nombre des mesures sur la qualité de la localisation estimée par l'algorithme de suivi. Ces mesures se fondent sur les bases de données [Benchmark 17] ou [CAVIAR] auquel est associée une vérité du terrain qui capture l'interprétation vraie de la scène en termes des objets à suivre.

Pour permettre l'évaluation des performances des algorithmes de suivi, plusieurs étapes doivent être prises en considération. Tout d'abord, les séquences vidéo sur lesquelles l'évaluation sera effectuée doivent être disponibles. Deuxièmement, il y a des résultats de suivi qui doivent être évalués. Troisièmement la vérité du terrain doit être générée et stockée. Puis, la vérité du terrain doit être comparée aux résultats produits utilisant des métriques de performance. Finalement, les résultats sont combinés pour chaque séquence vidéo pour une présentation à l'utilisateur. Par conséquent, nous distinguerons quatre thèmes principaux pour l'évaluation comme illustré sur la figure 1.8 ci-dessous, à savoir, la création des données de la vérité du terrain, l'ensemble des données disponibles pour l'évaluation, les métriques de performance et la présentation des résultats d'évaluation. L'évaluation de performance des algorithmes de suivi exige une comparaison des résultats des algorithmes avec des résultats optimaux nommés habituellement la vérité du terrain. De plus, pour fournir une évaluation fidèle, les données utilisées doivent être représentatives et contenir des scénarios typiques et difficiles. L'évaluation des performances des algorithmes de suivi a donné lieu à plusieurs projets de constitution de tels benchmarks [Benchmark 17]. Les premiers efforts de grande envergure dans l'évaluation des performances de suivi d'objets proviennent des workshops PETS (Performance Evaluation of Tracking and Surveillance) organisés par l'IEEE. En 2002

et 2001, le workshop a fourni des données de benchmarks de surveillance pour que les participants évaluent leurs systèmes. D'autres projets d'évaluation des performances des systèmes de suivi ont également été proposés [CAVIAR, CLEAR, CREDS, ETISEO, VACE] chacun avec un objectif spécifique.

1.9 Conclusion

Dans ce chapitre, nous avons dressé les modèles et les principes nécessaires pour le suivi d'objets en vision par ordinateur. Tout d'abord, nous avons expliqué le principe de suivi d'objets. Ensuite, les représentations d'objets basées sur leurs formes et leurs apparences et les caractéristiques les plus utilisées pour le suivi d'objets ont été présentées. Nous avons consacré une partie pour présenter les descripteurs statistiques de la couleur. Puis, nous avons exposé les méthodes de suivi d'objets les plus répandues dans la littérature et utilisant les techniques détaillées dans les parties précédentes. Par ailleurs, nous avons décrit les avantages et les inconvénients de l'ensemble, afin de développer une idée générale sur les différentes approches existantes et d'extraire les informations pertinentes pour le suivi d'objets. Finalement, nous avons présenté quelques techniques pour l'évaluation des algorithmes de suivi d'objets. L'ensemble de ces rappels préliminaires est l'occasion de présenter et d'organiser les travaux antérieurs sur lesquels se basent nos approches proposées dans les chapitres suivants.

Chapitre 2

Méthode de suivi basée sur les points d'intérêt couleurs

2.1 Introduction

Le suivi d'objets dans une séquence d'images est une problématique difficile qui se pose dans un grand nombre d'applications de la vision par ordinateur. En effet, les méthodes de suivi se basent principalement sur le choix de la représentation des objets et sur une méthode de suivi. Le choix de la représentation consiste à choisir des caractéristiques visuelles significatives et fiables qui décrivent le contenu visuel de ces objets tout en permettant de les reconnaître au cours du temps, malgré les changements des conditions environnementales et les variations de l'apparence le long de la séquence d'images. Cependant, parmi toutes les caractéristiques locales qui peuvent être extraites des vidéos nous trouvons les points d'intérêt couleurs qui sont particulièrement intéressants car ce sont des caractéristiques de bas niveau simples, robustes et qui permettent une bonne caractérisation des objets à suivre. Dans ce chapitre, nous présentons une nouvelle approche pour le suivi des personnes dans une séquence d'images basée sur les points d'intérêt couleurs.

2.2 Outils de suivi

Dans le domaine de suivi d'objets, les prétraitements diffèrent d'un système à un autre. Ils dépendent de l'environnement du travail, de la qualité des images et des objectifs visés par le système de suivi. Généralement, le suivi des personnes s'effectue dans la plupart des cas dans un environnement non-contrôlé à cause de différentes difficultés rencontrées comme les changements d'illumination, les mouvements des personnes, les occultations, etc. Ceci nous a amené à étudier plusieurs techniques de prétraitements pour faciliter la tâche de la localisation des personnes dans les images successives d'une vidéo. Premièrement, nous allons nous concentrer sur des prétraitements adaptables à l'environnement statique. Pour cette raison, la soustraction de l'arrière-plan [Stauffer 99] semble intéressante afin de détecter les personnes en mouvement et restreindre la zone de recherche. Deuxièmement, la couleur de la peau contient une forte composante rouge, ce qui donne la possibilité d'isoler les régions de la peau pour localiser plus précisément les personnes dans les images. D'où, la combinaison de ces

deux techniques de prétraitements peut mener à faciliter la localisation des personnes dans une séquence d'images. Nous présentons dans la suite de cette section les méthodes que nous avons utilisées pour la soustraction de l'arrière-plan, la détection de la couleur de la peau et la détection des points d'intérêt couleurs.

2.2.1 Soustraction de l'arrière-plan

La première étape lors d'un processus de suivi d'objets consiste en la détection de ces objets à suivre. En effet, la détection faut intervenir au début de la séquence d'images ou à tout moment si des objets peuvent potentiellement entrer dans le champ de la caméra. Cette étape permet de simplifier les traitements ultérieurs en localisant les régions d'intérêt dans l'image pour optimiser le temps de calcul. Ensuite, le nombre de fausses détections est également réduit puisque les régions de l'image où la probabilité de trouver un objet est très faible sont éliminées. Pour notre application, les régions d'intérêt détectées correspondent aux régions de l'image où il y a une forte probabilité qu'il y ait une personne. La détection d'objets peut être réalisée en construisant une représentation de la scène appelée le modèle de l'arrière-plan, puis en trouvant des écarts par rapport au modèle pour chaque image entrante. La modélisation de l'arrière-plan regroupe toutes les méthodes de détection des personnes qui consistent à créer un modèle de l'arrière-plan de la scène filmée sans aucun objet mobile, et la soustraction de l'arrière-plan est l'opération qui suit logiquement la modélisation de l'arrière-plan.

Dans la formulation la plus simple, le modèle de l'arrière-plan peut être une image correspondant à un instant où il n'y avait pas d'objet en mouvement dans la scène comme dans les travaux de [Cucchiara 03, Yang 04c]. Dans ce cas, une différence en valeur absolue entre ce modèle et l'image courante est effectuée afin de détecter les objets en mouvement. Cependant, cette formulation n'est évidemment pas adaptée pour les applications réelles car le modèle doit être capable de gérer certaines difficultés comme les changements d'illumination brusques ou progressifs, les ombres et les réflexions, les textures animées dans l'arrière-plan (les branches des arbres, les écrans de télévision, etc.), l'introduction ou le retrait des objets statiques dans la scène et les secousses de la caméra. De nombreux modèles de l'arrière-plan ont été présentés dans la littérature pour gérer ces difficultés, par exemple la modélisation de l'arrière-plan par un modèle statistique décrivant la fonction de distribution des niveaux de gris ou des couleurs en tout point [Wren 97, Stauffer 99, Elgammal 00a, Zivkovic 06b]. Dans ce cas, nous calculons la probabilité que chaque pixel appartienne à l'arrière-plan en testant la valeur observée dans le modèle. D'autres chercheurs utilisent une base d'images

caractéristiques qui constitue un sous-espace vectoriel dans lequel ils considèrent que les pixels représentant l'arrière-plan vont se trouver [Oliver 00]. Toutes ces méthodes sont basées sur l'hypothèse que l'objet en mouvement est d'une couleur différente de l'arrière-plan. Ces méthodes identifient, comme appartenant au premier-plan, les pixels à l'instant t dont la couleur diffère de la couleur de ce même pixel dans l'arrière-plan [Zhou 01a]. Nous appelons "arrière-plan" l'union de toutes les zones statiques de l'image et "premier-plan" les zones de l'image où un changement a été détecté. Nous nous focaliserons dans la suite de cette section à la modélisation paramétrique de l'arrière-plan pour réaliser la soustraction de ce dernier.

2.2.1.1 Modélisation gaussienne de l'arrière-plan

La modélisation gaussienne de l'arrière-plan [Zivkovic 06b, Wren 97] est exprimée d'un point de vue statistique. Il s'agit pour chaque pixel d'estimer la probabilité d'observer la couleur ou le niveau de gris en se basant sur un modèle appris qui représente l'arrière-plan de la scène. Le modèle consiste en un ensemble de fonctions de densité de probabilité (une par pixel de l'image). Les mesures dont la probabilité d'être observées est élevée correspondent à des pixels qui seront étiquetés comme arrière-plan, tandis que celles dont la probabilité d'être observées est faible correspondent à des pixels qui seront étiquetés comme premier-plan. Cependant, de nombreux auteurs modélisent chaque pixel de l'arrière-plan par une densité de probabilité déterminée à partir d'une série d'images d'apprentissage [Zivkovic 06b]. Dans ce cas, la soustraction de l'arrière-plan est réalisée en seuillant la fonction de densité de probabilité. Par exemple, les auteurs Wren et al. [Wren 97] proposent de modéliser l'intensité des points de l'arrière-plan par une distribution gaussienne. En tout point, la moyenne et l'écart-type sont mis à jour récursivement, et chaque nouvelle observation est déclarée comme appartenant à l'arrière-plan si elle se situe suffisamment près de la moyenne courante, compte tenu de l'estimation courante de l'écart-type. Néanmoins, la modélisation gaussienne peut être utilisée sur des images couleurs [McKenna 00]. Ce modèle permet d'obtenir des bons résultats pour les scènes à l'intérieur où l'arrière-plan est parfaitement statique, mais pour l'environnement extérieur, les phénomènes périodiques tels que l'ondulation d'une surface de l'eau ou le balancement d'une branche d'arbre peuvent le rendre totalement inopérant car la distribution de l'apparence de l'arrière-plan est alors multimodale.

2.2.1.2 Modélisation avec un mélange de gaussiennes

Modéliser un pixel de l'arrière-plan avec une distribution gaussienne n'est quelques fois pas suffisant. Un même pixel peut par exemple représenter le ciel à un instant donné puis une feuille d'arbre à un autre instant. Pour prendre en compte les arrière-plans qui ont des textures

animées (les vagues sur la mer, les branches d'un arbre agitées par le vent, les écrans de télévision ou d'ordinateur, etc.), des fonctions de densité de probabilité multimodales avec un mélange de gaussiennes ont été proposées [Lee 05a, Bouwmans 08]. Par exemple, les auteurs Stauffer et al. [Stauffer 99] modélisent chaque pixel de l'arrière-plan par un mélange de K gaussiennes. La probabilité d'occurrence d'une couleur d'un pixel s à l'instant t est représentée par :

$$P(I_{s,t}) = \sum_{i=1}^K \omega_{i,s,t} * f(I_{s,t}, \mu_{i,t,s}, \Sigma_{i,s,t}) \quad (2.1)$$

où K est le nombre de distributions compris entre 3 et 5 [Stauffer 99], $\omega_{i,s,t}$ est le poids accordé à la $i^{\text{ème}}$ gaussienne à l'instant t (quelle portion des données est prise en compte pour cette gaussienne), $\mu_{i,s,t}$ et $\Sigma_{i,s,t}$ sont respectivement la moyenne et la matrice de covariance pour la $i^{\text{ème}}$ gaussienne à l'instant t , $I_{s,t}$ est la valeur du pixel s à l'instant t et $f(I_{s,t}, \mu_{i,s,t}, \Sigma_{i,s,t})$ est la fonction gaussienne de densité de probabilité définie par :

$$f(I_{s,t}, \mu_{i,s,t}, \Sigma_{i,s,t}) = \frac{1}{\sqrt{(2\pi)^K |\Sigma_{i,s,t}|}} \exp \left[-\frac{1}{2} (I_{s,t} - \mu_{i,t,s})^T \Sigma_{i,s,t}^{-1} (I_{s,t} - \mu_{i,t,s}) \right] \quad (2.2)$$

Cependant, pour limiter les temps de calcul, Stauffer et Grimson [Stauffer 99] suppose que les trois composantes de couleur rouge, verte et bleue sont indépendantes et ont les mêmes écarts-types, alors la matrice de covariance $\Sigma_{i,s,t}$ est estimée par une matrice diagonale comme suit :

$$\Sigma_{i,s,t} = \sigma_{i,s,t}^2 \cdot Id \quad (2.3)$$

avec Id représente la matrice identité de taille 3×3 . Pour chaque nouvelle observation $I_{s,t}$, la gaussienne qui est susceptible d'expliquer $I_{s,t}$ avec la plus forte probabilité est mise à jour. Nous considérons que la gaussienne i explique $I_{s,t}$ si $\|I_{s,t} - \mu_{i,t}\| \leq 2.5\sigma_{i,t}$. La mise à jour de la gaussienne i se déroule de la manière suivante :

$$\omega_{i,s,t} = (1 - \alpha) \cdot \omega_{i,s,t-1} + \alpha \quad (2.4)$$

$$\mu_{i,s,t} = (1 - \rho) \cdot \mu_{i,s,t-1} + \rho \cdot I_{s,t} \quad (2.5)$$

$$\sigma_{i,s,t}^2 = (1 - \rho) \sigma_{i,s,t-1}^2 + \rho (I_{s,t} - \mu_{i,s,t-1})^T (I_{s,t} - \mu_{i,s,t-1}) \quad (2.6)$$

avec α correspond au taux d'apprentissage défini par l'utilisateur ($0 \leq \alpha \leq 1$) et ρ correspond à un second taux d'apprentissage défini par $\rho = \alpha f(I_{s,t}, \mu_{i,s,t}, \Sigma_{i,s,t})$ [Stauffer 99].

Après avoir mis à jour chaque gaussienne, les poids $\omega_{i,s,t}$ sont normalisés. Pour décider parmi les gaussiennes appartenant au mélange lesquelles modélisent l'arrière-plan et lesquelles modélisent les objets en mouvement, les K distributions sont ordonnées selon leur rapport de $\omega_{i,s,t}/\sigma_{i,s,t}$. Une distribution modélise l'arrière-plan si elle apparaît fréquemment ($\omega_{i,s,t}$ grand) et varie peu ($\sigma_{i,s,t}$ faible). Les distributions de l'arrière-plan les plus probables restent alors les premières. Donc, seules les H premières distributions les plus fiables sont attribuées à l'arrière-plan avec :

$$H = \underset{h}{\operatorname{argmin}} \left(\sum_{i=1}^h \omega_i > \theta \right) \quad (2.7)$$

où θ est un seuil fixé empiriquement. Si θ est faible, une distribution unimodale représentera l'arrière-plan. Inversement, si θ est grand, plusieurs distributions représenteront l'arrière-plan (dans la limite du nombre maximal de gaussiennes K). La détection est finalement réalisée par :

$$\mathcal{X}_t(s) = \begin{cases} 1 & \text{si } d_M(I_{s,t}, \mu_{s,t}) > \tau \quad \forall i \leq H \\ 0 & \text{sinon} \end{cases} \quad (2.8)$$

avec $\mathcal{X}_t(s)$ est le masque du premier-plan à l'instant t , d_M est la distance entre $I_{s,t}$ et $\mu_{s,t}$ et τ est un seuil. Les pixels de $\mathcal{X}_t$ à 1 correspondent au premier-plan et les pixels à 0 correspondent à l'arrière-plan.

Discussion : La plupart des méthodes de suivi d'objets par des caméras statiques [Haritaoglu 00a, Collins 01b] utilisent les méthodes de soustraction de l'arrière-plan pour détecter les régions d'intérêt. La limitation la plus importante de la soustraction de l'arrière-plan est l'exigence des caméras fixes. Le mouvement de la caméra généralement déforme les modèles de l'arrière-plan. Pour surmonter ces limites, certains chercheurs ont proposés de nombreuses solutions; par exemple la création d'un arrière-plan en forme mosaïque avec une caméra PTZ (Pan Tilt Zoom) [Irani 98, Suhr 11], ou en compensant le mouvement de la caméra [Murray 94]. Cependant, ces solutions nécessitent des hypothèses de scènes planes et des petits mouvements dans des images successives. En effet, parmi les techniques de soustraction de l'arrière-plan qui existent dans la littérature, nous avons présenté les méthodes les plus répandues et qui nous semblent adaptables à notre application de suivi. Après cette étude, nous retenons la modélisation de l'arrière-plan avec un mélange de gaussiennes sur laquelle nous ferons des tests approfondis par la suite. Ce choix est motivé par sa simplicité, sa robuste au changement d'illumination et de position des objets dans l'arrière-plan, et la

perspective d'une application sur une zone restreinte caractérisée par l'existence des personnes.

2.2.2 Détection de la couleur de la peau

La détection de la peau est une technique très importante utilisée dans de nombreuses applications de suivi des personnes [Pérez 02, Argyros 04]. Elle est souvent une étape préliminaire et indispensable pour des problèmes divers tels que la détection de visage qui composent la majorité des systèmes utilisant ce type d'algorithmes [Hsu 02], la détection des mains [Girondel 02], ou encore la détection des personnes dans une image [Fleck 96]. Un point commun entre ces applications est les régions de la peau, où la détection de la peau va permettre une réduction considérable de l'espace de recherche des personnes qui va se limiter uniquement aux régions de la peau. En effet, la détection de la peau n'est pas une tâche facile pour des raisons diverses telles que les conditions d'éclairage généralement inconnues (figure 2.1 (a)), la diversité ethnique avec des teintes qui varient selon la personne et selon les différentes races (figure 2.1 (b)), sans oublier qu'un décor complexe peut compliquer cette tâche (figure 2.1 (d)).



Figure 2.1 : Problèmes de détection de la peau. (a) changement d'illumination, (b) richesse ethnique avec des teintes variées et (c) décors complexes.

Dans une image donnée, il revient à trouver les pixels et les régions de l'image correspondant à la peau. Son objectif est de construire une règle de décision, qui discrimine entre les pixels de la peau et de non-peau. Généralement, ce processus se déroule en trois étapes : le choix d'un espace de couleur convenable (RGB, YCbCr, HSV, etc.), la modélisation de la couleur de la peau, c'est-à-dire la détermination d'un ensemble de règles qui permettent de définir les régions de la peau et la classification des pixels en pixels de peau et de non peau [El Fkihi 08, Phung 05]. L'étape de classification consiste à convertir l'image originale vers l'espace de couleur choisi, puis choisir un modèle de détection de la couleur de la peau, ensuite utiliser les règles (seuils) du modèle choisi pour discriminer entre les pixels de peau et celles de non peau.

2.2.2.1 Modélisation de la couleur de la peau

Pour modéliser la couleur de la peau, les méthodes statistiques ont été appliquées sur un ensemble d'images d'apprentissage de plusieurs personnes d'ethnies différentes (Africains, Européens, etc.) dans des espaces de couleurs différents. Les modélisations existantes peuvent être regroupées en trois catégories : explicites, paramétriques et non paramétriques. Les approches paramétriques représentent la distribution de la peau sous forme d'un modèle gaussien [Yang 99a, Phung 05], ou d'un mélange de gaussiennes [Sidibe 07]. Les approches non paramétriques (classifieur de Bayes ou table normale de consultation) utilisent les histogrammes des données d'apprentissage pour représenter la distribution de la couleur de la peau [Joenes 99]. Une des méthodes de classification des pixels peau la plus couramment utilisée et la plus simple consiste à définir explicitement les limites (frontières de décision) de la classe peau sur chaque canal de l'espace de couleur utilisé. Dans sa forme la plus élémentaire, cette frontière peut se définir comme une sous-région dans l'espace de représentation des couleurs. En pratique, cela se traduit par l'utilisation des intervalles (seuils maximum et minimum) sur chaque axe de cet espace. Ces seuils sont choisis de manière empirique dans la phase d'apprentissages en utilisant un ensemble de pixels de peau [Yang 99a, Phung 05, Singh 03]. La figure 2.2 ci-dessous montre un exemple de détection de la peau d'une image couleur et son masque binaire après la détection des régions de la peau.

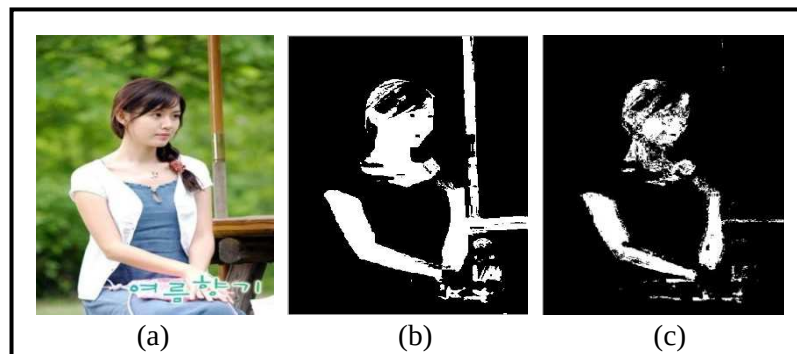


Figure 2.2 : Exemple de détection de la peau. (a) image originale, (b) image binaire montrant la peau détectée dans l'espace YCbCr en utilisant le modèle gaussien et (c) image binaire montrant la peau détectée dans l'espace YCbCr en utilisant le modèle non-paramétrique.

Notre objectif consiste à détecter les personnes afin de pré-localiser et de restreindre les régions d'intérêts. Pour cela, nous avons adopté dans ce travail la méthode explicite afin de détecter les régions de la peau dans les images pour sa simplicité et sa performance. En effet, cette méthode a l'avantage de pouvoir segmenter la peau des personnes en analysant les couleurs de la scène dans un espace donné (RGB normalisé, YCbCr, HSV). Ainsi elle ne

nécessite aucun apprentissage préalable. Celle-ci convient bien pour notre travail, avec un meilleur taux de précision.

2.2.2.2 Opérations morphologiques

Les opérations morphologiques jouent un rôle primordial pour la détection de la couleur de la peau, ces opérations permettent de modifier la valeur des pixels en fonction de ses voisins en utilisant un élément structurant (un ensemble de géométrie connue, composé d'un pixel central et de ses voisins). Ce qui permet de modifier la structure et la forme d'une région. Les opérations morphologiques de base sont l'érosion qui supprime les éléments de taille plus petite que la taille de l'élément structurant et la dilatation qui bouche les trous de taille plus petite que la taille de l'élément structurant [Boomgard 92].

— **Érosion** : L'érodé Y d'un ensemble X par un élément structurant A est l'ensemble des points x de X tel que $A_x(A$ centré sur x) est complètement inclus dans X [Boomgard 92].

$$Y = \{x \in X / A_x \subset X\} \quad (2.9)$$

Pour appliquer l'érosion sur une région, un élément structurant est centré sur chaque pixel et s'il est contenu entièrement dans cette région le pixel sera blanc sinon il sera noir.

— **Dilatation** : Le dilaté Z de X par l'élément structurant A est l'ensemble des points x de X tels qu'au moins un des points de $A_x(A$ centré sur x) soit en contact avec X [Boomgard 92].

$$Z = \{x \in X / X \cap A_x \neq \emptyset\} \quad (2.10)$$

Pour appliquer la dilatation sur une région, un élément structurant est centré sur chaque pixel et s'il a avec la région au moins un pixel en commun alors le pixel sera blanc sinon il sera noir.

2.2.2.3 Détermination des régions de la couleur de la peau

Plusieurs études ont montré que la variabilité de la couleur de la peau tenait plus de la différence d'intensité plutôt que de la chrominance [Phung 05]. Le choix de l'espace de couleur est crucial pour la détection de la peau, ainsi que les règles et les seuils de décision empiriques. Cependant les espaces de couleurs les plus convenables pour la segmentation des régions de la peau sont ceux qui séparent la luminance de la chrominance [Phung 05]. Dans ce travail, nous avons utilisé les trois espaces de couleur RGB normalisé, HSV et YCbCr :

— **Espace RGB normalisé** : Cet espace combine trois composantes de couleurs (Rouge, Vert et Bleu) et ne nécessite aucun modèle de peau et aucune transformation de couleurs [Kovac 03, Vezhnevets 03]. En effet, l'espace RGB n'est pas convenable pour la détection de

la peau, car il ne sépare pas la luminance de la chrominance, et les composantes R, G et B sont corrélées. Dans l'espace RGB normalisé la somme des trois éléments normalisés est connu ($r+g+b=1$), la troisième composante b ne contient plus d'information supplémentaire et peut-être omise, cela permet ainsi de réduire la dimension de l'espace.

— **Espace YCbCr** : Dans cet espace la couleur est représentée par la luminance Y et deux composantes de chrominances Cb et Cr (Bleu et Rouge). Cet espace sépare la luminance et la chrominance ce qui rend l'espace YCbCr attractif pour la modélisation de la couleur de la peau [Vezhnevets 03].

— **Espace HSV** (Hue, Saturation, value): Cet espace est basé sur l'espace de couleur RGB et il décrit la couleur avec des valeurs intuitives basé sur la teinte (H), la saturation (S) et la valeur d'intensité (V). La teinte définit la couleur dominante dans une région. La saturation mesure la coloration d'une région en proportion de sa brillance. Le point fort de l'utilisation de cet espace réside dans la discrimination entre la luminance et la chrominance [Vezhnevets 03, Tsekeridou 98].

Le tableau 2.1 ci-dessous présente les règles qui déterminent les régions de la peau dans ces trois espaces de couleurs. Ces règles ont été défini à partir d'un ensemble d'images de la base de données brésiliennes FEI [FEI 14] contenant 2800 images de 200 personnes prises dans des conditions différentes (éclairage, pose, etc.).

Espaces	Règles	
rgb normalisé	A	$0.45 \leq r \leq 0.7$
	B	$0.23 \leq g \leq 0.55$
YCbCr	C	$77 \leq Cb \leq 127$
	D	$133 \leq Cr \leq 173$
HSV	E	$0 \leq H \leq 0.5$
	F	$0.17 \leq S \leq 0.63$

Tableau 2.1 : Règles de segmentation des régions de la peau.

Dans ce travail, nous avons combiné les trois espaces de couleur RGB normalisé, HSV et YCbCr. La segmentation des images en régions de peau et de non peau est effectuée à l'aide d'un seuil calculé par la combinaison des règles proposées dans le tableau 2.1 ci-après. Ces règles sont combinées de la façon suivante :

$$((A)AND (B)) OR ((C)AND(D))OR((E)AND(F)) \quad (2.11)$$



Figure 2.3 : Exemple de détection de la peau avec la méthode explicite. (a) image originale, (b) pixels de la peau dans l'espace rgb normalisé, (c) pixels de la peau dans l'espace YCbCr, (d) pixels de la peau dans l'espace HSV, (e) pixels de la peau dans l'espace combiné rgb normalisé-HSV-YCbCr.

La figure 2.3 ci-après montre un exemple de détection de la peau dans les espaces de couleurs RGB normalisé, YCbCr, HSV et la combinaison de ces trois espaces de couleurs.

L'algorithme 2.1 ci-après présente les étapes de la segmentation des régions de la peau à l'aide des règles définies dans le tableau 2.1 ci-après.

Algorithme 2.1 : Segmentation de la couleur de la peau

Entrée : Image en couleur I

Règles de segmentation définie dans le tableau 2.1.

Sortie : Image binaire Ib

Début

Lire (I);

(N,M)=dimension(I);

Pour i=1 à N **faire**

Pour j=1 à M **faire**

Si (((A) And (B)) OR ((C) And (D)) OR ((E) And (F))) **alors**

 Ib(i,j)=1;

Sinon

 Ib(i,j)=0;

Fin si;

Fin pour;

Fin pour;

Fin.

2.2.3 Détection et appariement des points d'intérêt

L'extraction des points d'intérêt est une étape préliminaire de plusieurs processus dans le domaine de la vision par ordinateur. Les points d'intérêt ont des caractéristiques utiles pour de nombreuses applications, en particulier pour la détection des personnes, l'analyse du mouvement, le suivi d'objets ou encore la reconnaissance d'objets. Une qualité souhaitable d'un point d'intérêt est son invariance aux changements d'échelle et d'illumination aussi que de position de la caméra. Les détecteurs des points d'intérêt sont utilisés pour les trouver dans les images. Le premier détecteur que nous trouvons dans la littérature est celui de Moravec [Moravec 77] qui présente des problèmes sur quelques images, par la suite nous trouvons plusieurs détecteurs qui apportent plus d'améliorations et parmi ces détecteurs nous citons les détecteurs de Beaudet [Beaudet 78], SUSAN [Smith 97a], SIFT [Lowe 04], Harris [Harris 88] et KLT [Shi 94]. Pour une évaluation comparative de détecteurs des points d'intérêt, nous trouvons celles présentées dans [Mikolajczyk 05a, Schmid 98c].

Le détecteur de Moravec [Moravec 77] est basé sur une fonction d'auto-corrélation du signal. Il mesure les différences des niveaux de gris entre une fenêtre fixe et des fenêtres déplacées dans plusieurs directions. Quatre déplacements discrets dans des directions parallèles aux lignes et aux colonnes de l'image sont utilisés. Si le minimum de ces quatre différences est supérieur à un seuil, un point d'intérêt est détecté. Harris et Stephens ont amélioré l'approche de Moravec [Moravec 77] par l'utilisation de la matrice d'auto-corrélation. Ils ont déduit un détecteur de coins très populaire, c'est le détecteur de Harris [Harris 88]. L'utilisation des directions et des changements discrets sont ainsi évités. Au lieu d'utiliser une simple somme, une gaussienne est exploitée pour pondérer les dérivées à l'intérieur de la fenêtre. Les points d'intérêt sont détectés si la matrice d'auto-corrélation a deux valeurs propres élevées (proches et supérieures à un seuil) (Annexe A). Un détecteur très populaire SIFT est proposé par David Lowe [Lowe 04]. Il se base sur deux étapes : la détection des extrema dans l'espace d'échelles et la localisation des points d'intérêt. La première étape fait la convolution de l'image avec un noyau gaussien et la seconde permet de localiser les extrémums des images dans l'espace d'échelle : chaque point de l'image est comparé avec ses 26 voisins. Enfin, un test est réalisé sur les points détectés pour éliminer ceux qui ont un faible contraste et ceux qui sont proches d'un contour.

Les résultats obtenus après une étude comparative effectuée par Schmid [Schmid 98c], démontrent que le détecteur de Harris donne de meilleurs résultats, comparé au détecteur de Susan. Ceci concerne les transformations de l'image liées à la rotation, aux changements

d'échelle, aux points de vue et à la luminosité, ainsi qu'au bruit lié au capteur. En effet, les points d'intérêt couleurs sont des points présentant des changements significatifs dans le temps et dans l'espace et ils concentrent l'information initialement contenue dans toute l'image en quelques points spécifiques. Ainsi, ils sont aussi présents dans une grande majorité d'images par rapport aux contours. C'est pour cela nous avons utilisé dans ce travail les points d'intérêt couleurs afin de représenter et suivre les personnes dans une séquence d'images. Nous avons adopté l'algorithme de Harris couleur [Montesinos 98, Harris 88] détaillé dans l'annexe A (section A.1) pour détecter les points d'intérêt couleurs dans les régions de la peau. Après la détection de ces points, on fait appel à la mesure de corrélation *ZNCC* (Zero mean Normalized Cross Correlation) [Chambon 11, Stefano 05] détaillée dans l'annexe A (section A.2) afin de suivre les personnes dans les séquences des images.

2.3 Approche proposée

Dans le but d'améliorer et de développer les méthodes de suivi des personnes dans une séquence d'images, nous avons proposé une contribution dans ce domaine basée sur les points d'intérêt couleurs. La contribution est intitulée "Système de vidéosurveillance pour le suivi des personnes en mouvement en utilisant les points d'intérêt couleurs" [Laaroussi 14d, Laaroussi 14e]. Les principales étapes de notre approche pour le suivi des personnes basée sur les points d'intérêt couleurs sont présentées sur la figure 2.4 ci-après.

Notre approche est composée de deux étapes principales. Dans la première étape, nous avons effectué quelques prétraitements adaptables à notre application. Tout d'abord, nous avons segmenté l'arrière-plan et le premier-plan par le modèle de mélange gaussien GMM (Gaussian Mixture Model) [Stauffer 99] détaillé dans la section 2.2.1.2 afin d'extraire les personnes présentes dans la séquence d'images. Ensuite, nous avons déterminé des règles qui permettent de définir les régions de la peau par l'utilisation d'un ensemble d'images d'apprentissage dans les trois espaces de couleurs (RGB normalisée, YCbCr et HSV). Nous avons segmenté ces régions à l'aide de ces règles afin de simplifier les traitements ultérieurs en localisant les régions d'intérêt dans chaque image (section 2.2.2.3). Enfin, nous avons appliqué une combinaison d'opérations morphologiques sur les régions de la peau extraites pour boucher les trous et éliminer les petits segments (section 2.2.2.2). Dans la deuxième étape, nous détectons les points d'intérêt couleurs sur les régions de la peau trouvées dans l'étape précédente en utilisant l'algorithme de Harris couleur [Harris 88, Montesinos 98] détaillé dans l'annexe A (section A.1) dont le but de réduire le maximum

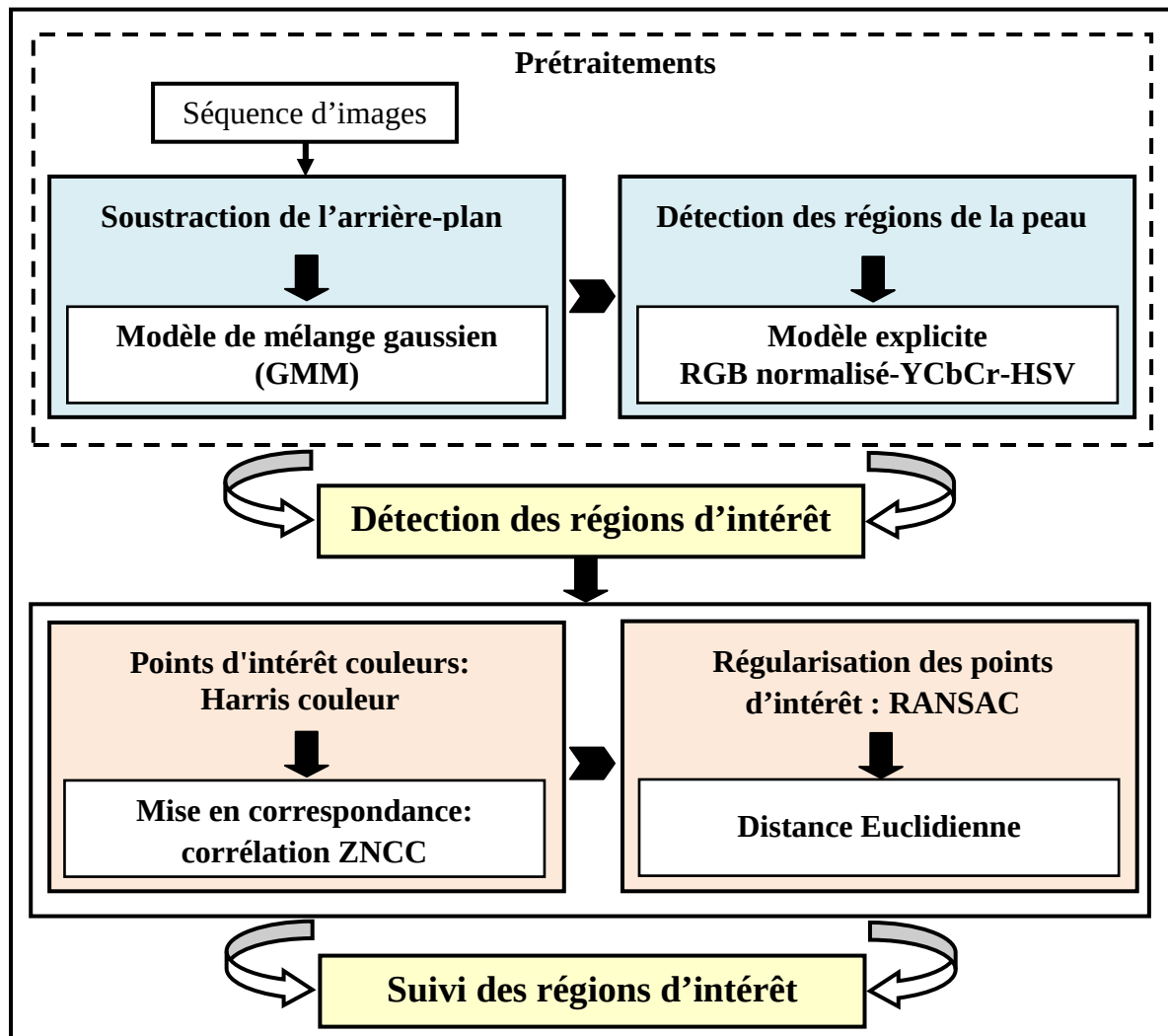


Figure 2.4 : Étapes de l'approche proposée pour le suivi des personnes.

d'informations d'une image en quelques points. L'utilisation de l'ensemble des points d'intérêt couleurs nous permet de suivre les régions de la peau qui représentent les personnes suivies, en faisant appel à la mesure de corrélation *ZNCC* (Zero mean Normalised Cross Correlation) détaillée dans l'annexe A (section A.2) [Stefano 05, Chambon 11]. En effet, lors de cette étape, plusieurs points sont des faux appariements. Ceci a pour effet d'altérer complètement le processus de suivi, d'où une étape de régularisation est très importante. Dans ce travail, nous avons utilisé la fonction *RANSAC* [Fischler 81] détaillée dans l'annexe A, où les meilleurs appariements sont identifiés en considérant le coefficient de corrélation maximum. Finalement, en calculant la distance euclidienne entre les meilleurs appariements et les autres points d'intérêt détectés dans chaque image, nous pouvons observer le mouvement des régions d'intérêt qui représentent les personnes suivies dans la scène.

Sachant que, la distance euclidienne entre deux points X et Y est définie comme étant la longueur du segment de droite qui les relie. Considérant deux points $M(x, y)$ et $N(u, v)$, la distance euclidienne entre ces deux points est définie par [Shih 04] :

$$d_e(M, N) = \sqrt{(x - u)^2 + (y - v)^2} \quad (2.12)$$

À noter que, les espaces de couleur YCbCr et HSV ont été utilisés seulement pour la segmentation des régions de la peau. Pour les autres étapes comme la soustraction de l'arrière-plan, la détection des points d'intérêt couleurs, l'appariement et la régularisation de ces points d'intérêt, nous avons utilisé l'espace de couleur RGB.

L'algorithme 2.2 ci-dessous présente les phases essentielles de notre approche pour le suivi des personnes basé sur les points d'intérêt couleurs:

Algorithme 2.2: Suivi des personnes basé sur les points d'intérêt couleurs.

Début

1. Lecture des images successives ;
2. Appliquer le modèle de mélange gaussien GMM pour segmenter l'arrière-plan et le premier-plan afin de détecter les personnes présentes dans les images ;
3. Appliquer le détecteur de peau par seuillage dans l'espace RGB normalisé-YCbCr-HSV pour segmenter les régions de la peau ;
4. Appliquer les opérations morphologiques dilatation et érosion avec un élément structurant sur les régions de la peau détectées dans l'étape 3 ;
5. Appliquer le détecteur de Harris couleur sur les régions d'intérêt trouvées dans les étapes précédentes ;
6. Apparier les points d'intérêt détectés en utilisant la mesure de corrélation $ZNCC$;
7. Régularisation des points d'intérêt détectés par la fonction $RANSAC$;
8. Calculer la distance euclidienne entre les meilleurs appariements et les autres points d'intérêt dans chaque image ;
9. Résultats du suivi où les régions d'intérêt qui représentent les personnes suivies sont entourées par des rectangles.

Fin

L'analyse de la complexité de cet algorithme consiste en l'étude formelle de la quantité des ressources nécessaires à l'exécution de cet algorithme. En effet, la complexité de cet

algorithme est très faible, car elle nécessite des opérations très simples comme la segmentation de la couleur de la peau, la détection des points d'intérêt couleur et la mise en correspondance de ces points. La complexité linéaire du temps de cet algorithme est donnée par $O(n)$ où n est le nombre de points d'intérêt couleurs utilisé.

2.4 Résultats expérimentaux

Dans cette section, nous allons présenter les simulations et les expérimentations de notre approche. Cette section est composée de deux parties principales. Dans la première, nous avons étudié des propriétés qui sont très importantes dans le cadre de l'analyse des vidéos, telles que la robustesse des points d'intérêt couleurs par rapport au bruit gaussien et au changement du contraste. Dans la deuxième partie, nous présentons les résultats expérimentaux sur différentes séquences d'images utilisées dans cette étude. Après nous allons faire une comparaison entre notre méthode et d'autres approches existantes dans la littérature.

2.4.1 Données synthétiques

Pour tester les performances de l'approche proposée, deux situations ont été examinées à savoir la modification du contraste et l'ajout d'un bruit gaussien. Ces deux opérations perturbent le plus la détection des points d'intérêt couleurs. Dans cette simulation, nous avons pris une image de taille 320×240 pixels prise à partir d'une vidéo d'un match de football et cette image est simulée pour tester la performance de l'approche proposée. Un bruit gaussien de déviation standard (écart type) $0 \leq \sigma \leq 2$ est ajouté aux pixels de l'image lors de l'étape de détection des points d'intérêt. Ces points d'intérêt sont détectés par l'algorithme de Harris couleur [Montesinos 98, Harris 88] (Annexe A). Le tableau 2.2 ci-après présente le nombre des points d'intérêt couleurs détectés en changeant le contraste et le bruit. L'évaluation est réalisée en observant les variations du nombre des points d'intérêt détectés par rapport à la situation initiale sans modification du contraste et sans bruit. Nous remarquons que la détection des points d'intérêt est très sensible à la modification du contraste lorsque celui-ci est très important comme indique le tableau 2.2 ci-après et la figure 2.5 (e) et (f) ci-après. Cependant, le nombre des points d'intérêt est relativement stable par rapport au changement du bruit gaussien lorsque celui-ci n'est pas très important (tableau 2.2 et figure 2.5 (c)). En effet, l'impact du bruit et du contraste sur les algorithmes de suivi ne peut pas être ignoré et la plupart des algorithmes de suivi peuvent être complètement perdus lors de suivi d'objets. À partir des simulations effectuées (le calcul du nombre de points d'intérêt couleurs détectés en

Contraste [Min Max]	Nombre de points	Bruit gaussien	Nombre de points
[0.99 1]	1	$\sigma = 0$ référence	52
[0.8 1]	7	$\sigma = 0.001$	62
[0.4 1]	37	$\sigma = 0.01$	954
[0 1] référence	52	$\sigma = 0.1$	1006
[0 0.5]	55	$\sigma = 1$	1071
[0 0.2]	64	$\sigma = 2$	1096

Tableau 2.2 : Influence du contraste et du bruit sur la détection de points d'intérêt couleurs.

changeant le contraste et le bruit), nous pouvons conclure que notre méthode donne des résultats satisfaisants lors de l'ajout du bruit et du contraste.

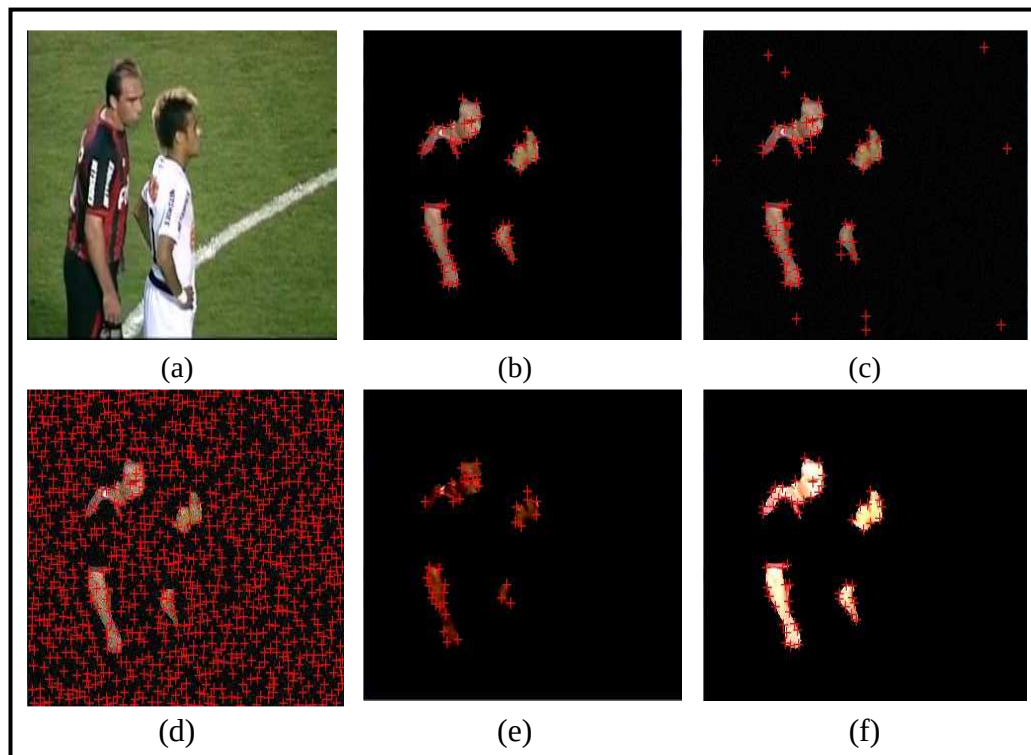


Figure 2.5 : Influence du bruit et du contraste sur la détection de points d'intérêt couleurs. (a) image originale, (b) résultat de détection des points d'intérêt couleurs sans modification du contraste et du bruit, (c) bruit gaussien avec $\sigma = 0.001$, (d) bruit gaussien avec $\sigma = 0.01$, (e) détection avec faible contraste et (f) détection avec fort contraste.

2.4.2 Données réelles

Pour caractériser les performances de la méthode proposée, nous avons utilisé deux séquences d'images avec des scénarios différents. Ces deux séquences contiennent des personnes qui se déplacent dans une scène avec un arrière-plan statique :

- **Classe:** Séquence de 245 images avec une résolution de 450×360 pixels présent à partir d'une vidéo en classe qui dure 25 minutes et 40 secondes, l'objet suivi est une personne qui se déplace librement devant la caméra.
- **Sport:** Séquence de 207 images avec une résolution de 420×356 pixels présent à partir d'une vidéo d'un match de football qui dure une minute et 5 secondes, les objets suivis sont deux personnes qui se croisent et marchent cote à cote.

Pour les deux séquences d'images, tout d'abord nous avons utilisé la soustraction de l'arrière-plan par le modèle de mélange gaussien GMM [Stauffer 99] afin de séparer les personnes en mouvement du premier-plan et les objets statiques de l'arrière-plan. Après cela, nous avons détecté les régions de la peau dans l'espace de couleur RGB normalisé-YCbCr-HSV. Une fois la classification des pixels de la peau effectuée, nous appliquons les deux opérations morphologiques l'érosion suivie par la dilatation pour enlever les pixels dispersés représentant des fausses classifications et renforcer les régions d'intérêt détectées. Après la détection de toutes les régions de la peau pour chaque image de la séquence, les plus petites sont éliminées. Dans notre application, une taille minimale des régions est de 400 pixels est bien convenable. Les points d'intérêt couleurs sont détectés par l'algorithme de Harris couleur [Harris 88, Montesinos 98] et appariés dans chaque couple d'images par la fonction de corrélation *ZNCC* [Stefano 05, Chambon 11]. Le résultat obtenu contient des faux appariements, pour les éliminer, nous avons utilisé l'algorithme de *RANSAC* [Fischler 81]. Dans ce qui suit, nous présentons nos résultats d'expérimentations pour les deux séquences d'images où les scénarios sont variés allant de la simple personne qui marche seule à deux personnes qui se croisent et marchent cote à cote.

Dans la première expérience, nous avons utilisé la séquence Classe et la cible à suivre est la personne qui se déplace librement dans la classe avec une caméra statique. L'arrière-plan de cette séquence est relativement stable et les images sont prises à partir de différents points de vue à savoir devant, profil et derrière. Un exemple des résultats qualitatifs de cette séquence sont présentés sur la figure 2.5 ci-après. La première ligne (a) présente les images originales considérées pour cette séquence, la deuxième ligne (b) montre le premier-plan détecté par le modèle de mélange gaussien GMM [Stauffer 99] et la troisième ligne (c) présente les résultats de segmentation des régions de la peau après l'application de deux opérations morphologiques l'érosion suivie par la dilatation. La ligne (d) montre les points d'intérêt couleurs détectés par Harris [Harris 88, Montesinos 98] sur les régions de la peau, les points d'intérêt sont affichés en rouge et vert dans chaque couple d'images. La ligne (e)

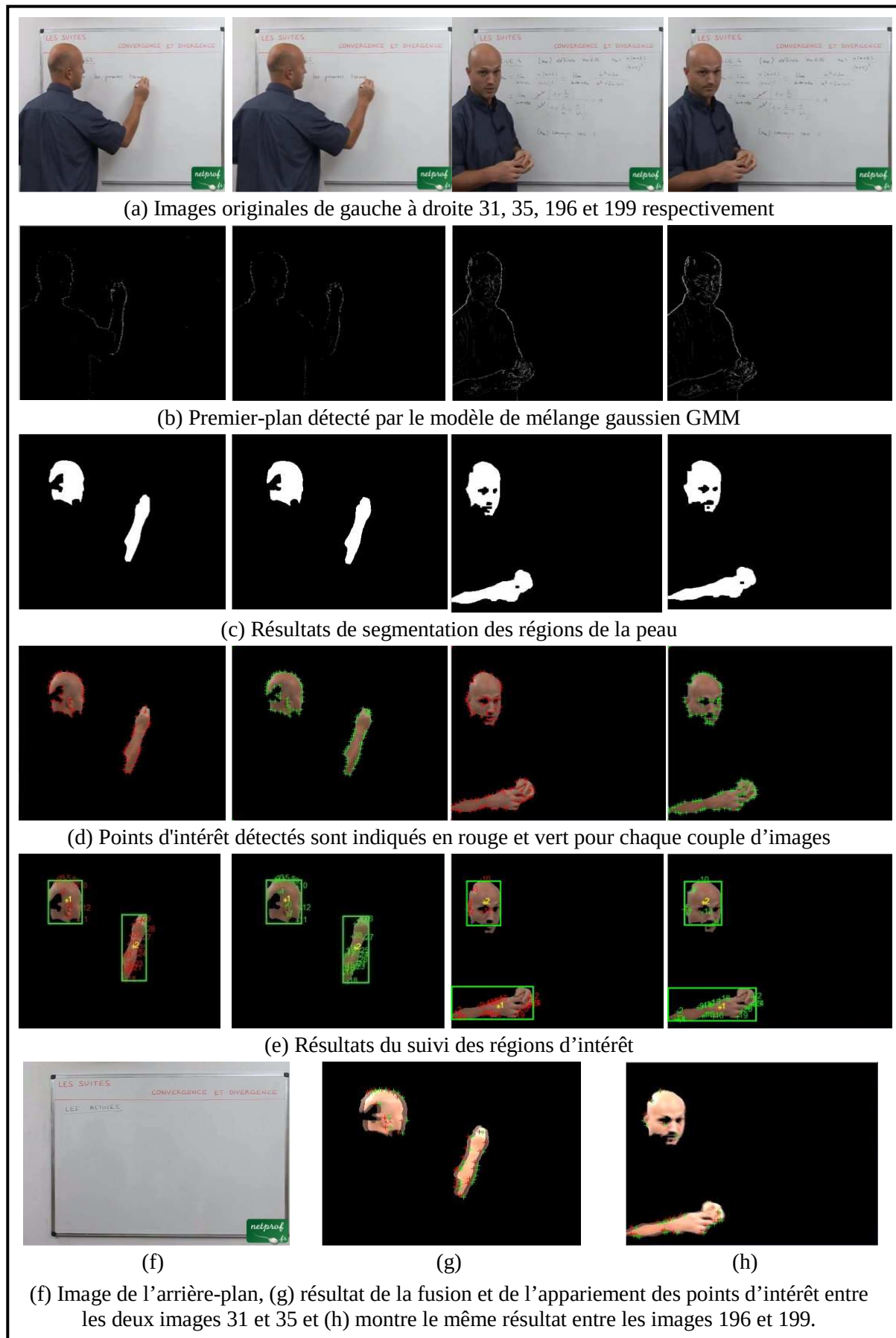


Figure 2.6 : Résultats du suivi pour la séquence Classe.

montre les résultats d'appariements par la mesure de corrélation *ZNCC* [Stefano 05, Chambon 11] et qui sont régularisés par la fonction *RANSAC* [Fischler 81] afin d'éliminer les faux appariements. Les appariements qui restent sont numérotés en rouge et vert dans chaque couple d'images. Cette étape montre aussi les résultats du suivi des régions d'intérêt qui représentent la personne suivie dans cette séquence où chaque région est entourée par un rectangle. La dernière ligne (f) montre l'arrière-plan identifié pour cette séquence et les résultats de la fusion pour chaque couple d'images.

Le vecteur de déplacement correspondant à ces images est donné par $V = (v_1, \dots, v_n)$, où n est le nombre d'appariements détecté dans chacune des paires d'images. Les distances euclidiennes traversées entre les meilleurs appariements et les autres points pour les images 31, 35, 196 et 199 respectivement sont $d_1=3.2463e+003$, $d_2=2.7077e+003$, $d_3=2.9711e+003$, $d_4=4.8297e+003$. À partir de ces résultats, nous pouvons observer le mouvement des régions d'intérêt qui représentent la personne suivie dans cette séquence. En résumé, les résultats qualitatifs présentés sur la figure 2.6 ci-avant montrent que notre méthode proposée arrive à détecter et à suivre correctement les régions d'intérêt qui représentent la personne suivie dans toute la séquence en utilisant la couleur de la peau et les points d'intérêt couleurs, malgré le changement d'illumination et de l'arrière-plan dans la fin de cette séquence comme montrent les images 196 et 199 de la figure 2.5 (a) ci-avant.

La deuxième expérience est une séquence plus complexe "Sport" avec 207 images et les cibles à suivre sont les deux personnes qui interagissent entre eux et se déplacent l'une vers l'autre. Les difficultés de cette séquence sont la capacité à prendre en compte le mouvement rapide des personnes, le changement d'illumination et l'occultation. Certains résultats finaux du suivi des personnes pour cette séquence sont présentés sur la figure 2.7 ci-après. La première ligne (a) montre quelques images originales de cette séquence où la plupart sont filmées de profil. La deuxième ligne (b) montre le premier-plan détecté par le modèle de mélange gaussien GMM [Stauffer 99] et la troisième ligne (c) présente les résultats de segmentation des régions de la peau après l'application des deux opérations morphologiques : l'érosion suivie par la dilatation. La ligne (d) présente les points d'intérêt couleurs détectés par Harris [Harris 88, Montesinos 98] dans les régions de la peau trouvées. Ces points sont affichés en rouge et vert dans chacune de deux images. La ligne (e) montre les résultats d'appariements par la mesure de corrélation *ZNCC* [Stefano 05, Chambon 11] et qui sont régularisés par *RANSAC* [Fischler 81] afin d'éliminer les faux appariements. Les appariements qui restent sont numérotés en rouge et vert dans chaque couple d'images.

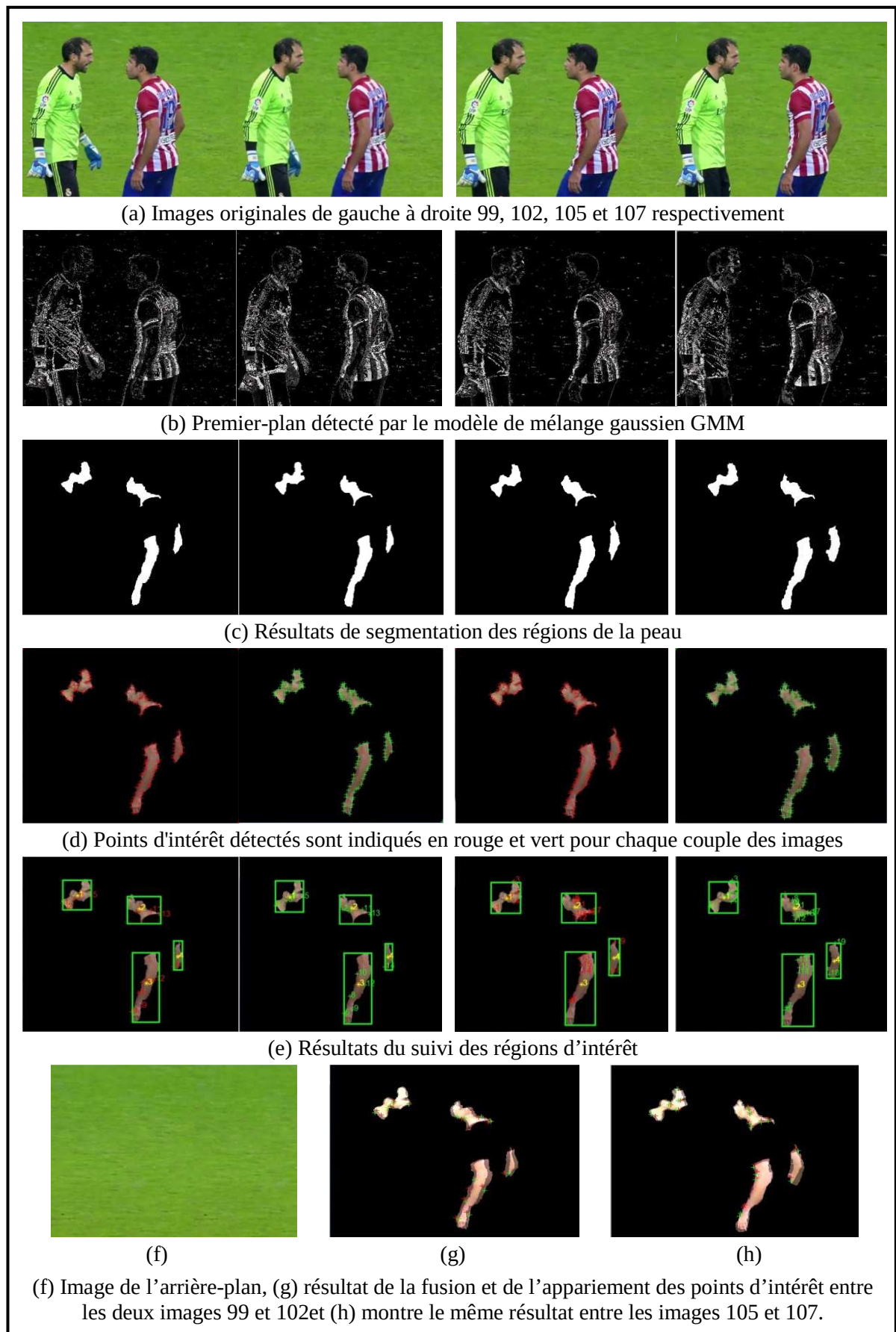


Figure 2.7 : Résultats du suivi pour la séquence Sport.

Cette étape montre aussi les résultats du suivi des régions d'intérêt qui représentent les personnes suivies dans cette séquence où chaque région est entourée par un rectangle. La dernière ligne (f) présente l'arrière-plan identifié pour cette séquence et les résultats de la fusion pour chaque couple d'images. Le vecteur de déplacement correspondant à ces images est donné par $U = (u_1, \dots, u_m)$, où m est le nombre d'appariements détecté dans chacune des paires d'images. Les distances euclidiennes traversées pour les images 99, 102, 105 et 107 sont respectivement $d_1=2.1019e+003$, $d_2=1.9172e+003$, $d_3=2.5233e+003$, $d_4=2.5925e+003$. À partir de ces résultats, nous pouvons observer le mouvement des régions d'intérêt qui représentent les deux personnes suivies dans cette séquence.

Les résultats qualitatifs présentés sur la figure 2.7 ci-avant montrent que les deux personnes sont correctement détectées et suivies dans toute la séquence. Les mains et les visages sont bien détectés et suivis dans la majorité d'images de la séquence même en présence des changements d'illumination et les mouvements des personnes suivies. Les résultats obtenus, même s'ils ne sont pas parfaits, sont vraiment très encourageants compte tenu de la complexité de la séquence. Ainsi que, notre méthode s'adapte bien avec les conditions du bruit et avec un faible contraste. En résumé, par la combinaison des points d'intérêt couleurs et la couleur de la peau nous avons développé une méthode de suivi des personnes qui prend en considération la qualité de suivi.

Pour évaluer la performance de notre approche par rapport aux autres méthodes de la littérature, nous avons comparé nos résultats avec les résultats obtenus par deux méthodes ([Lim 11] et [Gouet 04b]) bien établies et qui donnent de bons résultats. La première méthode utilisée est celle de Lim et al. [Lim 11]. Cette approche est basée sur l'utilisation des caractéristiques de la couleur des points d'intérêt et d'une fenêtre de recherche variable VSW (Variable Search Window) selon la taille des personnes suivies. La seconde méthode est basée sur les points d'intérêt et les contours [Gouet 04b] mais avec des images en niveau de gris et pour un objet unique dans la scène.

Pour estimer la précision de suivi nous avons utilisé la formule suivante [Lim 11] :

$$\text{Précision}(\%) = \frac{NT - NF}{NT} \times 100 \quad (2.13)$$

avec NT est le nombre total des images de la séquence et NF est le nombre d'images pour un faux suivi. Le tableau 2.3 ci-après présente la précision de suivi en pourcentage calculée par la formule (2.13) ci-dessus et le temps moyen d'exécution en seconde par image pour chaque méthode considérée.

	Notre méthode	Méthode1 [Lim 11]	Méthode2 [Gouet 04b]
Précision de suivi (%)	98.86	97.17	93.27
Temps moyen d'exécution (s/image)	0.15	0.14	0.80

Tableau 2.3 : Précision de suivi (%) et le temps moyen d'exécution (s/image) pour chaque méthode considérée.

D'après les résultats quantitatifs présentés dans le tableau 2.3 ci-dessus, nous pouvons remarquer que notre méthode a obtenu une précision plus élevée avec une moyenne de précision de 98.86%, suivie par la méthode de [Lim 11] avec une moyenne de précision de 97.17% et celle de [Gouet 04b] avec une moyenne de précision de 93.27%. Pour le temps moyen d'exécution notre méthode donne de bons résultats par rapport aux résultats obtenus par [Gouet 04b] et elle donne des résultats similaires à ceux obtenus par [Lim 11] selon le tableau 2.3 ci-dessus.

D'après cette étude comparative, nous avons tiré diverses conclusions. Tout d'abord, nous pouvons constater que les méthodes qui travaillent avec des vidéos en niveaux de gris sont moins performantes que celles travaillant avec des vidéos en couleurs comme le cas de la méthode [Gouet 04b]. Cependant, les limitations de la méthode de [Lim 11] sont : sa sensibilité aux objets ayant des couleurs non spécifiques, aux changements d'illumination et à la présence du bruit.

2.4.3 Discussion

La méthode que nous avons proposée pour le suivi des personnes basée sur les points d'intérêt couleurs a les avantages suivants :

- ✓ L'utilisation d'un ensemble de points d'intérêt couleurs détectés sur les régions de la peau pour le suivi des personnes nous permet de réduire le maximum d'information d'une image en quelques points ;
- ✓ La précision de suivi est plus élevée par rapport aux autres méthodes basées sur les points d'intérêt. En effet le taux de détection dans ces images est toujours supérieur à 98% et le taux de fausse détection est très faible dans toutes ces images. Ceci est dû à l'étape qui consiste à vérifier l'existence des personnes dans les régions ayant la couleur de la peau ;
- ✓ Résultats satisfaisants: différents points de vue (devant, profil et derrière), arrière-plans statiques, bruits, changements d'illumination et d'échelle ;

✓ Applicable sur les images couleurs ;

Néanmoins, notre méthode a quelques inconvénients :

✓ Non valide dans les cas d'occultation complexe et du mouvement de la caméra ;

✓ Il faut au moins une partie de la couleur de la peau reste visible.

2.5 Conclusion

Dans ce chapitre, nous avons présenté une nouvelle méthode pour le suivi des personnes dans une séquence d'images basée sur le modèle de mélange gaussien GMM, la couleur de la peau, et les points d'intérêt couleurs. Nous avons détaillé ses différentes étapes, puis nous avons présenté les résultats obtenus sur un ensemble d'images prises sur Internet. Nous avons aussi, comparé ces résultats avec ceux des autres méthodes récemment publiées. Les résultats des tests et de comparaison sont très satisfaisants. Les différentes expérimentations effectuées sur des données synthétiques et réelles montrent, d'une part, la qualité de notre travail et d'autre part, les bons résultats obtenus. Les résultats sont vraiment encourageants surtout que notre méthode s'adapte bien avec le changement d'illumination, la présence d'occultation partielle et les conditions du bruit. En effet, la combinaison de ces deux caractéristiques : la couleur de la peau et les points d'intérêt couleurs permettra de concevoir une méthode simple et robuste pour le suivi des personnes dans une séquence d'images. Le point faible de cette contribution est l'utilisation d'une caméra statique, qui n'est pas toujours présente dans la nature. Pour cette raison, nous avons proposé d'autres contributions qui s'adaptent avec le mouvement de la caméra. Ces contributions seront présentées dans les chapitres suivants.

Chapitre 3

Nouvelles méthodes de suivi basées sur les histogrammes

3.1 Introduction

Le modèle d'apparence d'un objet est un élément essentiel qui joue un rôle très important pour la performance d'un algorithme de suivi d'objets. La notion de modèle d'apparence est précisée sur la base de l'extraction de descripteurs visuels comparés à l'aide de similarités à une référence. Cependant, les techniques d'estimation fondées sur les densités à noyau ont été proposées dans les travaux de [Comaniciu 03c, Georgescu 04] pour construire des représentations statistiques de l'apparence de l'objet. Leur avantage est de proposer une description relativement flexible et générique de l'apparence et elles permettent une mise en œuvre rapide et en temps-réel dans certains cas. Tandis que, des progrès significatifs ont été réalisés dans la conception d'un modèle d'apparence sophistiqué avec une représentation efficace de la cible. En effet, le suivi d'objets reste un problème difficile, notamment lorsque la seule information fournie à l'algorithme de suivi est la région de l'objet cible annotée manuellement sur la première image, le plus souvent délimitée par une forme géométrique. Cette limitation accroît les difficultés de suivi pour de nombreuses raisons principales à savoir le manque d'informations sur l'apparence de la cible vue qu'une seule vue est disponible, l'imprécision dans la distinction entre l'objet et son arrière-plan, car des pixels de ce dernier sont souvent inclus dans la boîte englobante ou la forme géométrique délimitant l'objet et le changement d'apparence de la cible provoqué par divers facteurs de perturbation (bruit, occultation, mouvement, éclairage, etc.). Bien que, trouver les caractéristiques les plus pertinentes pour représenter la cible à suivre reste un problème difficile. Cependant, de nombreux algorithmes pour le suivi d'objets ont été proposés dans la littérature [Yilmaz 06a, Li 13c]. Parmi eux, l'algorithme Mean Shift [Comaniciu 03c] qui se révèle être robuste et précis par rapport aux autres algorithmes [Yilmaz 06a, Li 13c, He 13] et il a attiré une attention considérable pour le suivi d'objets.

Dans le but d'améliorer et de développer les méthodes de suivi des personnes dans une séquence d'images, nous avons proposé trois contributions qui seront présentées dans ce

chapitre et qui sont basées sur l'algorithme Mean Shift [Comaniciu 03c]. La première contribution [Laaroussi 16c] utilise l'histogramme pondéré et corrigé de l'arrière-plan [Ning 10a] et les descripteurs de moments du poids de l'image [Bradski 98, Ning 12c] afin d'estimer les changements de la position, de l'échelle et de l'orientation de la cible. La deuxième méthode [Laaroussi 16a] s'appuie sur les caractéristiques de couleur-texture et l'histogramme pondéré de l'arrière-plan pour une représentation plus distinctive et efficace de la cible. Le point faible de cette contribution est l'utilisation des caractéristiques moins robustes par rapport à la similarité d'apparence de la cible et son arrière-plan, ainsi qu'à la présence d'éléments distracteurs. Pour cette raison, nous avons proposé une troisième contribution [Laaroussi 17b] qui utilise la jointure des caractéristiques couleur-texture qui est une partie commune avec la précédente et l'histogramme pondéré du premier-plan afin de sélectionner uniquement les caractéristiques importantes de la cible.

Le reste de ce chapitre est organisé comme suit : la section 3.2 présente la première contribution, la section 3.3 expose la deuxième contribution et la section 3.4 examine la troisième contribution.

3.2 Méthode de suivi basée sur l'histogramme pondéré et corrigé de l'arrière-plan

Dans cette section, nous présentons une approche pour le suivi des personnes dans une séquence d'images intitulée « Suivi d'une personne basé sur l'histogramme pondéré de l'arrière-plan en estimant l'échelle et l'orientation (Scale and Orientation-based Background Weighted Histogram for human tracking SOBWH) » [Laaroussi 16c]. Cette méthode est basée sur l'histogramme pondéré et corrigé de l'arrière-plan [Ning 10a] et les descripteurs de moments du poids de l'image [Bradski 98, Ning 12c]. Ce travail vise à surmonter certaines difficultés de suivi liées aux changements d'échelle et d'orientation de la cible et réduire l'interférence des caractéristiques de l'arrière-plan dans la localisation de la cible où la seule connaissance disponible sur celle-ci est sa position sur la première image de la séquence vidéo. Deux axes sont ainsi développés dans cette approche. D'une part, les descripteurs de moments du poids de l'image [Bradski 98] ont été utilisés pour estimer la position, la taille et l'orientation de la cible (section 3.2.1). Dans ce contexte, nous soutenons que le poids de l'image calculé à partir du modèle de la cible et celui de la cible candidate représente la probabilité d'appartenance d'un pixel à la cible. De ce fait, la taille de l'objet est calculée par le moment de zéroth et le coefficient de Bhattacharyya [Kailath 67], ensuite, la largeur, la

hauteur et l'orientation ont été calculées par la région estimée de la cible et les moments d'ordre 2. D'autre part, l'histogramme pondéré et corrigé de l'arrière-plan CBWH (Corrected Background-Weighted Histogram) [Ning 10a] (section 3.2.2) a été incorporé dans la représentation de la cible afin de dériver une représentation simple des caractéristiques principales de l'arrière-plan et de réduire l'interférence de ces caractéristiques dans la localisation de la cible. Finalement, nous avons combiné ces deux concepts afin de générer une nouvelle approche pour le suivi des personnes dans une séquence d'images.

3.2.1 Estimation de l'échelle et de l'orientation

L'algorithme Mean Shift pour le suivi d'objets consiste à rechercher la position de l'objet d'intérêt dans l'image courante à partir de sa distribution de couleur qui est généralement approximée par un histogramme normalisé [Comaniciu 03c] (détaillé dans l'annexe B). En utilisant la procédure Mean Shift [Comaniciu 03c], la cible se déplace de la position courante $\hat{y}_0$ vers la nouvelle position $\hat{y}_1$ selon la relation suivante :

$$\hat{y}_1 = \frac{\sum_{i=1}^{n_h} x_i \omega_i}{\sum_{i=1}^{n_h} \omega_i} \quad (3.1)$$

avec $\{x_i\}_{i=1 \dots n_h}$ l'ensemble des coordonnées normalisées de n_h pixels de la cible candidate centré à y dans l'image courante, et les poids ω_i sont défini par :

$$\omega_i = \sum_{u=1}^m \sqrt{\frac{\hat{q}_u}{\hat{p}_u(\hat{y}_0)}} \beta[l(x_i) - u] \quad (3.2)$$

avec $\hat{q}_u$ est la probabilité de la caractéristique $u = 1 \dots m$ dans le modèle de la cible $\hat{q}$ calculée par l'équation (B.1) (Annexe B), $\hat{p}_u(\hat{y}_0)$ la probabilité de la caractéristique $u = 1 \dots m$ dans le modèle de la cible candidate $\hat{p}(\hat{y})$ calculée par l'équation (B.3) (Annexe B), et m est la taille de l'histogramme (c-à-d le nombre de classes utilisées). En utilisant l'équation (3.1) ci-dessus, l'algorithme Mean Shift cherche dans la nouvelle image la position de la région la plus semblable à la cible avec un nombre restreint d'itérations et la nouvelle position $\hat{y}_1$ est le centre de la région cible dans l'image courante.

En effet, à partir de l'équation (3.1), nous pouvons observer que les paramètres clés de l'algorithme Mean Shift sont les poids ω_i définis par l'équation (3.2). Dans la suite de cette section, nous analysons comment calculer les changements d'échelle et d'orientation de la cible en utilisant l'algorithme Mean Shift.

3.2.1.1 Estimation de la région cible

En se basant sur les analyses présentées dans les travaux de [Bradski 98, Mukundan 98, Ning 12c], nous pouvons considérer le poids de l'image dans l'algorithme Mean Shift [Comaniciu 03c] comme une fonction de densité de probabilité de la cible, où la valeur du poids d'un pixel représente la possibilité d'appartenance de ce pixel à la cible. Cependant, l'échelle et l'orientation de la cible peuvent être estimées en utilisant cette fonction de densité de probabilité conjointement avec les descripteurs de moments du poids de l'image. Par définition, la somme des poids de tous les pixels, c'est-à-dire le moment d'ordre zéroth, d'une image peut être considéré comme la zone pondérée de la cible [Ning 12c] :

$$M_{00} = \sum_{i=1}^n \omega(x_i) \quad (3.3)$$

Pour une estimation plus proche à la région de la cible, certains paramètres d'ajustement sont nécessaires afin de l'adapter aux changements d'échelle. Cependant, selon les travaux de Ning et al. [Ning 12c], pour garder la cible dans la région de recherche, ces auteurs augmentent la taille de la cible candidate par l'augmentation de la largeur et la longueur de la région de recherche basée sur les valeurs des images précédentes. D'autre part, ils ont utilisé le coefficient de Bhattacharyya ρ [Kailath 67] pour ajuster l'estimation de la région de la cible M_{00} . D'où la région de la cible est estimée par l'équation suivante :

$$A = e(\rho)M_{00} \quad (3.4)$$

avec $e(\rho)$ est une fonction gaussienne monotone et croissante et qui respecte le coefficient de Bhattacharyya ρ ($0 \leq \rho \leq 1$) [Kailath 67]. La fonction gaussienne $e(\rho)$ est définie par la formule suivante :

$$e(\rho) = \exp\left(\frac{\rho - 1}{\sigma}\right) \quad (3.5)$$

À partir de ces deux équations (3.4) et (3.5), nous pouvons constater que lorsque le modèle de la cible s'approche de la cible candidate, $e(\rho)$ se rapproche de 1 et dans ce cas il est plus fiable d'utiliser M_{00} comme estimation de la région cible. Lorsque ρ diminue, la similarité entre le modèle cible et la cible candidate diminue, M_{00} sera beaucoup plus grand que la région cible, mais $e(\rho)$ est inférieur à 1, de sorte que A peut éviter d'être trop biaisé de la région réelle de la cible. Notez que, lorsque ρ approche de 0, la cible suivie est perdue et $e(\rho)$ sera très faible pour que A soit proche de 0. Cependant, par expérimentation le paramètre σ est pris entre 1 et 2 [Ning 12c].

3.2.1.2 Estimation de la largeur, la hauteur et l'orientation de la cible

Les descripteurs de moments décrivent des quantités numériques à une certaine distance d'un point de référence ou d'un axe [Prokop 92]. Ils sont couramment utilisés dans l'analyse d'images [Mennesson 10], la reconnaissance des formes [Prokop 92] et le suivi d'objets [Bradski 98, Ning 12c]. Cela est dû essentiellement à leur simplicité, leur invariance et la signification géométrique des valeurs de moments d'ordre inférieur. Dans la suite de cette section, nous analysons les descripteurs de moments dans l'algorithme Mean Shift et ensuite les combiner avec la région estimée de la cible afin d'estimer davantage la largeur, la hauteur et l'orientation de la cible. Similaire à l'algorithme CAMSHIFT [Bradski 98], les moments d'ordre 1 du poids de l'image $\{M_{10}, M_{01}\}$, sont utilisés pour localiser le centre de la cible :

$$M_{10} = \sum_i^{n_h} w_i x_{i,1} \quad M_{01} = \sum_i^{n_h} w_i x_{i,2} \quad (3.6)$$

Également les moments d'ordre 2 du poids de l'image $\{M_{02}, M_{11}, M_{20}\}$ sont calculés par :

$$M_{20} = \sum_{i=1}^{n_h} w_i x_{i,1}^2 \quad M_{02} = \sum_{i=1}^{n_h} w_i x_{i,2}^2 \quad M_{11} = \sum_{i=1}^{n_h} w_i x_{i,1} x_{i,2} \quad (3.7)$$

avec le couple $(x_{i,1}, x_{i,2})$ représente les coordonnées du pixel i dans la région candidate. En comparant l'équation (3.1) avec les deux équations (3.3) et (3.6), la nouvelle position $\hat{y}_1$ de la cible trouvée par l'équation (3.1) est en fait le rapport entre le moment d'ordre 1 et le moment d'ordre 0 [Ning 12c] :

$$\hat{y}_1 = (\bar{x}_1, \bar{x}_2) = \left(\frac{M_{10}}{M_{00}}, \frac{M_{01}}{M_{00}} \right) \quad (3.8)$$

avec $(\bar{x}_1, \bar{x}_2)$ représente le centre de la région candidate de la cible. Selon [Bradski 98], le moment d'ordre 2 décrit la forme et l'orientation de la cible (grand axe). En utilisant les équations (3.1), (3.3), (3.7) et (3.8), nous pouvons convertir l'équation (3.1) vers le moment d'ordre 2 comme suit :

$$\begin{aligned} \mu_{20} &= M_{20}/M_{00} - \bar{x}_1^2 & \mu_{02} &= M_{02}/M_{00} - \bar{x}_2^2 \\ \mu_{11} &= M_{11}/M_{00} - \bar{x}_1 \bar{x}_2 \end{aligned} \quad (3.9)$$

L'équation (3.9) peut être réécrite par la matrice de covariance suivante :

$$Cov_1 = \begin{bmatrix} \mu_{20} & \mu_{11} \\ \mu_{11} & \mu_{02} \end{bmatrix} \quad (3.10)$$

En utilisant la région estimée de la cible A (section 3.2.1.1) et les descripteurs de moment présentés ci-dessus, la largeur, la hauteur et l'orientation de la cible peuvent être bien

estimées. La matrice de covariance dans l'équation (3.10) peut être décomposée en utilisant la décomposition en valeurs singulières (SVD) [Horn 91a] de la manière suivante :

$$Cov_1 = U \times S \times U^T = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix} \times \begin{bmatrix} \lambda_1^2 & 0 \\ 0 & \lambda_2^2 \end{bmatrix} \times \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix}^T \quad (3.11)$$

avec $U = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix}$ et $S = \begin{bmatrix} \lambda_1^2 & 0 \\ 0 & \lambda_2^2 \end{bmatrix}$. λ_1^2 et λ_2^2 sont les valeurs propres de la matrice de covariance [Horn 91a]. Les vecteurs $(u_{11}, u_{21})^T$ et $(u_{12}, u_{22})^T$ représentent l'orientation des deux axes principaux de la cible dans la région candidate. Dans l'algorithme CAMSHIFT [Bradski 98], les valeurs propres λ_1 et λ_2 sont directement utilisées en tant que largeur et hauteur de la cible [Mukundan 98]. Cependant, Ning et al. [Ning 12c] ont proposé un nouveau schéma pour estimer avec plus de précision la largeur et la hauteur de l'objet. La cible à suivre est représentée par une ellipse dont les longueurs de demi-grand axe et demi-petit axe sont désignées par a et b , respectivement. En effet, Mukundan et al. [Mukundan 98] ont montré que le rapport λ_1 et λ_2 peut bien approximer le rapport de a à b , c-à-d $\lambda_1/\lambda_2 \approx a/b$, par conséquent, les termes a et b peuvent être définis comme suit : $a = k\lambda_1$ et $b = k\lambda_2$, où k est un facteur d'échelle. En utilisant les moments géométriques de l'ellipse [Mukundan 98] et la région estimée de la cible A calculée par l'équation (3.4) ci-avant, nous pouvons écrire que $A = \pi ab = \pi(k\lambda_1)(k\lambda_2)$. Ensuite, nous pouvons facilement déduire que :

$$k = \sqrt{A/(\pi\lambda_1\lambda_2)} \quad (3.12)$$

$$a = \sqrt{\lambda_1 A/(\pi\lambda_2)} \quad b = \sqrt{\lambda_2 A/(\pi\lambda_1)} \quad (3.13)$$

Or, la matrice de covariance devient :

$$Cov_1 = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix} \times \begin{bmatrix} a^2 & 0 \\ 0 & b^2 \end{bmatrix} \times \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \end{bmatrix}^T \quad (3.14)$$

avec a^2 et b^2 sont les valeurs propres de cette matrice de covariance qui représentent respectivement la hauteur et la largeur de la région ellipsoïde suivie. D'après ces résultats, nous avons combiné la région estimée de la cible A avec la matrice de covariance pour estimer la largeur, la hauteur et l'orientation de la cible.

Après l'estimation de la position, l'échelle et l'orientation de la cible dans l'image courante, nous devons déterminer la taille de la cible candidate dans l'image suivante. Selon les travaux de [Ning 12c] et l'équation (3.14) ci-dessus, nous définissons la matrice de covariance suivante pour représenter la taille de la région de la cible candidate dans l'image suivante par :

$$Cov_2 = U \times \begin{bmatrix} (a + \Delta d)^2 & 0 \\ 0 & (b + \Delta d)^2 \end{bmatrix} \times U^T \quad (3.15)$$

avec Δd correspond à l'incrément de la région de la cible candidate dans l'image suivante.

3.2.2 Histogramme pondéré et corrigé de l'arrière-plan

L'information de l'arrière-plan joue un rôle très important puisqu'elle est souvent présentée dans la région sélectionnée de la cible, ainsi que cette information reste un concept critique pour le suivi d'objets. Cependant, les raisons pour lesquelles nous avons utilisé ce type d'information sont : D'une part, si certaines caractéristiques de la cible sont également présentées dans l'arrière-plan, leur pertinence pour la localisation de la cible sera diminuée. D'autre part, dans de nombreuses applications [Pan 16, Yilmaz 06a], il est difficile de délimiter exactement la région de la cible sans aucune information de l'arrière-plan, d'où pour une meilleure représentation de celle-ci, le modèle de l'arrière-plan a été utilisé pour améliorer la localisation de la cible.

Un modèle de représentation de l'arrière-plan nommé l'histogramme pondéré de l'arrière-plan BWH (Background-Weighted Histogram) a été proposé par Comaniciu et al. [Comaniciu 03c] afin de dériver une représentation simple des caractéristiques de l'arrière-plan et de ne sélectionner que les parties saillantes des représentations du modèle de la cible et des cibles candidates. Le modèle original de l'arrière-plan $\hat{b}$ est modélisé par une représentation discrète (histogramme normalisé) $\{\hat{b}_u\}_{u=1 \dots m}$ (avec $\sum_{u=1}^m \hat{b}_u = 1$) dans l'espace de caractéristiques comme suit :

$$\begin{cases} \hat{b} = \{\hat{b}_u\}_{u=1 \dots m} \\ \hat{b}_u = C_1 \sum_{i=1}^n k(\|x_i^*\|^2) \beta[l(x_i^*) - u] \end{cases} \quad (3.16)$$

avec $\hat{b}_u$ représente la probabilité de la caractéristique, $u = 1 \dots m$, dans ce modèle, m est la taille de l'histogramme, β est la fonction Kronecker où la somme de ces fonctions pour $u = 1 \dots m$ est égal à 1 [Comaniciu 03c]. La fonction $l : R^2 \rightarrow \{1, \dots, m\}$ associe à chaque pixel à la position x_i^* l'indice $l(x_i^*)$ de sa classe dans le m-histogramme. $k(x)$ est une fonction de profil convexe, monotone et décroissante, attribuant un poids plus faible aux coordonnées éloignées du centre de la région. Le rayon de cette fonction de profil est égal à 1. $\|x_i^*\|$ est la distance entre le pixel x_i^* et le centre de la région. La constante de normalisation C_1 est obtenue en imposant $\sum_{u=1}^m \hat{b}_u = 1$ [Comaniciu 03c] d'où :

$$C_1 = \frac{1}{\sum_{i=1}^n k(\|x_i^*\|^2)} \quad (3.17)$$

Ensuite, les coefficients des poids ont été définis par la formule suivante [Comanciu 03c] :

$$\left\{ \vartheta_u = \min \left(\frac{\hat{b}^*}{\hat{b}_u}, 1 \right) \right\}_{u=1 \dots m} \quad (3.18)$$

avec $\hat{b}^*$ la valeur minimale non nulle de la représentation. Ces poids sont utilisés pour définir une transformation pour les représentations du modèle de la cible et les cibles candidates. Cette transformation réduit l'importance des caractéristiques qui ont des poids ϑ_u faibles, c-à-d, celles qui sont dominantes dans l'arrière-plan et moins importantes pour la représentation de la cible. En revanche, cette représentation est calculée dans une zone elliptique autour de la cible d'une superficie égale à trois fois la taille de celle-ci.

Par la suite, Ning et al. [Ning 10a] ont montré que les poids attribués aux pixels dans la région candidate de la cible par l'histogramme pondéré de l'arrière-plan BWH sont proportionnels à ceux sans information de l'arrière-plan. Ensuite, ces auteurs ont proposé un nouveau modèle de l'arrière-plan appelé l'histogramme pondéré et corrigé de l'arrière-plan CBWH (Corrected Background-Weighted Histogram) en transformant seulement le modèle de la cible, mais pas les cibles candidates. D'où, le modèle de la cible transformé est obtenu par la formule suivante :

$$\left\{ \begin{array}{l} \hat{q}' = \{\hat{q}'_u\}_{u=1 \dots m} \\ \hat{q}'_u = C_2 \vartheta_u \sum_{i=1}^n k(\|x_i^*\|^2) \beta[l(x_i^*) - u] \end{array} \right. \quad (3.19)$$

avec $\hat{q}'_u$ représente la probabilité de la caractéristique, $u = 1 \dots m$, dans ce modèle, et C_2 est une constante de normalisation définie par :

$$C_2 = \frac{1}{\sum_{i=1}^n k(\|x_i^*\|^2) \sum_{u=1}^m \vartheta_u \beta[l(x_i^*) - u]} \quad (3.20)$$

Cependant, l'histogramme pondéré et corrigé de l'arrière-plan CBWH permet de réduire l'importance des caractéristiques de l'arrière-plan seulement dans le modèle de la cible. D'où, le modèle de la cible candidate utilisé est le modèle original comme suit :

$$\left\{ \begin{array}{l} \hat{p}(y) = \{\hat{p}_u(y)\}_{u=1 \dots m} \\ \hat{p}_u(y) = C_3 \sum_{i=1}^{n_h} k\left(\left\|\frac{y - x_i}{h}\right\|^2\right) \beta[l(x_i^*) - u] \end{array} \right. \quad (3.21)$$

avec $\hat{p}_u(y)$ représente la probabilité de la caractéristique, $u = 1 \dots m$, dans ce modèle, et C_3 est une constante de normalisation définie par :

$$C_3 = \frac{1}{\sum_{i=1}^{n_h} k(\|\frac{y - x_i}{h}\|^2) \sum_{u=1}^m \beta[l(x_i^*) - u]} \quad (3.22)$$

La nouvelle formule des poids ω'_i calculée par l'histogramme CBWH est définie par :

$$\omega'_i = \sum_{u=1}^m \sqrt{\frac{\hat{q}'_u}{\hat{p}_u(y)}} \beta[l(x_i^*) - u] \quad (3.23)$$

Dû aux avancées importantes réalisées par l'histogramme CBWH [Ning 10a] pour le suivi d'objets, nous avons adopté cet histogramme dans ce travail afin d'exploiter les caractéristiques principales de l'arrière-plan pour la représentation de la cible.

3.2.3 Algorithme proposé

Dans ce qui suit, nous présentons les phases essentielles de notre approche pour le suivi d'une personne basée sur l'histogramme pondéré et corrigé de l'arrière-plan et les descripteurs de moments du poids de l'image :

Algorithme : Suivi d'une personne basé sur l'histogramme pondéré et corrigé de l'arrière-plan.

Nous disposons de la distribution de probabilité du modèle de la cible $\{\hat{q}_u\}_{u=1 \dots m}$ calculée par l'équation (B.1) (Annexe B) et l'estimation de sa position initiale $\hat{y}_0$ dans la première image de la séquence.

Début

1. Calculer l'histogramme pondéré de l'arrière-plan $\{\hat{b}_u\}_{u=1 \dots m}$ et les poids $\{\vartheta_u\}_{u=1 \dots m}$ selon les équations (3.16) et (3.18) respectivement, ensuite calculer le modèle de la cible transformé $\{\hat{q}'_u\}_{u=1 \dots m}$ selon (3.19).
2. Initialiser la position de la cible candidate $\hat{p}(\hat{y}_0)$ dans l'image courante avec $\hat{y}_0$, définir le nombre d'itérations $k = 0$, puis calculer la distribution $\{\hat{p}_u(\hat{y}_0)\}_{u=1 \dots m}$ selon (3.21) dans l'image courante, puis évaluer le coefficient de Bhattacharyya $\rho[\hat{p}(\hat{y}_0), \hat{q}'] = \sum_{u=1}^m \sqrt{\hat{p}_u(\hat{y}_0) \hat{q}'_u}$.
3. Calculer la nouvelle formule des poids $\{\omega'_i\}_{i=1 \dots n_h}$ selon l'équation (3.23).
4. Trouver la nouvelle position $\hat{y}_1$ de la cible candidate selon l'équation (3.1).
5. Calculer la distribution de probabilité $\hat{p}_u(\hat{y}_1)_{u=1 \dots m}$, puis évaluer le coefficient de

Bhattacharyya $\rho[\hat{\mathbf{p}}(\hat{\mathbf{y}}_1), \hat{\mathbf{q}}'] = \sum_{u=1}^m \sqrt{\hat{\mathbf{p}}_u(\hat{\mathbf{y}}_1) \hat{\mathbf{q}}'_u}$.

6. Tant que $\rho[\hat{\mathbf{p}}(\hat{\mathbf{y}}_1), \hat{\mathbf{q}}'] < \rho[\hat{\mathbf{p}}(\hat{\mathbf{y}}_0), \hat{\mathbf{q}}']$

Faire $\hat{\mathbf{y}}_1 \leftarrow \frac{1}{2}(\hat{\mathbf{y}}_0 + \hat{\mathbf{y}}_1)$

Evaluer $\rho[\hat{\mathbf{p}}(\hat{\mathbf{y}}_1), \hat{\mathbf{q}}']$

7. Soit $\mathbf{d} \leftarrow \|\hat{\mathbf{y}}_1 - \hat{\mathbf{y}}_0\|$

Si $\mathbf{d} < \varepsilon$ ou $k \geq N_{max}$

Arrêter l'itération et aller à l'étape 10.

Sinon

$\hat{\mathbf{y}}_0 \leftarrow \hat{\mathbf{y}}_1, k \leftarrow k + 1$ et aller à l'étape 3.

8. Calculer le poids de l'image ω'_i selon l'équation (3.23), puis trouver la région estimée de la cible $\mathbf{A}$ en utilisant l'équation (3.4).

9. Analyser les descripteurs de moments du poids de l'image selon les équations (3.6) et (3.7) respectivement.

10. Estimer la largeur, la hauteur et l'orientation de la cible candidate en utilisant l'équation (3.14).

11. Estimer la taille de la région cible candidate dans l'image suivante en utilisant l'équation (3.15).

Fin

La valeur par défaut pour le critère d'arrêt ε utilisé dans l'étape 7 est $\varepsilon = 0.1$ [Comanciu 03c]. La complexité de cet algorithme est très faible car elle nécessite des opérations très simples comme le calcul de la distribution de probabilité, le coefficient de Bhattacharyya et l'estimation de la largeur, la hauteur et l'orientation de la cible. En effet, la complexité du temps de celui-ci est donnée par $O(Tn^2)$ où T est le nombre d'itérations et n est le nombre des points dans l'ensemble de données.

3.3 Méthode de suivi basée sur les caractéristiques couleur-texture

Dans cette section, nous allons présenter une autre méthode intitulée « Suivi d'une personne basé sur un modèle d'apparence (Human tracking based on appearance model) [Laaroussi 16a] ». Cette méthode combine les caractéristiques de couleur-texture et l'information de l'arrière-plan pour limiter les ambiguïtés liées à la représentation par histogramme de couleurs et diminuer l'influence de l'arrière-plan dans la localisation de la

cible. Les principales contributions de ce travail sont doubles, premièrement, les caractéristiques de texture de l'objet ont été utilisées en combinaison avec les caractéristiques de couleurs et qui sont extraites par la technique des motifs binaires locaux [Ojala 02b] pour représenter l'objet (section 3.3.1). Deuxièmement, nous avons utilisé un modèle de représentation de l'arrière-plan intitulé l'histogramme pondéré de l'arrière-plan BWH (Background-Weighted Histogram) [Comaniciu 03c]. Enfin, nous avons combiné ces deux concepts afin de générer de nouvelles caractéristiques pour la représentation de la cible et le suivi des personnes dans une séquence d'images.

3.3.1 Description de la texture avec les motifs binaires locaux

Les motifs binaires locaux LBP (Local Binary Patterns) [Ojala 96a, Ojala 02b] sont des descripteurs qui décrivent chaque pixel d'une image en se basant sur la texture. Cette dernière est calculée par rapport au niveau de gris des pixels voisins. Les motifs binaires locaux sont largement utilisés pour la soustraction de l'arrière-plan [Heikkilä 06] et le suivi d'objets [Ning 09b]. Ces descripteurs ont été initialement proposés par Ojala et al. [Ojala 96a] afin de caractériser les textures présentées dans les images en niveaux de gris. Le point fort de la technique des motifs binaires locaux est sa robustesse par rapport à la variation des conditions d'éclairage obtenue grâce à la relation locale entre le pixel et ses voisins. Pour chaque pixel le descripteur LBP est obtenu par l'algorithme présenté sur la figure 3.1 ci-après comme suit, premièrement, les différences entre les intensités du pixel central et ses voisins sont calculées en soustrayant l'intensité de ceux-ci par l'intensité du pixel central. Ensuite, nous codons par le bit 1 les voisins qui ont une intensité plus forte que le pixel central et par le bit 0 les autres. Grâce à l'étape de pondération, chaque pixel est décrit par un nombre entre 2^0 et 2^7 (8 bits). Enfin, la somme de ces nombres donne le code LBP du pixel central et le descripteur final est une matrice de codes LBP qui peut être représentée par un histogramme. Cependant, la version originale de l'opérateur LBP ne considère que les huit voisins d'un pixel [Ojala 96a], mais la définition peut être facilement étendue pour inclure tous les voisins circulaires avec n'importe quel nombre de pixels [Ojala 02b]. Formellement, le codage par la méthode LBP classique du pixel (x_c, y_c) est alors exprimé par la formule suivante [Ojala 02b] :

$$LBP_{P,R}(x_c, y_c) = \sum_{j=0}^{P-1} s(e_j - e_c) 2^j \quad (3.24)$$

avec P désigne le nombre de points échantillonnés, R le rayon du cercle, e_c et e_j représentent, respectivement, le niveau de gris du pixel central et celui du $j^{ème}$ pixel dans le

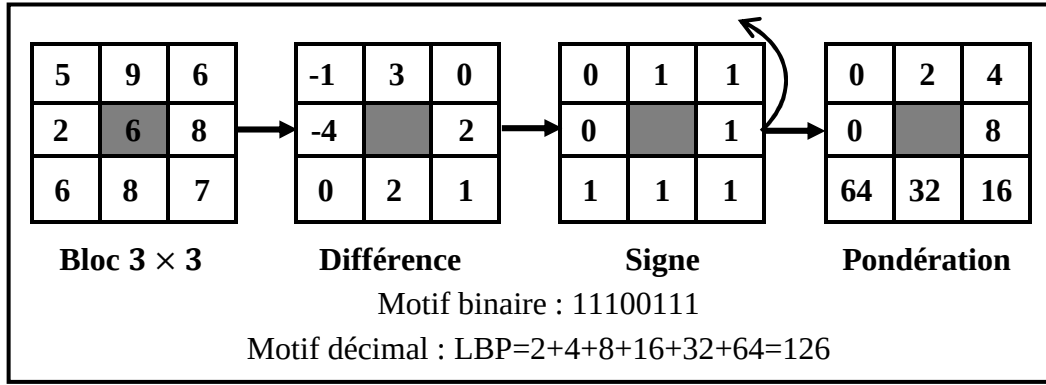


Figure 3.1 : Processus de calcul d'un motif binaire local (LBP).

voisinage circulaire de rayon R , et $s(x)$ est la fonction binaire d'Heaviside [Ito 91] définie par :

$$s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (3.25)$$

Les motifs binaires locaux $LBP_{P,R}$ obtenus par l'équation (3.24) sont invariants seulement à la transformation monotone de l'échelle, ce qui permet de traiter la texture à différentes échelles. Une autre extension des motifs binaires locaux concerne la définition des motifs uniformes qui permettent de réduire la taille du vecteur descripteur et de rendre les motifs binaires invariants par rapport à la rotation [Ojala 02b]. Un motif binaire uniforme est défini comme étant tout motif possédant exactement 0 ou 2 transitions (01 ou 10) dans un parcours circulaire [Ojala 02b]. Formellement, les motifs binaires uniformes et invariants en rotation qui ont une valeur de transition U au plus 2 $LBP_{P,R}^{riu2}$ sont définis par [Ojala 02b] :

$$LBP_{P,R}^{riu2} = \begin{cases} \sum_{j=0}^{P-1} s(e_j - e_c) & \text{si } U(LBP_{P,R}) \leq 2 \\ P + 1 & \text{Sinon} \end{cases} \quad (3.26)$$

Où

$$U(LBP_{P,R}) = |s(e_{P-1} - e_c) - s(e_0 - e_c)| + \sum_{j=1}^{P-1} |s(e_j - e_c) - s(e_{j-1} - e_c)| \quad (3.27)$$

L'une des limitations des motifs binaires locaux est qu'ils ne fonctionnent pas correctement sur les régions planes comme le ciel, où les valeurs des niveaux de gris des pixels voisins sont très proches à la valeur du pixel central. En pratique, ce type de codage s'est avéré sensible aux images bruitées. Ceci à cause de l'effet du seuillage adopté par la technique des motifs binaires locaux [Pietikäinen 11]. Pour remédier à cet inconvénient, les auteurs Heikkilä et al. [Heikkilä 06] ont modifié la stratégie de seuillage de l'opérateur

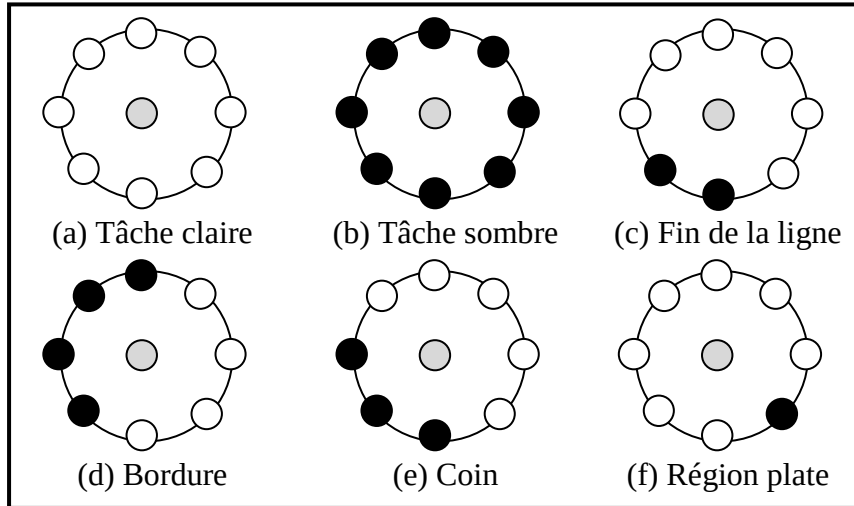


Figure 3.2 : Quelques exemples des motifs binaires locaux uniformes et invariants à la rotation $LBP_{8,1}^{riu2}$.

LBP en remplaçant le terme $s(e_j - e_c)$ dans les équations (3.24), (3.26) et (3.27) par le terme $s(e_j - e_c + a)$. La plus grande valeur de $|a|$ est le plus grand changement dans les valeurs des pixels qui sont autorisées sans affecter autant les résultats de seuillage [Heikkilä 06]. La figure 3.2 ci-dessus, présente quelques exemples de labels pour les motifs binaires uniformes et invariants à la rotation obtenus dans le cas d'un rayon unitaire $R = 1$ et de huit voisins = 8 $LBP_{8,1}^{riu2}$. Généralement, le modèle $LBP_{8,1}^{riu2}$ a neuf motifs uniformes et invariants à la rotation pour représenter la texture de la cible [Ojala 02b]. Par exemple, les motifs '0' et '8' correspondent respectivement des tâches claires et sombres, le motif '2' représente la fin de la ligne, le motif '3' est un coin et le motif '4' est une bordure. Sur la figure 3.2 ci-dessus, les '1' sont représentés par des cercles noirs, et les '0' par des cercles blancs. Cependant, pour la représentation de la cible en utilisant les caractéristiques de texture, les motifs suivants : les bordures, la fin de la ligne et les coins sont appelés les motifs uniformes majeurs qui décrivent les principales caractéristiques de la cible, alors que les tâches claires et sombres représentent les motifs uniformes mineurs [Ojala 02b]. Les motifs uniformes majeurs et invariants à la rotation sont extraits par l'équation suivante :

$$LBP_{8,1}^{riu2} = \begin{cases} \sum_{j=0}^7 s(e_j - e_c + a) & U(LBP_{8,1}) \leq 2 \text{ et} \\ \sum_{j=0}^7 s(e_j - e_c + a) \in \{2,3,4,5,6\} & \\ 0 & \text{Sinon} \end{cases} \quad (3.28)$$

Dans l'opérateur $LBP_{8,1}^{riu2}$, les étiquettes correspondant à des motifs uniformes mineurs sont 0, 1, 7 et 8, respectivement, les étiquettes des motifs non uniformes sont 9, et les

étiquettes correspondant aux motifs uniformes majeurs sont $\{2, 3, 4, 5, 6\}$, qui sont cinq motifs [Ning 09b].

3.3.2 Histogramme pondéré de l'arrière-plan

Nous utilisons le modèle de représentation de l'arrière-plan BWH (Background-Weighted Histogram) proposé par Comaniciu et al. [Comaniciu 03c] afin de dériver une représentation simple des caractéristiques de l'arrière-plan et de ne sélectionner que les parties saillantes des représentations du modèle de la cible et des cibles candidates. Pour calculer le modèle original de l'arrière-plan $\hat{b}$ et les coefficients de poids ϑ_u , nous avons utilisé les formules des équations (3.16) et (3.22) de la section (3.2.2) respectivement. Cependant, la représentation de l'arrière-plan est calculée dans une zone rectangulaire autour de la cible d'une superficie égale à trois fois la taille de celle-ci [Comaniciu 03c]. Dans l'histogramme pondéré de l'arrière-plan BWH, les poids ϑ_u sont utilisés pour définir une transformation pour les représentations du modèle de la cible et des cibles candidates. D'où, le modèle de la cible transformé est calculé par l'équation (3.19) de la section (3.2.2) et le modèle de la cible candidate transformé est obtenu par l'équation suivante :

$$\begin{cases} \hat{p}'(y) = \{\hat{p}'_u(y)\}_{u=1 \dots m} \\ \hat{p}'_u(y) = C_4 \vartheta_u \sum_{i=1}^{n_h} k\left(\left\|\frac{y - x_i}{h}\right\|^2\right) \beta[l(x_i^*) - u] \end{cases} \quad (3.29)$$

avec $\hat{p}'_u(y)$ représente la probabilité de la caractéristique, $u = 1 \dots m$, dans ce modèle, et C_4 est une constante de normalisation définie par :

$$C_4 = \frac{1}{\sum_{i=1}^n k\left(\left\|\frac{y - x_i}{h}\right\|^2\right) \sum_{u=1}^m \vartheta_u \beta[l(x_i^*) - u]} \quad (3.30)$$

La nouvelle formule des poids ω_i'' calculée par l'histogramme pondéré de l'arrière-plan (BWH) est définie par :

$$\omega_i'' = \sum_{u=1}^m \sqrt{\frac{\hat{q}'_u}{\hat{p}'_u(y)}} \beta[l(x_i^*) - u] \quad (3.31)$$

Dû aux avancées importantes réalisées par l'histogramme pondéré de l'arrière-plan (BWH) [Comaniciu 03c] pour le suivi d'objets, nous avons adopté cet histogramme dans ce travail afin d'exploiter les caractéristiques principales de l'arrière-plan pour la représentation de la cible.

3.3.3 Algorithme proposé

Dans ce qui suit, nous présentons les phases essentielles de notre approche pour le suivi d'une personne basée sur les caractéristiques couleur-texture :

Algorithme : Suivi d'une personne basé sur les caractéristiques couleur-texture

Nous disposons de la distribution de probabilité du modèle de la cible $\{\hat{q}_u\}_{u=1\dots m}$ calculée par l'équation (B.1) (Annexe B), et l'estimation de sa position initiale $\hat{y}_0$ dans la première image de la séquence.

Début

1. Calculer la distribution des caractéristiques couleur-texture du modèle de la cible $\{\hat{q}_u\}_{u=1\dots m}$ dans la position initiale $\hat{y}_0$ selon les équations (B.1) (Annexe B) et (3.28). Ensuite, calculer l'histogramme pondéré de l'arrière-plan $\{\hat{b}_u\}_{u=1\dots m}$ et les coefficients des poids $\{\vartheta_u\}_{u=1\dots m}$ par les équations (3.16) et (3.22) respectivement, puis calculer le modèle de la cible transformé $\{\hat{q}'_u\}_{u=1\dots m}$ selon l'équation (3.19).
2. Initialiser la position de la cible candidate $\hat{p}(\hat{y}_0)$ dans l'image courante avec $\hat{y}_0$, définir le nombre d'itérations $k = 0$, puis calculer la distribution des caractéristiques couleur-texture de la cible candidate $\{\hat{p}'_u(\hat{y}_0)\}_{u=1\dots m}$ selon (3.29) dans l'image courante, puis évaluer le coefficient de Bhattacharyya $\rho[\hat{p}(\hat{y}_0), \hat{q}'] = \sum_{u=1}^m \sqrt{\hat{p}_u(\hat{y}_0) \hat{q}'_u}$.
3. Calculer la nouvelle formule des poids $\{\omega'_i\}_{i=1\dots n_h}$ selon l'équation (3.31).
4. Trouver la nouvelle position $\hat{y}_1$ de la cible candidate selon l'équation (B.10) (Annexe B).
5. Calculer $\{\hat{p}'_u(\hat{y}_1)\}_{u=1\dots m}$ et évaluer le coefficient de Bhattacharyya $\rho[\hat{p}'(\hat{y}_1), \hat{q}'] = \sum_{u=1}^m \sqrt{\hat{p}'_u(\hat{y}_1) \hat{q}'_u}$.
6. Tant que $\rho[\hat{p}'(\hat{y}_1), \hat{q}'] < \rho[\hat{p}'(\hat{y}_0), \hat{q}']$
 Faire $\hat{y}_1 \leftarrow \frac{1}{2}(\hat{y}_0 + \hat{y}_1)$
 Evaluer $\rho[\hat{p}'(\hat{y}_1), \hat{q}']$
7. Soit $d \leftarrow \|\hat{y}_1 - \hat{y}_0\|$
 Si $d < \varepsilon$ ou $k \geq N_{max}$
 Arrêter l'itération.

Sinon

$\hat{\mathbf{y}}_0 \leftarrow \hat{\mathbf{y}}_1, \mathbf{k} \leftarrow \mathbf{k} + 1$ et aller à l'étape 3.

Fin

La valeur par défaut pour le critère d'arrêt ε utilisé dans l'étape 7 est $\varepsilon = 0.1$ [Comanciu 03c]. La complexité de cet algorithme est très faible car elle nécessite des opérations très simples comme le calcul de la distribution des probabilités. En effet, la complexité du temps de celui-ci est donnée par $O(Tn^2)$ où T est le nombre d'itérations et n est le nombre des points dans l'ensemble de données.

3.4 Nouvelle approche de suivi basée sur l'histogramme pondéré du premier-plan

Dans cette section, nous proposons une nouvelle approche pour le suivi des personnes dans une séquence d'images intitulée « suivi d'une personne en utilisant la jointure des caractéristiques couleur-texture et l'histogramme pondéré du premier-plan (Human tracking using joint color-texture features and foreground-weighted histogram CTFWH) » [Laaroussi 17b]. Ce travail vise à renforcer notre précédente contribution présentée dans la section 3.3 [Laaroussi 16a] en améliorant considérablement la précision de suivi, le temps de calcul, et la robustesse aux diverses interférences de l'arrière-plan et des scènes similaires à la cible suivie.

Cette approche consiste en trois étapes principales. Dans la première étape, nous avons utilisé une méthode simple et efficace pour l'extraction des caractéristiques de la texture basée sur la technique des motifs binaires locaux (LBP) [Ojala 02b]. Outre, les caractéristiques classiques de l'histogramme de couleurs, les motifs uniformes majeurs LBP [Ojala 02b] sont également exploités pour former un histogramme combiné des caractéristiques couleur-texture. En effet, l'histogramme combiné des caractéristiques couleur-texture est calculé pour les régions de la cible et celles du premier-plan, afin de générer une représentation plus distinctive et efficace de la cible (voir section 3.4.1). Dans la deuxième étape, nous proposons un modèle de représentation de la région du premier-plan nommé l'histogramme pondéré du premier-plan FWH (Foreground-Weighted Histogram), afin d'exploiter les caractéristiques importantes du premier-plan et de les utiliser pour sélectionner uniquement les parties saillantes du modèle de la cible. Cependant, en se basant sur le modèle du premier-plan, nous calculons les coefficients de pondération, puis nous transformons la représentation du

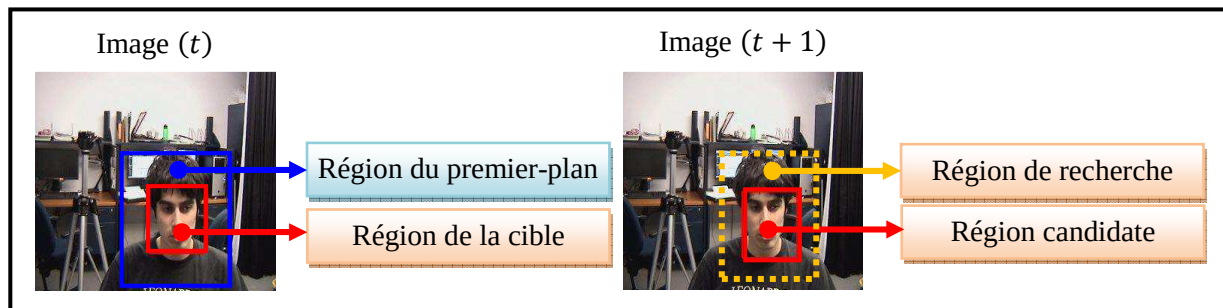


Figure 3.3 : Ensemble de toutes les régions utilisées : la région de la cible, la région du premier-plan, la région de recherche et la région candidate.

modèle de la cible, telle que les caractéristiques avec des poids plus grands sont les plus importantes pour la représentation de la cible (voir la section 3.4.2). Dans la troisième étape, nous proposons une méthode simple pour la mise à jour du premier-plan à cause des changements drastiques de celui-ci durant le processus de suivi où ce modèle est calculé pour chaque image. Pour ce faire, le coefficient de Bhattacharyya [Kailath 67] est calculé et puis seuillé entre le modèle courant du premier-plan calculé dans l'image courante et celui calculé dans l'image précédente (voir la section 3.4.3). Enfin, en combinant toutes ces étapes, nous générons une nouvelle représentation du modèle de la cible, à savoir la combinaison des caractéristiques couleur-texture et l'histogramme pondéré du premier-plan (CTFWH). L'algorithme Mean Shift est ensuite appliqué pour rechercher la position la plus proche de la cible dans l'image courante. Afin de décrire clairement notre méthode, nous illustrons tous les types de régions utilisées sur la figure 3.3 ci-dessus. Cependant, toutes les régions sont représentées par des rectangles dans l'image. Pour éliminer l'influence de différentes dimensions d'une région, toutes les régions sont normalisées.

3.4.1 Histogramme combiné des caractéristiques couleur-texture

La représentation de l'apparence par un histogramme de couleurs est largement utilisée pour le suivi d'objets [Comaniciu 03c]. Cette représentation peut être considérée comme une fonction de densité de probabilité PDF (Probability Density Function) de la région cible. Cependant, la description de l'apparence en utilisant uniquement les caractéristiques de couleurs pour caractériser la cible ne peut pas donner assez d'information. L'aberration de la couleur est causée par de nombreuses situations comme les changements d'illumination et quand la cible et son arrière-plan ont une couleur similaire. Pour surmonter ces limites, nous introduisons une combinaison des caractéristiques couleur-texture dans l'algorithme Mean Shift comme utilisé par Ning et al. [Ning 09b]. La figure 3.4 ci-après présente une illustration de l'histogramme combiné des caractéristiques couleur-texture correspondant au

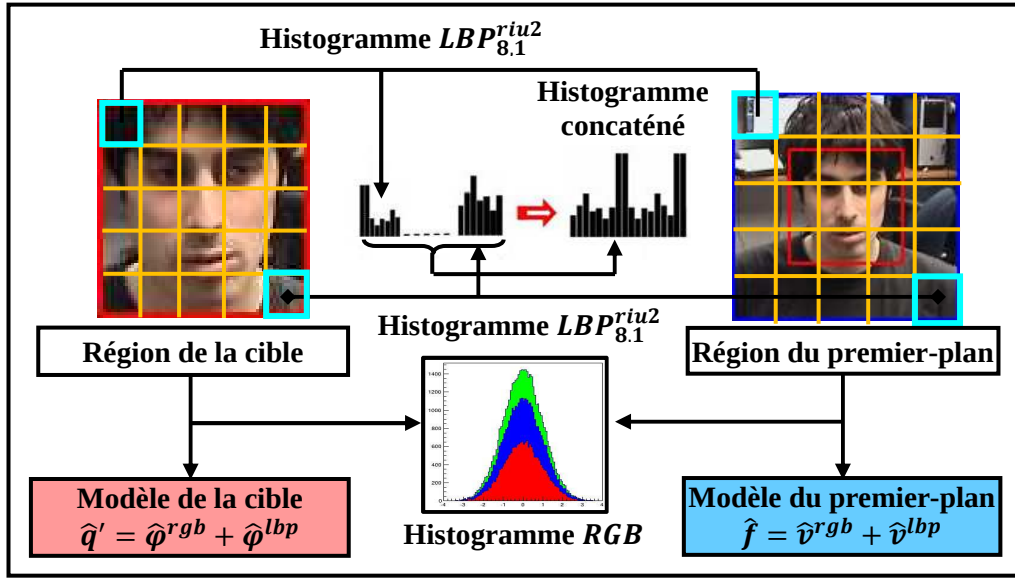


Figure 3.4 : Calcul des histogrammes combinés des caractéristiques couleur-texture $\hat{q}'$ et $\hat{f}$ correspondants au modèle de la cible et celui du premier-plan.

modèle de la cible et celui du premier-plan.

Pour caractériser le modèle de la cible, il faut d'abord, choisir un espace de caractéristiques. Dans ce travail, nous choisissons l'espace de couleur RGB comme espace de caractéristiques et il a été quantifié en $Rbins \times Gbins \times Bbins$ classes, avec $Rbins$, $Gbins$ et $Bbins$ représentent les classes quantifiés pour les canaux de couleur R , G et B respectivement. Soit $\{x_i^*\}_{i=1 \dots n}$ l'ensemble des coordonnées normalisées de n pixels du modèle de la cible centré à 0. Basé sur les travaux de [Comaniciu 03c], le modèle de la cible couleur $\hat{\varphi}^{rgb}$ pour la région de la cible est calculé par :

$$\begin{cases} \hat{\varphi}^{rgb} = \{\hat{\varphi}_u^{rgb}\}_{u=1 \dots m} \\ \hat{\varphi}_u^{rgb} = C_5 \sum_{i=1}^n k(\|x_i^*\|^2) \beta[l(x_i^*) - u] \end{cases} \quad (3.32)$$

avec $\hat{\varphi}_u^{rgb}$ représente la probabilité de la caractéristique, $u = 1 \dots m$, dans le modèle de la cible, m est la taille de l'histogramme. Les termes β , l et $k(x)$ sont les mêmes que ceux utilisés dans l'équation (3.16). La constante de normalisation C_5 est obtenue en imposant $\sum_{u=1}^m \hat{\varphi}_u^{rgb} = 1$:

$$C_5 = \frac{1}{\sum_{i=1}^n k(\|x_i^*\|^2)} \quad (3.33)$$

Dans ce travail, nous avons utilisé les motifs binaires locaux uniformes et invariants à la rotation $LBP_{8,1}^{riu2}$ [Ojala 02b], afin d'exploiter les caractéristiques importantes de la texture du

modèle de la cible. Néanmoins, le vecteur des caractéristiques de la région de la cible $LBP_{8,1}^{riu2}$ est créé en divisant la fenêtre en blocs (3×3 pixels pour chaque bloc), puis l'opérateur $LBP_{8,1}^{riu2}$ est appliqué pour obtenir le motif binaire pour chaque bloc comme expliqué dans la section 3.3.1. L'histogramme $LBP_{8,1}^{riu2}$ est calculé pour chaque bloc et ensuite, nous concaténons les histogrammes normalisés de tous les blocs. Cela donne un vecteur des caractéristiques pour toute la région. Un exemple d'histogramme $LBP_{8,1}^{riu2}$ est montré sur la figure 3.4 ci-avant. D'où, l'histogramme $LBP_{8,1}^{riu2}$ est présenté par :

$$\hat{\phi}^{lbp} = \{\hat{\phi}_u^{lbp}(lbpBins)\}_{u=1\dots m} \quad (3.34)$$

où $lbpBins$ représente la classe quantifiée des motifs $LBP_{8,1}^{riu2}$ extraits par l'équation (3.28). D'après les analyses présentées précédemment, l'histogramme combiné des caractéristiques couleur-texture est obtenu en calculant l'opérateur $LBP_{8,1}^{riu2}$ sur les trois canaux de l'espace des couleurs RGB indépendamment, puis en concaténant tous les résultats obtenus. Le modèle de la cible $\hat{q}'$ correspondant à la région de la cible est calculé comme suit:

$$\begin{cases} \hat{q}' = \hat{\phi}^{rgb} + \hat{\phi}^{lbp} \\ \hat{q}'_u = \{\hat{\phi}_u^{rgb} + \hat{\phi}_u^{lbp}\}_{u=1\dots m} \end{cases} \quad (3.35)$$

avec $\hat{q}'_u$ représente la probabilité des couleurs, $u = 1 \dots m$, dans le modèle de la cible. De même, soit $\{x_i\}_{i=1\dots n_h}$ l'ensemble des coordonnées normalisées des n_h pixels de la cible candidate centré sur y dans l'image courante. En utilisant le même profil de noyau $k(x)$ mais avec un rayon h . Le modèle de la cible candidate $\hat{\delta}^{rgb}$ correspondant à la région candidate est calculé par :

$$\begin{cases} \hat{\delta}(y) = \{\hat{\delta}_u(y)\}_{u=1\dots m} \\ \hat{\delta}_u(y) = C_6 \sum_{i=1}^{n_h} k\left(\left\|\frac{y - x_i}{h}\right\|^2\right) \beta[l(x_i) - u] \end{cases} \quad (3.36)$$

avec $\hat{\delta}_u^{rgb}$ représente la probabilité de la caractéristique, $u = 1 \dots m$, dans le modèle de la cible candidate et C_6 est une constante de normalisation définie par :

$$C_6 = \frac{1}{\sum_{i=1}^{n_h} k\left(\left\|\frac{y - x_i}{h}\right\|^2\right) \sum_{u=1}^m \beta[l(x_i) - u]} \quad (3.37)$$

D'où, l'histogramme combiné des caractéristiques couleur-texture du modèle de la cible candidate $\hat{p}'(y)$ est donné par :

$$\begin{cases} \hat{p}'(y) = \hat{o}^{rgb}(y) + \hat{o}^{lbp}(y) \\ \hat{p}'_u(y) = \{\hat{o}_u^{rgb}(y) + \hat{o}_u^{lbp}(y)\}_{u=1\dots m} \end{cases} \quad (3.38)$$

Au niveau de l'implémentation, les deux histogrammes *RGB* et $LBP_{8,1}^{riu2}$ sont stockés dans un seul vecteur de taille $m = Rbins \times Gbins \times Bbins \times lbpBins$. Dans ce travail, nous calculons l'histogramme combiné des caractéristiques couleur-texture également pour la région du premier-plan, comme illustré sur la figure 3.4 ci-avant. Dans la section suivante, nous présentons en détail la représentation de la cible basée sur l'histogramme pondéré du premier-plan et son intégration dans l'algorithme Mean Shift [Comaniciu 03c].

3.4.2 Histogramme pondéré du premier-plan

L'information du premier-plan joue un rôle très important, puisqu'elle est souvent présentée dans la région sélectionnée de la cible. En effet, un modèle d'apparence efficace devrait être capable de mettre en valeur les caractéristiques principales du premier-plan afin d'améliorer la précision de suivi d'objets. Cependant, si certaines caractéristiques principales du premier-plan sont également importantes dans la cible, l'élimination ou la négligence de ces caractéristiques peuvent réduire efficacement la performance de l'algorithme de suivi. Ainsi, en considérant seulement les caractéristiques de la cible dans le modèle de la représentation peuvent ne pas être suffisantes et alors l'information du premier-plan reste également un concept critique pour la représentation de la cible et le suivi d'objets comme les autres caractéristiques.

Dans ce travail, nous avons proposé un modèle de représentation de la région du premier-plan nommé l'histogramme pondéré du premier-plan (Foreground-Weighted Histogram FWH) pour dériver une représentation simple des caractéristiques du premier-plan et de les utiliser pour ne sélectionner que les caractéristiques importantes dans la représentation de la cible. La représentation du premier-plan est calculée dans une région autour de la cible. L'étendue de la région du premier-plan est égale à deux fois la région de la cible. Pour une illustration claire de l'approche proposée, nous avons présenté sur la figure 3.5 ci-après, un exemple pour trouver la nouvelle position de la cible à l'image $(t + 1)$. Dans ce travail, nous avons adopté l'opérateur $LBP_{8,1}^{riu2}$ [Ojala 02b] pour représenter également les caractéristiques de texture du premier-plan et ensuite nous utilisons l'histogramme combiné des caractéristiques couleur-texture de la région du premier-plan pour une représentation plus distinctive et efficace de la cible, comme illustré sur la figure 3.4 ci-après. Soit $\{x_i^*\}_{i=1\dots n}$ l'ensemble des coordonnées normalisées de n pixels dans le modèle du premier-plan centré

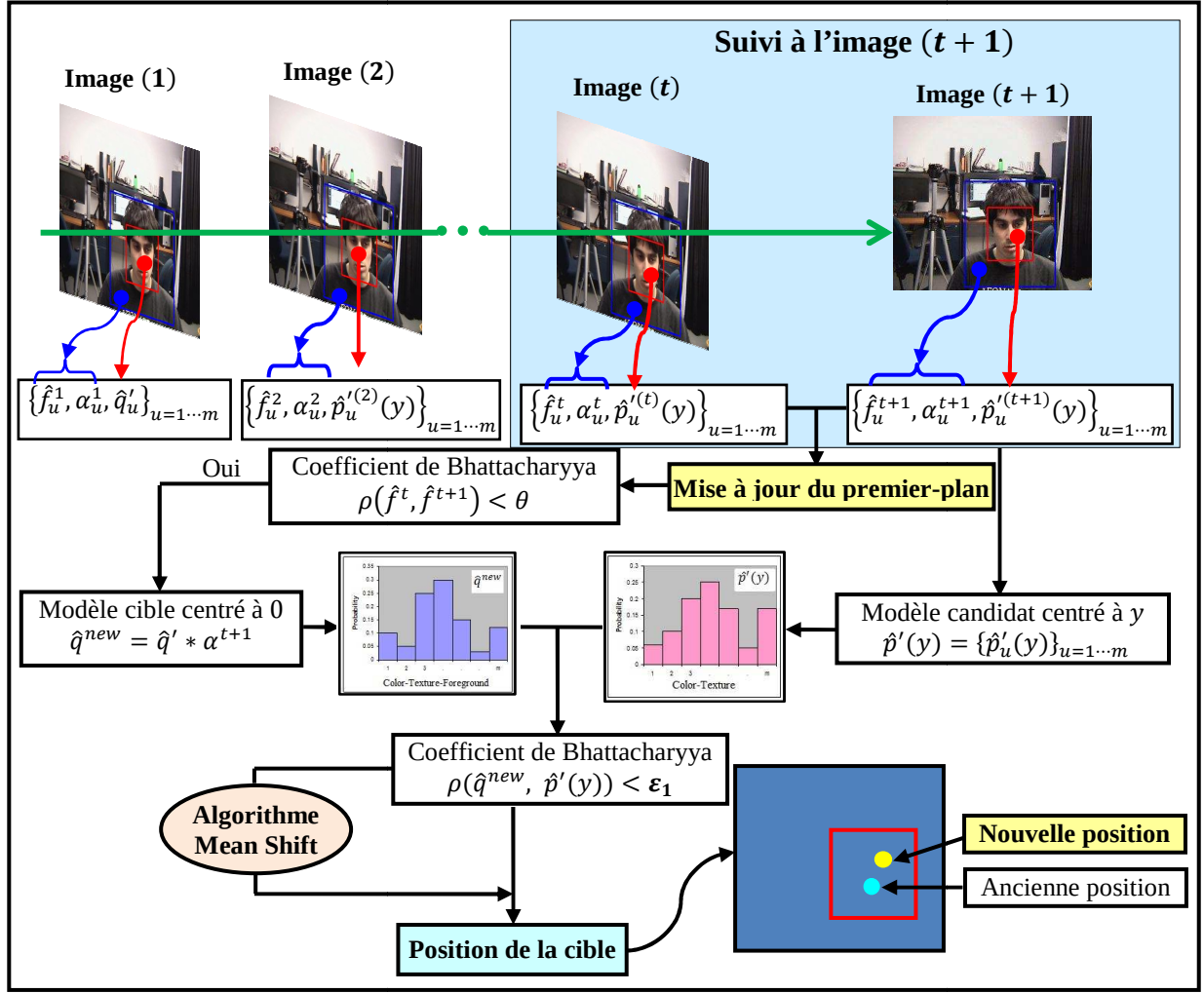


Figure 3.5 : Étapes de l'approche proposée. Cette figure illustre un exemple pour trouver la nouvelle position de la cible à l'image (t + 1) en utilisant l'approche proposée CTFWH.

à 0. D'où, le modèle du premier-plan $\hat{v}^{rgb}$ correspondant à la région de celui-ci est calculé comme suit :

$$\begin{cases} \hat{v}^{rgb} = \{\hat{v}_u^{rgb}\}_{u=1 \dots m} \\ \hat{v}_u^{rgb} = C_7 \sum_{i=1}^n k(\|x_i^*\|^2) \beta[l(x_i^*) - u] \end{cases} \quad (3.39)$$

où $\hat{v}_u^{rgb}$ représente la probabilité de la caractéristique $u = 1 \dots m$ dans le modèle du premier-plan, m est la taille de l'histogramme et les termes β , l et $k(x)$ sont les mêmes que ceux utilisés dans l'équation (3.16). La constante de normalisation C_7 est obtenue en imposant $\sum_{u=1}^m \hat{v}_u^{rgb} = 1$ comme suit :

$$C_7 = \frac{1}{\sum_{i=1}^{n_h} k(\|x_i^*\|^2) \sum_{u=1}^m \beta[l(x_i) - u]} \quad (3.40)$$

L'histogramme $LBP_{8,1}^{riu2}$ de la région du premier-plan est représenté par :

$$\hat{v}^{lbp} = \{\hat{v}_u^{lbp}(lbpBins)\}_{u=1\dots m} \quad (3.41)$$

avec $lbpBins$ représente la classe quantifiée des motifs $LBP_{8,1}^{riu2}$ extrait par l'équation (3.28). D'où, l'histogramme combiné des caractéristiques couleur-texture correspondant à la région du premier-plan est obtenu par :

$$\begin{cases} \hat{f} = \hat{v}^{rgb} + \hat{v}^{lbp} \\ \hat{f}_u = \{\hat{v}_u^{rgb} + \hat{v}_u^{lbp}\}_{u=1\dots m} \end{cases} \quad (3.42)$$

L'idée principale de ce travail est de donner plus d'importance aux caractéristiques du premier-plan pour sélectionner uniquement les plus importantes pour la représentation de la cible. D'où, nous définissons les coefficients de pondération α_u comme suit :

$$\left\{ \alpha_u = \max \left(\frac{\hat{f}_u}{\hat{f}^*}, 1 \right) \right\}_{u=1\dots m} \quad (3.43)$$

avec $\hat{f}^*$ est la valeur maximale non nulle pour le modèle du premier-plan, qui représente la densité discrète des caractéristiques les plus importantes dans ce modèle. Cependant, les coefficients des poids α_u sont utilisés pour définir une transformation pour la représentation du modèle de la cible. Cette transformation augmente l'importance des caractéristiques qui ont des poids α_u plus grands, c'est-à-dire celles qui sont dominantes dans le modèle du premier-plan et les plus importantes pour la représentation de la cible. Ensuite, le nouveau modèle de la cible $\hat{q}^{new}$ correspondant à la région cible est calculé par la formule suivante :

$$\begin{cases} \hat{q}^{new} = \{\hat{q}_u^{new}\}_{u=1\dots m} & \sum_{u=1}^m \hat{q}_u^{new} = 1 \\ \hat{q}_u^{new} = C_8 \alpha_u \sum_{i=1}^n k(\|x_i^*\|^2) \beta[l(x_i^*) - u] \end{cases} \quad (3.44)$$

avec $\hat{q}_u^{new}$ représente la probabilité de la caractéristique, $u = 1 \dots m$ dans le modèle de la cible et C_8 est une constante de normalisation obtenue en imposant $\sum_{u=1}^m \hat{q}_u^{new} = 1$ définie par :

$$C_8 = \frac{1}{\sum_{i=1}^n k(\|x_i^*\|^2) \sum_{u=1}^m \alpha_u \beta[l(x_i^*) - u]} \quad (3.45)$$

En effet, les poids α_u sont utilisés pour transformer seulement le modèle de la cible, c'est-à-dire que nous avons augmenté l'importance des caractéristiques uniquement dans le modèle de la cible et non pas dans les cibles candidates. D'où, le modèle de la cible candidate utilisé c'est le modèle

original présenté dans l'équation (3.38). Par conséquent, la nouvelle formule des poids ω_i^{new} est donnée par :

$$\omega_i^{new} = \sum_{u=1}^m \sqrt{\frac{\hat{q}_u^{new}}{\hat{p}'_u(y)}} \beta[l(x_i) - u] \quad (3.46)$$

À partir de l'équation (B.10) (Annexe B), nous pouvons observer que les paramètres clés de l'algorithme Mean Shift sont les poids ω_i^{new} de chaque point x_i . En outre, nous avons calculé ces poids par une nouvelle formule qui prend en considération seulement le modèle de la cible ainsi qu'elle incorpore les caractéristiques du premier-plan.

3.4.3 Mise à jour du premier-plan

Dans la méthode proposée, le modèle du premier-plan est initialisé au début de suivi. En revanche, durant le processus de suivi, le premier-plan change souvent en raison des scènes dynamiques, les changements d'illumination et du point de vue, les mouvements de la caméra et de l'objet et l'occultation. Si le modèle original du premier-plan est toujours utilisé sans mise à jour, la précision de suivi peut être réduite parce que le premier-plan dans l'image courante est très différent du modèle de premier-plan dans l'image précédente. De sorte que, les caractéristiques importantes du premier-plan doivent être détectées avec ces changements. Il est nécessaire de mettre à jour dynamiquement le modèle du premier-plan afin d'améliorer les performances de suivi. Dans ce travail, nous avons proposé une méthode simple pour la mise à jour du modèle du premier-plan où ce modèle a été calculé pour chaque image. Ensuite, nous avons calculé les caractéristiques du premier-plan $\{\hat{f}'_u\}_{u=1\dots m}$ et les poids $\{\alpha'_u\}_{u=1\dots m}$ dans l'image courante. Après cela, le coefficient de Bhattacharyya [Kailath 67] entre le modèle du premier-plan dans l'image courante $\{\hat{f}'_u\}_{u=1\dots m}$ et celui dans l'image précédente $\{\hat{f}_u\}_{u=1\dots m}$ est calculé par la formule suivante :

$$\theta = \sum_{u=1}^m \sqrt{\hat{f}_u \hat{f}'_u} \quad (3.47)$$

Si θ est inférieur à un seuil, cela implique qu'il existe des changements considérables dans le modèle du premier-plan, puis nous mettons à jour le modèle du premier-plan dans l'image précédente $\{\hat{f}_u\}_{u=1\dots m}$ par le modèle du premier-plan dans l'image courante $\{\hat{f}'_u\}_{u=1\dots m}$ et aussi les poids $\{\hat{\alpha}_u\}_{u=1\dots m}$ sont mis à jour par les nouveaux poids $\{\hat{\alpha}'_u\}_{u=1\dots m}$. Finalement, le nouveau modèle de la cible $\hat{q}_u^{new}$ est ensuite calculé par l'équation (3.44) en utilisant les nouveaux poids $\{\hat{\alpha}'_u\}_{u=1\dots m}$. Sinon, nous gardons l'ancien modèle du premier-plan.

3.4.4 Algorithme proposé

Dans ce qui suit, nous présentons l'ensemble des étapes essentielles de notre approche pour le suivi d'une personne basée sur l'histogramme pondéré du premier-plan :

Algorithme : Suivi d'une personne en utilisant l'histogramme pondéré du premier-plan.

Nous disposons de la distribution de probabilité du modèle de la cible couleur $\{\hat{\varphi}_u^{rgb}\}_{u=1\dots m}$ calculée par l'équation (3.32), et l'estimation de sa position initiale $\hat{\mathbf{y}}_0$ dans la première image de la séquence.

Début

1. Calculer le modèle de la cible $\{\hat{q}'_u\}_{u=1\dots m}$ selon l'équation (3.35) et le modèle du premier-plan $\{\hat{f}_u\}_{u=1\dots m}$, les poids $\{\alpha_u\}_{u=1\dots m}$ par les équations (3.42) et (3.43) respectivement, puis calculer le modèle de la cible transformé $\{\hat{q}_u^{new}\}_{u=1\dots m}$ selon l'équation (3.44).
2. Initialiser le nombre d'itération $\mathbf{k} = 0$.
3. Initialiser la position de la cible dans l'image courante avec $\hat{\mathbf{y}}_0$. Calculer la cible candidate $\{\hat{p}'_u(\hat{\mathbf{y}}_0)\}_{u=1\dots m}$ selon l'équation (3.38) dans l'image courante, puis évaluer le coefficient de Bhattacharyya $\rho[\hat{p}'(\hat{\mathbf{y}}_0), \hat{q}^{new}] = \sum_{u=1}^m \sqrt{\hat{p}'_u(\hat{\mathbf{y}}_0) \hat{q}_u^{new}}$.
4. Calculer la nouvelle formule des poids $\{\omega_i^{new}\}_{i=1\dots n_h}$ selon l'équation (3.46).
5. Trouver la nouvelle position $\hat{\mathbf{y}}_1$ de la cible candidate selon l'équation (B.10) (Annexe B).
6. Calculer $\{\hat{p}'_u(\hat{\mathbf{y}}_1)\}_{u=1\dots m}$ et évaluer $\rho[\hat{p}'(\hat{\mathbf{y}}_1), \hat{q}^{new}] = \sum_{u=1}^m \sqrt{\hat{p}'_u(\hat{\mathbf{y}}_1) \hat{q}_u^{new}}$.
7. Tant que $\rho[\hat{p}'(\hat{\mathbf{y}}_1), \hat{q}^{new}] < \rho[\hat{p}'(\hat{\mathbf{y}}_0), \hat{q}^{new}]$
 Faire $\hat{\mathbf{y}}_1 \leftarrow \frac{1}{2}(\hat{\mathbf{y}}_0 + \hat{\mathbf{y}}_1)$
 Evaluer $\rho[\hat{p}'(\hat{\mathbf{y}}_1), \hat{q}^{new}]$
8. Soit $\mathbf{d} \leftarrow \|\hat{\mathbf{y}}_1 - \hat{\mathbf{y}}_0\|$
 Si $\mathbf{d} < \varepsilon_1$ ou $\mathbf{k} \geq N_{max}$
 Calculer le modèle du premier-plan $\{\hat{f}'_u\}_{u=1\dots m}$ et les poids $\{\alpha'_u\}_{u=1\dots m}$ dans l'image courante selon les équations (3.42) et (3.43) respectivement.

Si $\theta < \varepsilon_2$ où θ est calculé par l'équation (3.47) puis

$$\{\hat{f}_u\}_{u=1\dots m} \leftarrow \{\hat{f}'_u\}_{u=1\dots m}$$

$$\{\alpha_u\}_{u=1\dots m} \leftarrow \{\alpha'_u\}_{u=1\dots m}$$

Le modèle de la cible $\{\hat{q}_u^{new}\}_{u=1\dots m}$ est mis à jour par l'équation (3.44).

Arrêter l'itération.

Sinon

$\hat{y}_0 \leftarrow \hat{y}_1, k \leftarrow k + 1$ et aller à l'étape 4.

Fin

Dans cette contribution, nous avons choisi l'espace de couleur RGB comme espace de caractéristiques et qu'il a été quantifié en $m = 16 \times 16 \times 16$ classes. Dans l'étape 1, pour calculer l'histogramme combiné des caractéristiques couleur-texture de la cible $\{\hat{q}_u\}_{u=1\dots m}$ selon l'équation (3.35), nous avons utilisé $m = 16 \times 16 \times 16 \times 5$ comme taille de l'histogramme, où les trois premières dimensions représentent les classes quantifiées des canaux de couleurs R , G et B , respectivement. La quatrième dimension est la classe quantifiée des motifs binaires locaux uniformes et invariants à la rotation extraient par le modèle $LBP_{8,1}^{riu2}$ [Ojala 02b] et qui sont cinq motifs [Ning 09b]. Pour calculer les principaux motifs uniformes extraient par l'équation (3.28), nous avons adopté l'opérateur LBP modifié proposé dans [Heikkilä 06]. Donc nous avons utilisé $a = 3$. Dans l'implémentation de cet algorithme, les deux histogrammes RGB et $LBP_{8,1}^{riu2}$ sont empilés dans un seul vecteur de taille $m = 16 \times 16 \times 16 \times 5$. La valeur par défaut pour le critère d'arrêt ε_1 utilisé dans l'étape 8 est $\varepsilon_1 = 0.1$ [Comanciu 03c].

Basé sur les expériences, le seuil ε_2 utilisé dans l'étape 8 pour mettre à jour le modèle du premier-plan a été fixé à 0.5, c'est-à-dire, si la mesure de similarité est inférieure à 0.5, cela implique qu'il y a un changement considérable dans le modèle du premier-plan, puis nous mettons à jour l'ancien modèle du premier-plan par celui calculé dans l'image courante. La complexité de calcul de l'algorithme proposé est très faible, car elle nécessite des opérations très simples, tel que la complexité du temps de celui-ci est donnée par $O(Tn^2)$ où T est le nombre d'itérations et n est le nombre de points.

3.5 Conclusion

Dans ce chapitre, nous avons présenté trois approches. Il s'agit de nos contributions aux méthodes de suivi des personnes dans une séquence d'images basées sur les histogrammes et sous le cadre de

l'algorithme Mean Shift. Nous avons introduit dans ces approches de nouvelles caractéristiques pour la représentation de la cible et les algorithmes de suivi. La nouveauté de la première contribution réside dans la fusion de descripteurs de moments du poids de l'image et l'histogramme pondéré et corrigé de l'arrière-plan afin de concevoir un nouvel algorithme de suivi qui estime les changements de la position, l'échelle et l'orientation de la cible. Dans la deuxième contribution, nous avons exploité les caractéristiques de couleur-texture et celles de l'arrière-plan. Dans la troisième contribution, nous avons utilisé les caractéristiques de couleur-texture et celles du premier-plan. Cette approche réduit l'impact des caractéristiques partagées par la cible et l'arrière-plan ce qui permet d'établir une distinction entre la cible et son arrière-plan, ainsi qu'elle permet d'améliorer les caractéristiques principales du modèle de la cible en incorporant les coefficients du premier-plan dans la représentation de celle-ci. Dans le chapitre suivant, nous présentons l'ensemble des expérimentations effectuées pour les trois contributions SOBWH, AMT et CTFWH présentées dans ce chapitre.

Chapitre 4

Mise en œuvre pratique et expérimentations

4.1 Introduction

Dans ce chapitre, nous exposons les résultats expérimentaux, les comparaisons et les discussions pour les trois contributions présentées dans le chapitre 3. Les différentes approches proposées s'inscrivent dans le cadre de suivi des personnes dans une séquence d'images basées sur les histogrammes en utilisant l'algorithme Mean Shift [Comaniciu 03c]. Les expériences sont effectuées en utilisant un ordinateur Intel ® Core™ 2 Duo CPU T5800 avec 2.00 GHz CPU et 3 Go de RAM. Le noyau d'Epanechnikov [Qingchang 07] a été utilisé pour le suivi par l'algorithme Mean Shift [Comaniciu 03c] et le coefficient de Bhattacharyya [Kailath 67] a été utilisé comme mesure de similarité entre l'histogramme du modèle de la cible et celui de la cible candidate. Pour les contraintes du temps réel, nous avons également limité le nombre d'itérations maximum de Mean Shift à $N_{max} = 20$ [Comaniciu 03c].

Le reste de ce chapitre est organisé comme suit : la section 4.2 présente les résultats expérimentaux et les discussions pour la première contribution SOBWH [Laaroussi 16c]. Les expérimentations de la deuxième AMT [Laaroussi 16a] et la troisième CTFWH [Laaroussi 17b] contribution sont présentées dans la section 4.3 et la section 4.4 conclut le chapitre.

4.2 Expérimentation 1

Dans cette section, nous réalisons des évaluations confirmant la robustesse de l'approche proposée SOBWH (Scale Orientation-based Background Weighted Histogram) [Laaroussi 16c] détaillée dans la section 3.2 du chapitre 3 pour le suivi d'une personne basée sur l'histogramme pondéré et corrigé de l'arrière-plan [Ning 10a] et les descripteurs de moments du poids de l'image [Bradski 98, Ning 12c].

4.2.1 Méthodes comparées

Pour évaluer la performance de notre approche SOBWH par rapport aux approches récentes de la littérature, nous avons comparé nos résultats avec les résultats obtenus par trois autres méthodes présentées dans la section 1.6.4 du chapitre 1, à savoir l'algorithme d'échelle adaptative Adaptive Scale [Comaniciu 03c], l'algorithme EM-Shift (Expectation-

Maximisation Shift) [Zivkovic 04a] et l'algorithme SOAMST (Scale and Orientation Adaptive Mean Shift Tracking) [Ning 12c]. En effet, tous ces algorithmes sont basés sur l'utilisation des représentations de la cible par modèle d'apparence, ainsi ils visent à répondre aux changements d'échelle et d'orientation de la cible dans le cadre de l'algorithme Mean Shift [Comaniciu 03c]. Pour une comparaison équitable de toutes ces approches présentées, les histogrammes sont calculés dans l'espace de couleur RGB qui a été choisi comme espace de caractéristique, et ils sont quantifiés en $16 \times 16 \times 16$ classes. Ainsi, nous avons utilisé une ellipse pour représenter la forme géométrique de la cible et un histogramme pondéré et normalisé pour son apparence.

Nous avons évalué quantitativement les performances des approches présentées en utilisant l'erreur de la localisation du centre CLE (Center Location Error). En effet, l'erreur de la localisation du centre pour une image donnée est approximée par la distance entre le centre du résultat de suivi et celui de la vérité du terrain. Pour calculer cette erreur CLE, nous avons utilisé la formule suivante [Senior 06] :

$$CLE = \frac{1}{N} \sum_{\exists i \ g(t_i) \& r(t_i)} \sqrt{(xg_i - xr_i)^2 + (yg_i - yr_i)^2} \quad (4.1)$$

avec N est le nombre total des images de la séquence, (xg_i, yg_i) et (xr_i, yr_i) sont respectivement les positions de la vérité du terrain et celles des résultats du suivi obtenues à l'image i . À noter qu'une méthode optimale devrait avoir une erreur minimale.

4.2.2 Base de données

Pour donner une évaluation objective de l'approche proposée SOBWH, nous avons utilisé quatre séquences de la base de données Benchmark [Benchmark 17] à savoir Human9, Diving, David et Woman. Le nombre total des images évaluées est 1604. Tout l'ensemble des données et la vérité du terrain sont disponibles dans [Benchmark 17]. La catégorisation des séquences utilisées selon dix difficultés principales est présentée dans le tableau 4.1 ci-après. Notez qu'une séquence vidéo peut présenter plus d'une difficulté. L'étape d'initialisation de l'objet à suivre dans la première image de la séquence est effectuée manuellement en marquant une boîte englobante sur celui-ci. Une fois que la cible est localisée, l'algorithme Mean Shift [Comaniciu 03c] (Annexe B) est utilisé pour rechercher itérativement la meilleure région correspondante à l'objet pour chaque prochaine image en utilisant une nouvelle représentation de la cible proposée par notre approche SOBWH.

Séquence	#Images	OCC	IV	SV	DEF	MB	FM	IPR	OPR	DS	CM
Human9	305		✓	✓	✓	✓	✓				✓
Diving	231			✓	✓			✓			✓
David	471	✓	✓	✓	✓	✓		✓	✓		✓
Woman	597	✓	✓	✓	✓	✓	✓		✓	✓	✓

Tableau 4.1 : Difficultés principales qui caractérisent les séquences d'images utilisées. OCC: Occultation, IV: Variation d'illumination, SV: Variation d'échelle, DEF: Déformation, MB: Flou du mouvement, FM: Mouvement rapide, IPR: Rotation dans le plan de l'image, OPR: Rotation hors le plan de l'image, DS: Élément distracteur, CM: Mouvement de la caméra.

4.2.3 Résultats expérimentaux

Les tableaux 4.2, 4.3 et 4.4 ci-après présentent respectivement, l'erreur de la localisation du centre CLE en pixels, le nombre moyen d'itérations et le temps moyen d'exécution (s/image) pour toutes les méthodes présentées, où l'approche proposée SOBWH réalise la meilleure performance dans toutes les séquences utilisées. La police rouge gras dans les tableaux 4.2, 4.3 et 4.4 indique le meilleur résultat obtenu. La performance globale de toutes les approches est résumée par la moyenne des valeurs obtenues dans les tableaux 4.2, 4.3 et 4.4 (les dernières lignes de ces tableaux). En effet, en calculant la moyenne des erreurs de la localisation du centre CLE pour chaque méthode (la dernière ligne du tableau 4.2), nous remarquons que, l'approche proposée SOBWH permet d'obtenir une précision de localisation minimale avec une erreur moins de 16 pixels, suivie par les algorithmes SOAMST, Adaptive Scale et EM-Shift respectivement. Elle a aussi besoin de moins du nombre d'itérations, environ 4.84 itérations par rapport aux autres algorithmes comparés (la dernière ligne du tableau 4.3), ce qui signifie que, la méthode proposée converge plus rapidement. Selon les résultats présentés dans la dernière ligne du tableau 4.4, l'approche SOBWH nécessite moins de temps d'exécution, environ 0.43 s pour traiter une image par rapport aux autres algorithmes comparés.

Pour analyser en profondeur les méthodes comparées sur plusieurs séquences d'images, nous avons également préparé une courbe pour chaque séquence qui montre l'erreur de la localisation du centre CLE par rapport au nombre d'images utilisées. Cette courbe est utile pour mieux comprendre les détails des comportements des méthodes comparées puisque la moyenne des erreurs de localisation du centre CLE résume simplement la performance de l'algorithme considéré sur une séquence donnée. Les résultats du suivi sont présentés par des ellipses rouges pour notre méthode SOBWH et nous avons utilisé les couleurs vert brillant, cyan et jaune pour présenter les résultats du suivi pour les méthodes Adaptive

Séquence	Adaptive Scale	EM-Shift	SOAMST	SOBWH
Human9	60.48	74.70	68.26	11.55
Diving	22.98	52.38	24.91	22.06
David	44.25	65.55	31.58	23.38
Woman	77.57	19.07	15.91	6.12
Moyenne	51.32	52.92	38.58	15.78

Tableau 4.2 : Erreur de la localisation du centre CLE (en pixels) pour notre approche SOBWH et les trois méthodes comparées.

Séquence	Adaptive Scale	EM-Shift	SOAMST	SOBWH
Human9	15	5	6.41	6.90
Diving	15	5	3.94	3.67
David	15	5	4.77	4.56
Woman	15	5	4.41	4.22
Moyenne	15	5	4.88	4.84

Tableau 4.3 : Nombre moyen d'itérations pour notre approche SOBWH et les trois méthodes comparées.

Séquence	Adaptive Scale	EM-Shift	SOAMST	SOBWH
Human9	1.19	0.82	3.26	0.53
Diving	1.06	0.77	1.69	0.63
David	1.73	0.88	1.05	0.36
Woman	1.15	0.88	0.57	0.19
Moyenne	1.28	0.84	1.64	0.43

Tableau 4.4 : Temps moyen d'exécution (s/image) pour notre approche SOBWH et les trois méthodes comparées.

Scale, EM-Shift et SOAMST respectivement.

Dans la première expérience, nous avons utilisé la séquence Human9 avec 305 images de 320×240 pixels et qui présente une personne qui marche sur la route. Cette séquence comporte de nombreux défis cruciaux tels que le changement d'illumination, la diminution progressive de l'échelle, la déformation, le flou du mouvement et le mouvement rapide de la personne et de la caméra. La figure 4.1 ci-après présente certains exemples de résultats qualitatifs du suivi pour les méthodes présentées. En raison du mouvement rapide de la personne et les changements drastiques de l'échelle et de l'arrière-plan, les algorithmes Adaptive Scale, EM-Shift et SOAMST perdent la cible très rapidement après l'image 72 comme le montre les figures 4.1 (a), (b) et (c) respectivement. Néanmoins, notre



Figure 4.1 : Résultats du suivi pour la séquence Human9 et pour les images de gauche à droite 6, 72, 108, 154 et 287 respectivement.

l'approche SOBWH est capable de suivre correctement la personne dans toute la séquence (Figure 4.1 (d)), puisque le modèle de représentation utilisé exploite les caractéristiques de l'arrière-plan pour diminuer l'interférence de ces caractéristiques dans la localisation de la cible. Pour l'évaluation quantitative, la première ligne des tableaux 4.2, 4.3 et 4.4 font la synthèse des résultats obtenus pour les différentes méthodes considérées. D'où, l'approche proposée SOBWH permet d'obtenir une meilleure précision de localisation de la cible avec une erreur inférieure à 12 pixels par rapport aux autres algorithmes comparés qui dépassent 60 pixels comme erreur de la localisation du centre. L'approche proposée a aussi besoin de moins du nombre d'itérations par rapport aux algorithmes Adaptive Scale et EM-Shift (tableau 4.3), ce qui signifie que notre approche SOBWH converge plus rapidement

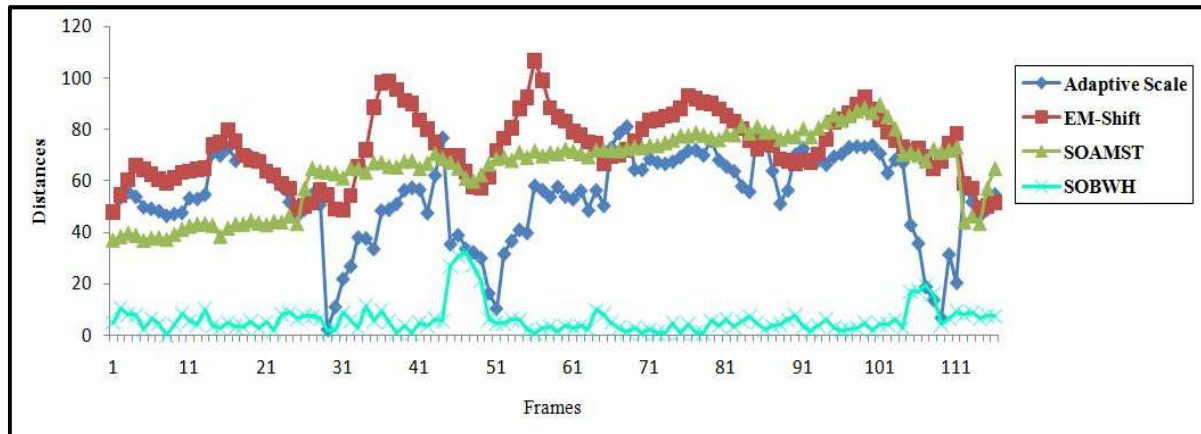


Figure 4.2 : Erreur de la localisation du centre CLE (en pixels) par image pour la séquence Human9.

et nécessite moins de temps d'exécution, environ 0.53 s par image (tableau 4.4). La courbe de l'erreur de la localisation du centre est présentée sur la figure 4.2 ci-dessus. En effet, ces résultats expérimentaux montrent que l'approche proposée SOBWH permet de diminuer l'interférence des caractéristiques de l'arrière-plan dans la localisation de la cible et aide à faire la discrimination entre la cible et son arrière-plan, ce qui permet d'améliorer la qualité de suivi en présence d'éléments perturbateurs de la scène et aux changements d'échelle par rapport aux autres algorithmes comparés à savoir Adaptive Scale, EM-Shift et SOAMST.

La deuxième expérience est la séquence Diving avec 231 images de 400×224 pixels. Les difficultés de la séquence sont nombreuses à savoir la capacité à prendre en compte les mouvements rapides de la personne tout autant que ceux de la caméra, les différents points de vue, les changements de l'échelle et les rotations de la cible dans le plan de l'image (tableau 4.1). Certains exemples des résultats qualitatifs du suivi pour toutes les méthodes considérées sont présentés sur la figure 4.3 ci-après. À cause du changement d'échelle, les rotations et les déformations importantes de la cible, les deux algorithmes Adaptive Scale et EM-Shift perdent la cible rapidement à partir de l'image 168 comme illustre la figure 4.3 (a) et (b) ci-après. Néanmoins, les deux approches SOAMST et SOBWH arrivent à suivre correctement la cible dans toute la séquence selon la figure 4.3 (c) et (d) ci-après malgré toutes les difficultés de la séquence. En effet, l'approche proposée SOBWH estime avec plus de précision l'échelle de la cible même si celle-ci se déplace à des allures différentes au cours de la séquence (figure 4.3 (d)). Faire référence aux résultats statistiques présentés dans les tableaux 4.2, 4.3 et 4.4 respectivement, l'approche proposée SOBWH permet d'obtenir une meilleure précision de la localisation de la cible avec une erreur moins de 22 pixels par rapport aux autres algorithmes Adaptive Scale, EM-Shift et SOAMST.

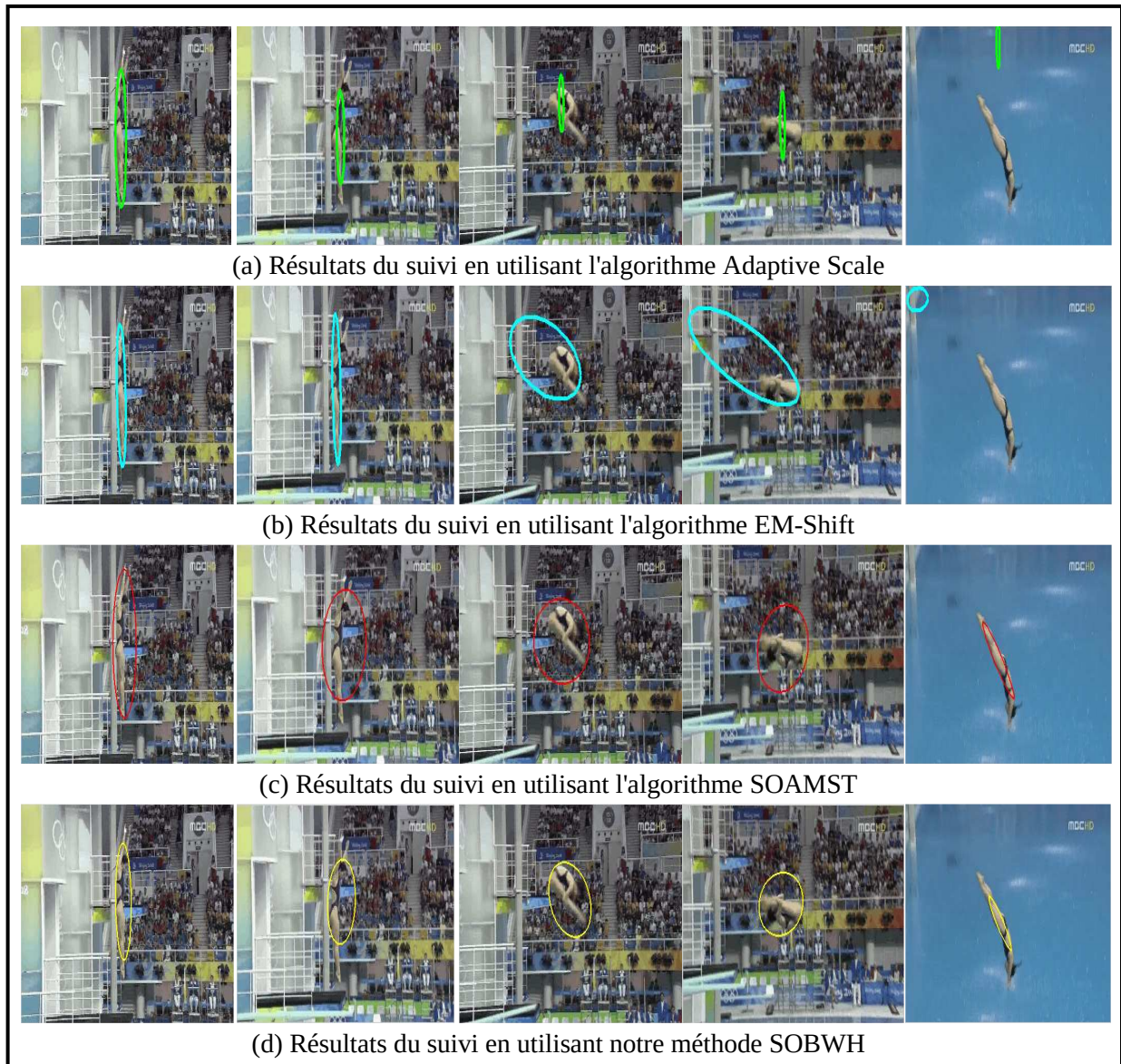


Figure 4.3 : Résultats du suivi pour la séquence Diving et pour les images de gauche à droite 2, 26, 120, 168 et 214 respectivement.

L'approche proposée a aussi besoin de moins du nombre d'itérations, environ 3.67 par rapport aux algorithmes Adaptive Scale, EM-Shift et SOAMST, ce qui signifie qu'elle converge plus rapidement et nécessite moins de temps d'exécution 0.63 s pour traiter une image. La courbe de l'erreur de la localisation du centre par rapport au nombre d'images utilisée est présentée sur la figure 4.4 ci-après. D'où, ces résultats expérimentaux valident la robustesse de la méthode proposée SOBWH, tout en ayant un modèle d'apparence invariant aux rotations, aux changements d'échelle et il est capable de faire la discrimination entre la cible et son arrière-plan.

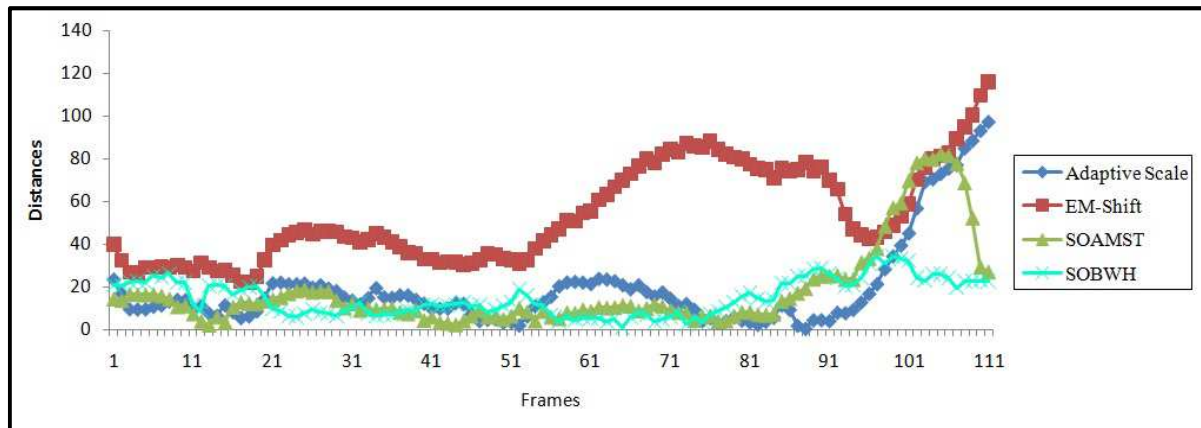


Figure 4.4 : Erreur de la localisation du centre CLE (en pixels) par image pour la séquence Diving.

La troisième expérience est la séquence David avec 471 images de 320×240 pixels. La séquence présente une personne qui se déplace librement dans une scène et la cible à suivre est le visage de cette personne. Cette séquence comporte de nombreux défis cruciaux tels que les changements importants d'illumination et d'échelle, l'occultation, la déformation, le flou du mouvement et les rotations de la cible dans et hors le plan de l'image (Tableau 4.1). Certains exemples des résultats qualitatifs du suivi sont présentés sur la figure 4.5 ci-après. À cause du mouvement rapide de la personne et les changements importants de l'arrière-plan et de l'illumination, ni l'algorithme Adaptive Scale, ni l'algorithme EM-Shift permet d'obtenir de bons résultats pour le suivi de cette personne comme illustre la figure 4.5 (a) et (b) et la figure 4.6 ci-après. Cependant, la méthode proposée SOBWH arrive à suivre correctement la cible tout au long de la séquence et estime avec plus de précision les changements de l'échelle par rapport à la méthode SOAMST selon la figure 4.5 (d) et (c) ci-après. Les résultats statistiques présentés dans les tableaux 4.2, 4.3 et 4.4 ci-avant montrent que l'approche SOBWH permet d'obtenir une meilleure précision de la localisation du centre avec une erreur inférieure à 24 pixels par rapport aux autres algorithmes qui dépassent 32 pixels comme erreur de la localisation du centre. L'approche proposée a besoin de moins du nombre d'itérations, environ 4.56 par rapport aux algorithmes Adaptive Scale et EM-Shift, ce qui signifie que, notre approche SOBWH converge plus rapidement et nécessite moins de temps d'exécution, environ 0.36 s par image. La courbe de l'erreur de la localisation du centre CLE par image est présentée sur la figure 4.6 ci-dessous. D'après ces résultats expérimentaux, l'approche proposée SOBWH est moins sensible aux changements importants de l'illumination et de l'arrière-plan par rapport aux algorithmes Adaptive Scale, EM-Shift et SOAMST.



Figure 4.5 : Résultats du suivi pour la séquence David et pour les images de gauche à droite 2, 9, 296, 439 et 471 respectivement.

Dans la quatrième expérience, nous avons utilisé une longue séquence Woman de 597 images et de 352×288 pixels. La cible à suivre est la femme qui marche librement sur la route où son échelle augmente progressivement à la fin de la séquence. En effet, la cible subit de nombreuses difficultés à savoir les occultations partielles plusieurs fois tout en se déplaçant derrière les structures de l'arrière-plan, les problèmes de déformation, le flou du mouvement, les rotations hors le plan de l'image, la présence d'éléments distrayeurs causée par l'arrière-plan, le mouvement de la caméra, ainsi que la femme se déplace rapidement dans la scène. Certains exemples de résultats finaux du suivi sont présentés sur la figure 4.7 ci-après. La courbe de l'erreur de la localisation du centre CLE par rapport au nombre d'images utilisées est présentée sur la figure 4.8 ci-après.

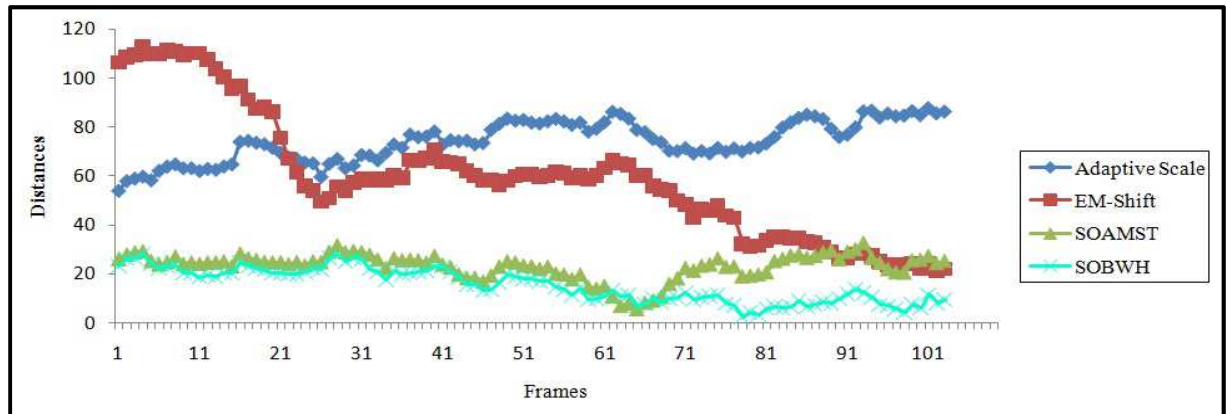


Figure 4.6 : Erreur de la localisation du centre CLE (en pixels) par image pour la séquence David.

Pour la séquence Woman, la femme suivie est partiellement occultée par une voiture dont la couleur est similaire à celle de sa chemise. Comme le montre les figures 4.7 (a), (b) et (c) ci-après et la courbe de l'erreur de la localisation du centre sur la figure 4.8 ci-après, les trois algorithmes comparés Adaptive Scale, EM-Shift et SOMAST dérivent la cible à partir de l'image 135 (figure 4.7 (a), (b) et (c)) et après avoir capturé transitoirement la cible, ils ne la récupèrent pas à la fin de la séquence, qui est principalement dû à la contamination du modèle d'apparence utilisé par des caractéristiques appartenant à l'objet occulté. En effet, l'algorithme Adaptive Scale basé sur l'incrémentation selon trois niveaux différents de l'échelle ne fonctionne pas bien dans le cas des changements brusques de la taille de l'objet. D'autre part, les algorithmes EM-Shift et SOMAST ignorent l'information de l'arrière-plan qui aide à faire la discrimination entre la cible et son arrière-plan. Par conséquent, l'approche proposée SOBWH arrive à suivre correctement la femme tout au long de la séquence selon la figure 4.7 (d) et la figure 4.8 ci-après. Les tableaux 4.2, 4.3 et 4.4 montrent que l'approche proposée SOBWH permet d'obtenir une meilleure précision de localisation de la cible par rapport aux autres algorithmes avec une moyenne des erreurs inférieure à 7 pixels. L'approche proposée a aussi besoin de moins du nombre moyen d'itérations environ 4.22 itérations par rapport aux algorithmes Adaptive Scale, EM-Shift et SOMAST, ce qui signifie que notre approche SOBWH converge plus rapidement et nécessite moins de temps d'exécution, environ 0.19 s par image.

D'après ces résultats expérimentaux, l'approche proposée SOBWH permet de réduire l'interférence des caractéristiques de l'arrière-plan dans la localisation de la cible, tout en ayant un modèle d'apparence invariant aux occultations partielles, aux changements d'échelle et d'orientation, aux mouvements de la caméra et finalement il est capable de faire la discrimination entre la cible et son arrière-plan.



Figure 4.7 : Résultats du suivi pour la séquence Woman et pour les images de gauche à droite 8, 135, 194, 369 et 587 respectivement.

4.2.4 Discussion

Nos expérimentations qualitatives et quantitatives sur les quatre séquences utilisées montrent que l'exploitation efficace de l'information de l'arrière-plan avec les descripteurs de moments du poids de l'image [Bradski 98] sous l'algorithme Mean Shift [Comaniciu 03c] permettent de suivre correctement la cible dans toutes les séquences. En effet, en apprenant l'arrière-plan avec un mécanisme de pondération par l'utilisation de l'histogramme pondéré et corrigé de l'arrière-plan, nous pouvons masquer les valeurs d'intensités reliées à l'arrière-plan et aider à faire la discrimination entre la cible et son arrière-plan, ce qui permet d'améliorer l'estimation de la position, l'échelle et l'orientation de la cible. Ainsi, l'approche proposée valide sa robustesse à travers des scénarios concrets du suivi à savoir le changement

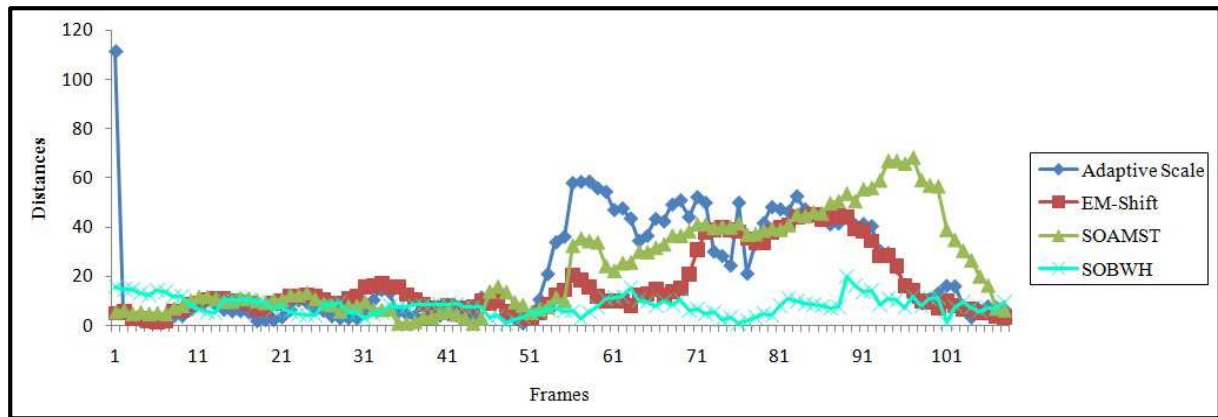


Figure 4.8 : Erreur de la localisation du centre CLE (en pixels) par image pour la séquence Woman.

de l'arrière-plan, l'occultation partielle, les changements d'échelle et d'orientation, la distinction entre l'objet et son arrière-plan, le mouvement de la cible, l'apparition d'éléments distracteurs, le changement d'illumination, les rotations et les déformations. En résumé, nous pouvons conclure que notre méthode SOBWH donne de résultats satisfaisants par rapport aux résultats obtenus par rapport à l'algorithme SOAMST [Ning 12c] et de bons résultats par rapport aux résultats obtenus par les algorithmes Scale Adaptive Mean Shift [Comaniciu 03c] et EM-Shift [Zivkovic 04a].

4.3 Expérimentation 2

Dans cette section, nous réalisons des tests confirmant la robustesse de nos approches proposées telles que l'approche AMT (Apperance Model Tracking) [Laaroussi 16a] et l'approche CTFWH (Color-Texture Foreground-Weighted Histogram) [Laaroussi 17b] pour le suivi d'une personne dans une séquence d'images. Ces deux approches ont été détaillé dans les sections 3.3 et 3.4 du chapitre 3, respectivement.

4.3.1 Méthodes comparées

Nous avons évalué nos approches proposées par une comparaison avec huit algorithmes récents de la littérature. Parmi les algorithmes de suivi comparés, deux sont les mêmes existant dans la littérature et qui utilisent un modèle d'apparence basé sur l'algorithme Mean Shift [Comaniciu 03c] déjà discuté dans la section 1.6.5 du chapitre 1, à savoir l'histogramme pondéré de l'arrière-plan BWH (Background-Weighted Histogram) [Comaniciu 03c] et l'algorithme saliance du premier-plan basé sur l'histogramme pondéré de l'arrière-plan FSBWH (Foreground Saliency-based Background-Weighted Histogram) [Wang 16c]. Les six autres algorithmes sont Adaptive Structural Tracker (AST) [Jia 12], Co-training Multiple

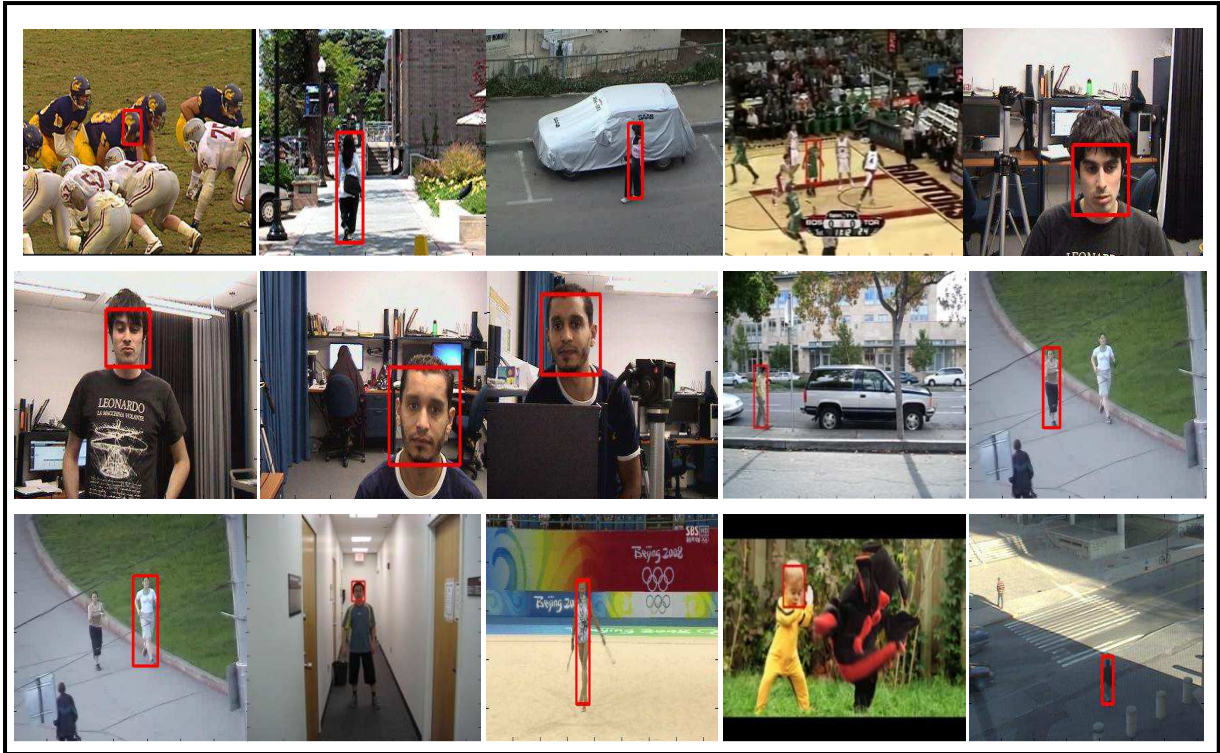


Figure 4.9 : Annotation de la première image des séquences vidéo utilisées pour l'expérimentation. De gauche à droite, de haut en bas: Football1, Human7, Woman, Basketball, Jp1, Jp2, Wbook, Wdesk, David3, Jogging1, Jogging2, Boy, Gym, DragonBaby et Crossing.

Instance Learning algorithm (CoMIL) [Lu 11], Sparsity-based Collaborative Model Tracker (SCMT) [Zhong 14], Online Multiple Support Instance Tracker (OMSIT) [Zhou 11b], Spatio-Temporal Context (STC) [Zhang 10a] et le huitième algorithme est Weighted Local Cosine Similarity (WLCS) [Wang 15b]. Afin d'assurer une comparaison équitable, les histogrammes sont calculés dans l'espace de couleur RGB et sont quantifiés en $16 \times 16 \times 16 \times 5$ classes pour les algorithmes AMT et CTFWH, tandis que, pour les algorithmes BWH et FSBWH, les histogrammes sont quantifiés en $16 \times 16 \times 16$ classes.

4.3.2 Base de données

Nous évaluons les approches comparées sur quinze séquences d'images différentes. Onze d'entre eux proviennent de suivi d'objets Benchmark couramment utilisé par la communauté Visual Tracker Benchmark [Wu 15]. Les séquences d'images et leurs vérités du terrain sont disponibles dans [Benchmark 17]. Les quatre autres séquences Jp1, Jp2, Wdesk et Wbook proviennent de la base de données LITIV utilisés par Bouachir et al. [Bouachir 15] et sont disponibles avec leurs vérités du terrain dans [LITIV 17]. La zone a été encombrée par des bureaux, des chaises et des équipements techniques des vidéos dans l'arrière-plan. La cible a été initialisée dans la première image en utilisant la position de la vérité du terrain pour tous

Séquence	#Images	OCC	IV	SV	DEF	MB	FM	OV	IPR	OPR	BC	DS	CM
Football1	81	✓							✓	✓	✓	✓	✓
Human7	250	✓	✓	✓	✓	✓	✓						✓
Woman	597	✓	✓	✓	✓	✓	✓			✓		✓	✓
Basketball	725	✓	✓		✓					✓	✓	✓	✓
jp1	609	✓										✓	✓
jp2	229	✓										✓	✓
Wbook	581	✓											✓
Wdesk	709	✓											✓
David3	252	✓			✓					✓	✓		✓
Jogging1	307	✓			✓					✓			✓
Jogging2	307	✓			✓					✓			✓
Boy	602			✓		✓	✓		✓	✓			✓
Gym	767			✓	✓				✓	✓			✓
DragonBaby	113	✓		✓		✓	✓	✓	✓	✓			✓
Crossing	120			✓	✓		✓			✓	✓		

Tableau 4.5 : Difficultés principales qui caractérisent les séquences d'images utilisées. OCC: Occultation, IV: Variation d'illumination, SV: Variation d'échelle, DEF: Déformation, MB: Flou du mouvement, FM: Mouvement rapide, OV: Sortie du champ visuel, IPR: Rotation dans le plan de l'image, OPR: Rotation hors le plan de l'image, BC: Encombrement de l'arrière-plan, DS: Élément distracteur, CM: Mouvement de la caméra.

les algorithmes comparés. Le nombre total des images évaluées est 6249. Dans ce travail, l'initialisation du suivi est définie en marquant manuellement une boîte englobante dans la première image. La figure 4.9 ci-avant présente la première image de chaque séquence. Notez que, nous avons annoté deux cibles pour la séquence Jogging. Afin de mieux comprendre les résultats quantitatifs de nos algorithmes proposés, nous avons catégorisé les séquences d'images selon les principales difficultés qui peuvent survenir dans chaque séquence. La catégorisation des séquences selon douze propriétés principales est présentée dans le tableau 4.5 ci-dessus. Cela nous permet de construire des sous-groupes de vidéos afin d'évaluer quantitativement les algorithmes de suivi dans plusieurs situations. Notez qu'une séquence d'images peut présenter plus d'une difficulté.

4.3.3 Méthodes d'évaluation

Nous avons évalué quantitativement les performances de nos approches proposées AMT et CTFWH en utilisant le taux de la réussite SR (Success Rate) et l'erreur de la localisation du centre CLE (Center Location Error). Pour mesurer le taux de la réussite SR, nous avons

Séquence	BWH	FSBWH	AST	CoMIL	SCMT	OMSIT	STC	WLCS	AMT	CTFWH
Football1	83.78	33.78	35.14	6.76	40.54	13.51	33.78	21.62	85.14	86.49
Human7	26.40	15.60	30.20	9.60	40.20	9.60	33.20	21.60	26.00	40.26
Woman	12.06	68.01	17.92	2.35	82.08	2.18	19.26	7.37	13.57	85.26
Basketball	68.97	87.45	78.62	1.79	41.38	1.38	20.69	0.83	35.03	90.24
Jp1	72.86	100.00	100.00	14.14	100.00	8.88	68.42	48.68	94.41	100.00
Jp2	75.11	69.87	55.46	30.13	55.46	15.72	91.27	55.02	87.04	87.12
Wbook	91.22	57.83	100.00	19.10	100.00	20.83	30.29	86.44	78.83	87.73
Wdesk	72.78	83.92	100.00	8.04	100.00	5.36	33.57	6.77	79.41	100.00
David3	55.56	67.86	33.73	1.98	32.14	1.98	32.94	30.95	71.43	91.32
Jogging1	36.81	22.48	22.48	20.52	20.20	21.82	20.85	6.84	41.69	96.46
Jogging2	14.66	96.74	17.92	6.84	100.00	6.19	16.94	16.61	94.14	100.00
Boy	98.50	98.34	65.61	14.78	64.62	15.95	65.12	43.85	56.98	100.00
Gym	33.64	28.81	40.17	1.56	4.95	2.35	20.21	1.56	43.02	75.76
DragonBaby	83.19	89.38	26.55	8.85	19.47	14.16	21.24	23.89	42.48	70.18
Crossing	48.33	87.17	99.17	65.83	100.00	30.00	18.33	100.00	78.33	89.17
Moyenne	58.26	67.15	56.13	14.15	60.14	11.33	35.07	31.47	61.85	86.67

Tableau 4.6 : Pourcentage des images correctement suivies (le taux de la réussite SR) pour nos approches AMT, CTFWH et les huit méthodes comparées.

calculé le taux de chevauchement OR (Overlap Ratio) [Wu 15] pour chaque image par la formule suivante :

$$OR = \frac{area(P_r \cap G_r)}{area(P_r \cup G_r)} \quad (4.2)$$

avec P_r est la région de la cible prédite et G_r est la région cible de la vérité du terrain, $\cap$ et $\cup$ représentent respectivement l'intersection et l'union des deux régions. Le suivi de la cible est considéré comme un succès pour une image donnée, si le taux de chevauchement $OR \geq 0.5$. Cependant, l'erreur de la localisation du centre CLE pour une image donnée est approximée par la distance entre le centre du résultat de suivi et celui de la vérité du terrain. Pour calculer l'erreur de la localisation du centre CLE, nous avons utilisé la même formule de l'équation (4.1) de la section 4.2.1.

4.3.4 Résultats expérimentaux

Les tableaux 4.6 ci-dessus et 4.7 ci-après présentent respectivement le taux de la réussite SR et l'erreur de localisation du centre CLE pour toutes les méthodes comparées, où les

Séquence	BWH	FSBWH	AST	CoMIL	SCMT	OMSIT	STC	WLCS	AMT	CTFWH
Football1	7.12	52.05	50.89	49.64	12.98	37.07	73.97	49.81	7.13	6.79
Human7	49.03	24.90	43.55	44.00	44.52	41.54	32.88	58.25	53.52	19.69
Woman	108.04	14.92	160.06	163.16	10.88	168.81	26.43	144.07	116.55	9.36
Basketball	58.40	17.78	11.18	98.37	123.31	122.77	75.90	189.13	112.56	7.98
Jp1	12.44	10.88	3.32	115.81	3.40	115.05	2.87	55.66	8.68	3.40
Jp2	16.93	37.31	53.72	64.57	42.24	77.30	8.34	23.44	11.30	10.12
Wbook	16.22	26.25	8.45	80.38	5.29	79.87	88.51	17.53	19.17	16.08
Wdesk	17.42	14.90	5.02	89.43	6.02	80.85	76.33	129.42	14.47	8.60
David3	55.86	84.13	103.86	338.95	72.87	335.41	30.88	86.83	21.12	12.72
Jogging1	50.37	91.30	107.79	70.10	167.60	76.07	150.25	109.94	60.30	13.79
Jogging2	87.95	5.89	169.92	36.39	3.71	28.93	161.21	150.33	6.90	5.05
Boy	3.59	3.49	69.94	101.28	69.34	89.38	25.75	56.04	48.48	3.13
Gym	9.92	15.34	13.25	103.62	39.82	100.96	26.23	63.12	8.17	7.99
DragonBaby	12.33	8.17	59.67	82.66	71.56	81.51	248.59	52.44	37.52	20.37
Crossing	12.45	4.47	4.59	4.89	1.58	59.24	32.72	1.68	6.79	4.19
Moyenne	34.54	27.12	57.68	96.22	45.01	99.65	70.72	79.18	35.51	9.95

Tableau 4.7 : Erreur de la localisation du centre CLE (en pixels) pour nos approches AMT, CTFWH et les huit méthodes comparées.

méthodes proposées AMT et CTFWH réalisent la première, la deuxième ou la troisième meilleure performance dans toutes les séquences utilisées. Dans les tableaux 4.6 et 4.7 la police rouge gras indique le meilleur résultat, la police bleue gras indique le deuxième meilleur résultat et la police verte gras indique le troisième meilleur résultat. La performance globale pour tous les algorithmes comparés est résumée par la moyenne des valeurs dans les dernières lignes des tableaux 4.6 et 4.7. Toutes les mesures utilisées pour l'évaluation globale de la performance montrent que notre méthode CTFWH dépasse tous les algorithmes comparés, en atteignant un taux moyen de réussite de 86.67% et une erreur de la localisation du centre CLE inférieure à 10 pixels, suivie par l'algorithme FSBWH [Wang 16c] qui a réalisé la deuxième meilleure performance avec un taux de réussite de 67.15% et une erreur de la localisation du centre de 27.12 pixels. Ainsi que, notre approche AMT [Laaroussi 16a] réalise la troisième meilleure performance globale avec un taux moyen de réussite de 61.85% et la moyenne des erreurs de la localisation du centre de 35.51 pixels. Alors que l'algorithme OMSIT [Zhou 11b] a eu le dernier rang, le classement des autres algorithmes est différent selon la métrique considérée.

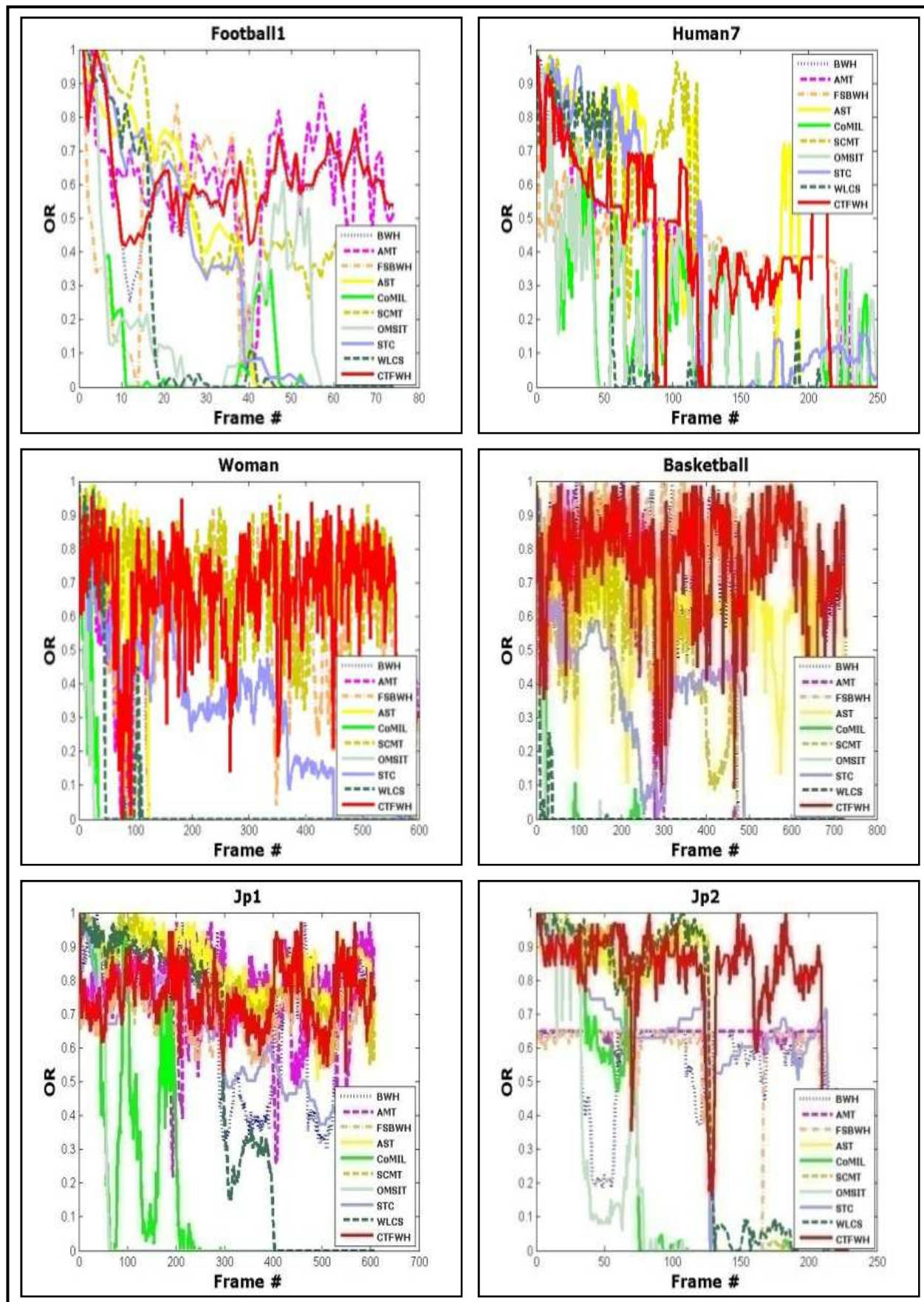


Figure 4.10 : Taux de chevauchement (OR) pour les séquences d'images Football1, Human7, Woman, Basketball, Jp1 et Jp2 respectivement.

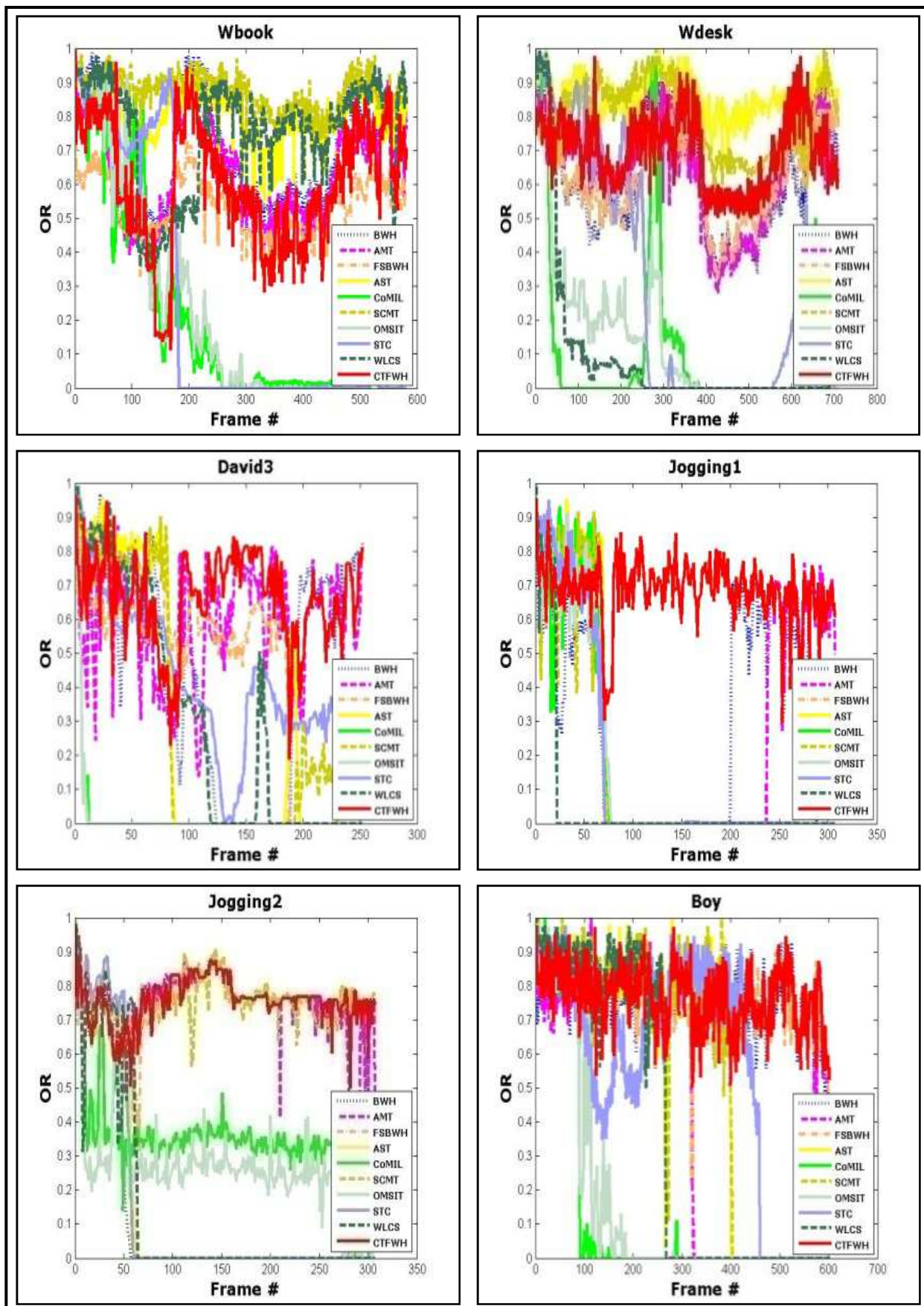


Figure 4.11 : Taux de chevauchement (OR) pour les séquences d'images Wbook, Wdesk, David3, Jogging1, Jogging2 et Boy respectivement.

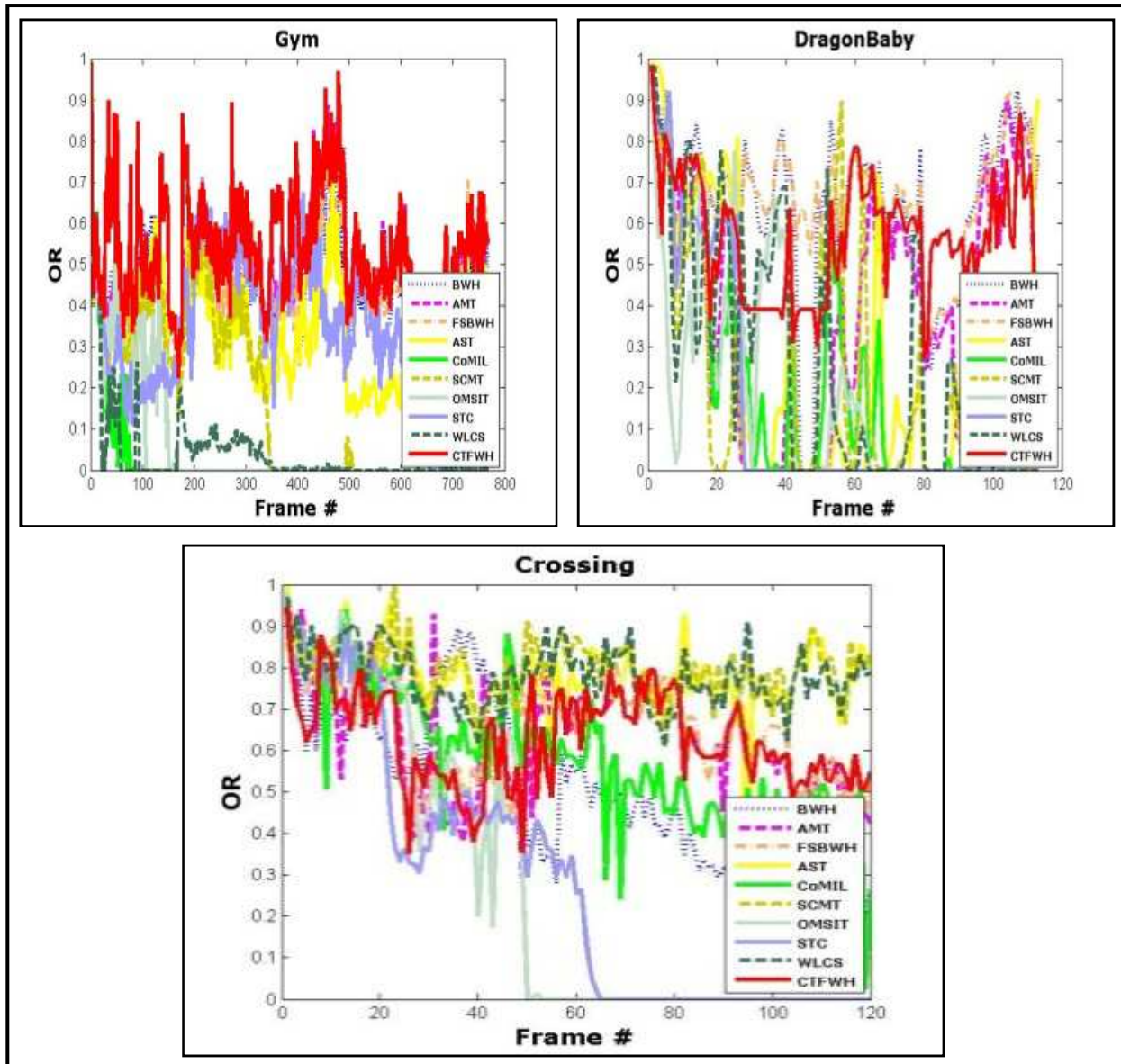


Figure 4.12 : Taux de chevauchement (OR) pour les séquences d'images Gym, DragonBaby et Crossing respectivement.

Afin d'analyser en profondeur les méthodes comparées sur plusieurs séquences d'images, nous avons également préparé deux courbes pour chaque séquence d'images. D'une part, la première courbe présente le taux de chevauchement OR par rapport au nombre d'images de la séquence. Les courbes de taux de chevauchement OR pour toutes les séquences d'images utilisées sont présentées sur les figures 4.10, 4.11 et 4.12 ci-avant. D'autre part, la deuxième courbe présente l'erreur de la localisation du centre CLE par rapport au nombre d'images de la séquence. Les courbes de l'erreur de la localisation du centre CLE pour toutes les séquences d'images utilisées sont présentées sur les figures 4.13, 4.14 et 4.15 ci-après. Ces courbes sont utiles pour mieux comprendre les détails des comportements des méthodes comparées puisque le taux de chevauchement OR et l'erreur de localisation du centre CLE résument simplement la performance de l'algorithme considéré sur une séquence donnée.

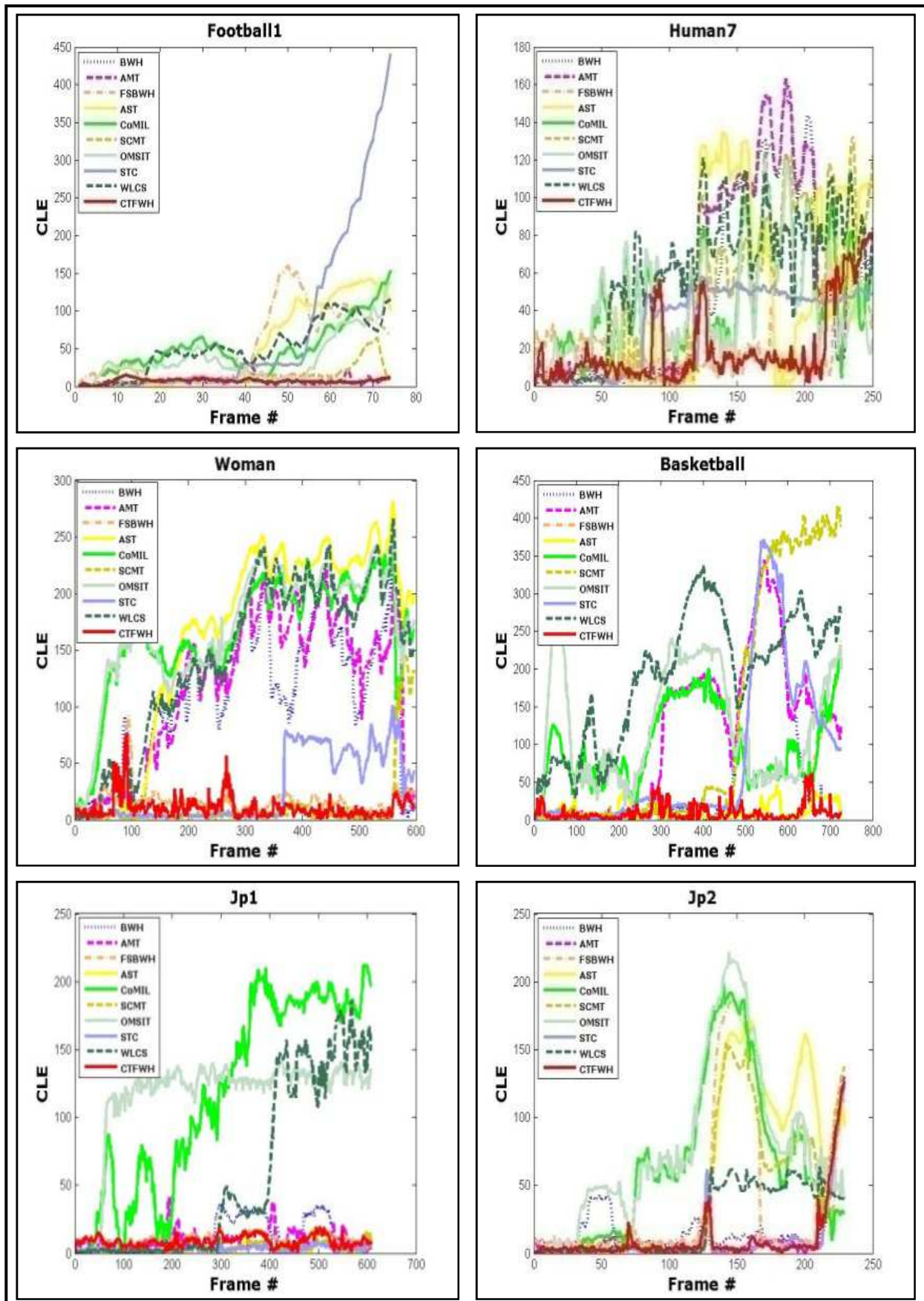


Figure 4.13 : Erreur de la localisation du centre CLE (en pixels) pour les séquences d'images Football1, Human7, Woman, Basketball, Jp1 et Jp2 respectivement.

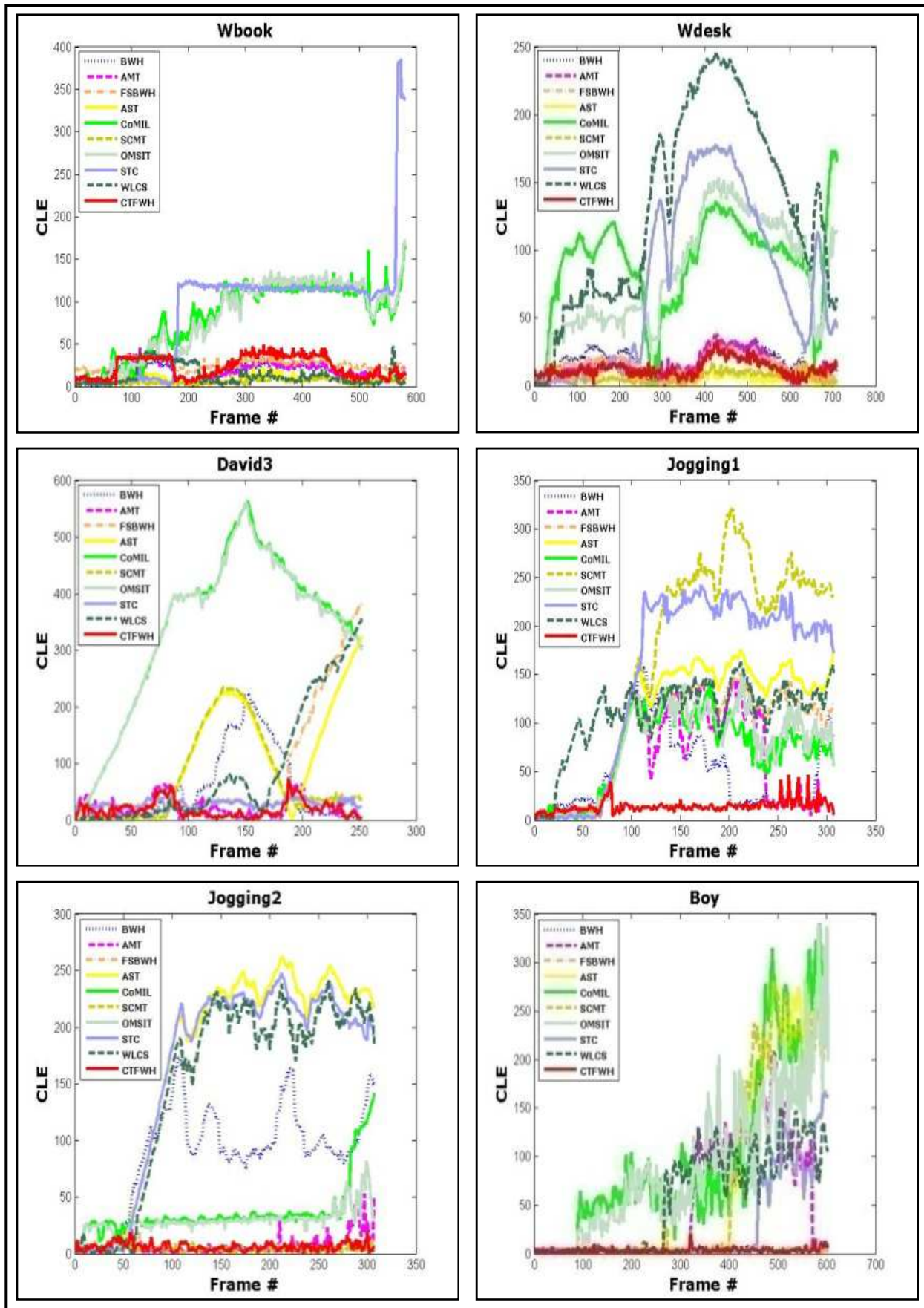


Figure 4.14 : Erreur de la localisation du centre CLE (en pixels) pour les séquences d'images Wbook, Wdesk, David3, Jogging1, Jogging2 et Boy respectivement.

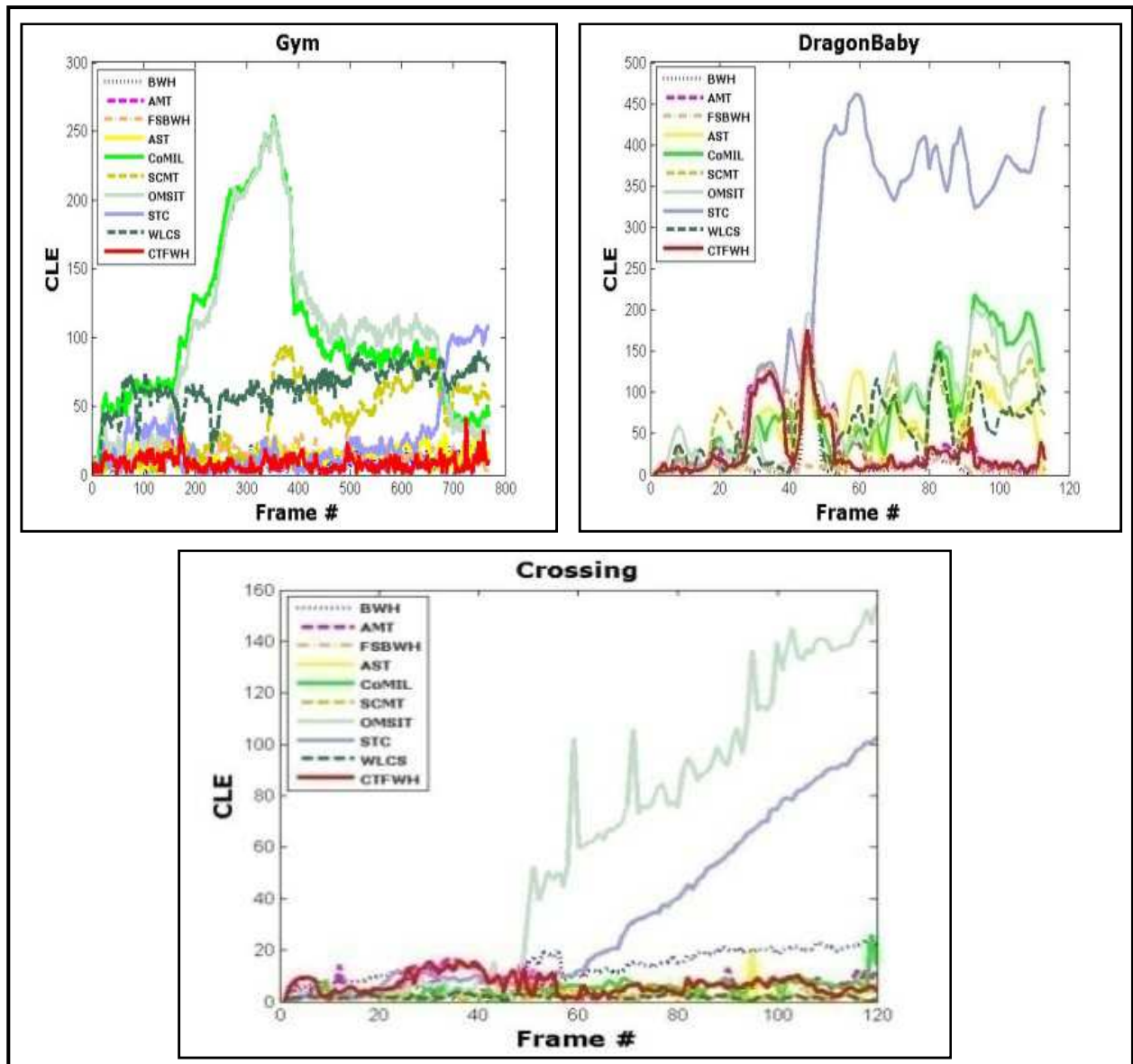


Figure 4.15 : Erreur de la localisation du centre CLE (en pixels) pour les séquences d'images Gym, DragonBaby et Crossing respectivement.

Les résultats du suivi sont présentés par des rectangles rouges pour toutes les méthodes présentées. Dans la suite, les résultats expérimentaux sont organisés et analysés par catégories selon le tableau 4.5 ci-avant.

- **Occultation** : L'occultation est l'un des facteurs critiques affectant la performance d'un algorithme de suivi. Cependant, de nombreuses méthodes de la littérature [Comaniciu 03c, Jia 12, Lu 11, Wang 15b, Wang 16c, Zhang 13b, Zhong 14, Zhou 11b], ainsi les approches proposées AMT [Laaroussi 16a] et CTFWH [Laaroussi 17b] sont développées pour surmonter ce problème. Nous avons évalué les dix méthodes comparées lors de l'occultation sur onze séquences d'images : Human7, Woman, Basketball, Jp1, Jp2, Wbook, Wdesk, David3, Jogging1, Jogging2 et DragonBaby.

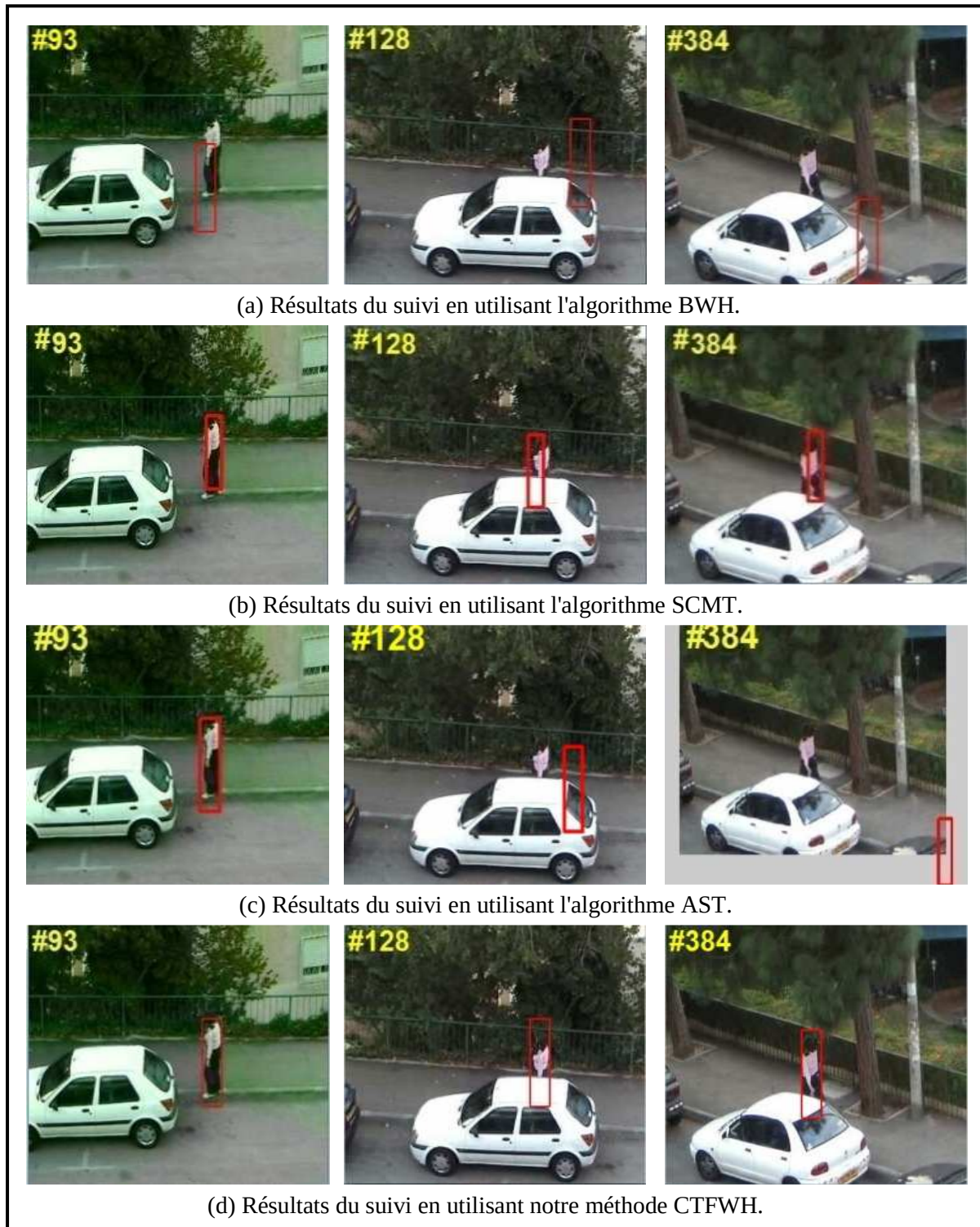


Figure 4.16 : Résultats du suivi pour la séquence d'images Woman avec occultation.

La figure 4.16 ci-avant montre certains exemples de résultats finaux du suivi dans le cas d'occultation pour la séquence Woman. La cible suivie est partiellement occultée par une voiture dont la couleur est similaire à celle de sa chemise. Comme le montre la figure 4.16 (a) et (c) ci-avant et la courbe du taux de chevauchement OR et celle de l'erreur de la localisation du centre CLE sur les figures 4.10 et 4.13, respectivement. La plupart des algorithmes

comparés BWH, AMT, AST, CoMIL, OMSIT, STC, WLCS dérivent la cible à partir de l'image 128 selon la figure 4.16 (a) et (c) et après avoir capturé transitoirement la cible, ils ne la récupèrent pas à la fin de la séquence, ce qui est principalement dû à la contamination du modèle d'apparence utilisé par des caractéristiques appartenant à l'objet occulté. Cependant, les algorithmes FSBWH et SCMT ainsi que l'approche proposée CTFWH fonctionnent mieux que les autres algorithmes et peuvent suivre correctement la cible dans toute la séquence comme le montre la figure 4.16 (b) et (d) ci-avant. L'algorithme FSBWH intègre à la fois la saillance de l'arrière-plan et du premier-plan pour améliorer l'approximation du modèle de la cible. Donc il est plus discriminatif et robuste à l'occultation partielle. D'autre part, la méthode SCMT est également conçue pour traiter les occultations à travers un schéma en considérant les patches sans occultation et une méthode basée sur le vote qui prédit le centre de la cible. Pour cette raison, elle est spécifiquement conçue pour gérer les occultations partielles grâce à son modèle d'apparence basé sur les caractéristiques du premier-plan. Notre méthode CTFWH arrive à suivre la cible correctement dans toute la séquence d'images et elle a obtenu le meilleur taux de réussite de 85.26% selon le tableau 4.6 avec une erreur de la localisation du centre CLE inférieure à 10 pixels.

Dans la séquence Wdesk (Figure 4.17 ci-après), la cible à suivre est le visage d'une personne qui se déplace librement dans une scène. En effet, la cible subit de nombreuses occultations partielles plusieurs fois tout en se déplaçant derrière les structures de l'arrière-plan. Les figures 4.11, 4.14 et 4.17 montrent que ces algorithmes CoMIL, OMSIT, STC et WLCS échouent à suivre le visage correctement. Néanmoins, les algorithmes BWH, FSBWH, AST, SCMT et nos approches AMT et CTFWH arrivent à suivre correctement la cible dans toute la séquence, malgré que la personne suivie effectue de grands déplacements et se cache derrière un bureau. Faire référence aux résultats statistiques présentés dans les tableaux 4.6 et 4.7 pour la séquence Wdesk, tous ces algorithmes AST, SCMT et CTFWH réalisent le meilleur taux de réussite de 100% avec une erreur de la localisation du centre CLE inférieure à 9 pixels.

Dans la séquence Jogging2 (Figure 4.18 ci-après), tous les algorithmes, à l'exception FSBWH, SCMT et nos algorithmes AMT et CTFWH, perdent la cible rapidement en raison d'une forte occultation entre la personne suivie et un objet de la scène. Concernant les résultats statistiques présentés dans les tableaux 4.6 et 4.7 pour la séquence Jogging2, les deux algorithmes SCMT et CTFWH réalisent le meilleur taux de réussite de 100% avec une erreur de la localisation du centre inférieure à 6 pixels, dès qu'une petite partie de la cible devient

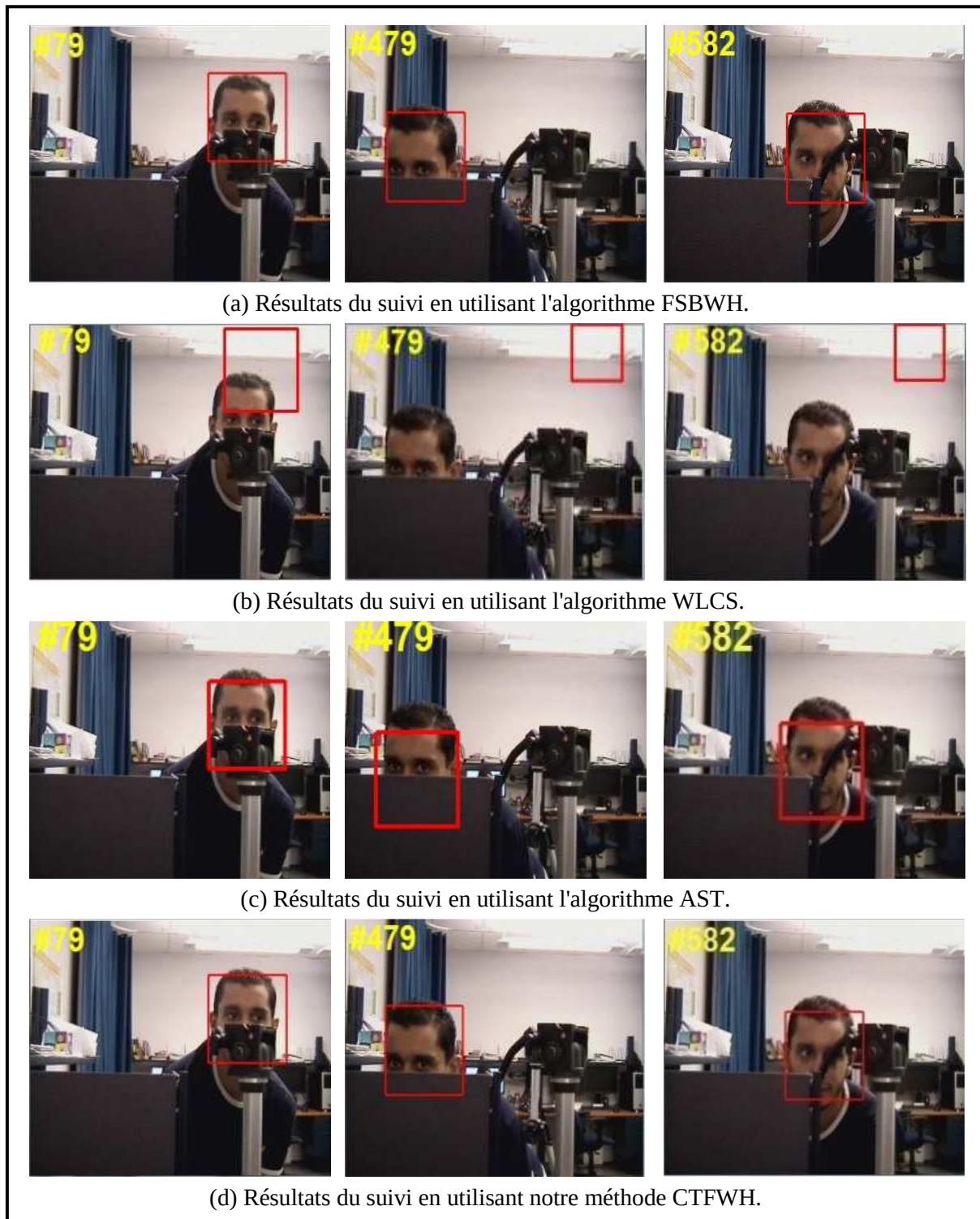


Figure 4.17 : Résultats du suivi pour la séquence d'images Wdesk avec occultation.

visible (image 53 de la figure 4.18 (a)). En résumé, les résultats satisfaisants obtenus par notre méthode CTFWH soulignent l'importance d'utiliser la jointure des caractéristiques couleur-texture afin de limiter les ambiguïtés liées à la représentation par histogramme de couleurs. De plus, les caractéristiques du premier-plan sont plus invariantes que les autres caractéristiques utilisées par les méthodes comparées lorsqu'une telle situation se produite.

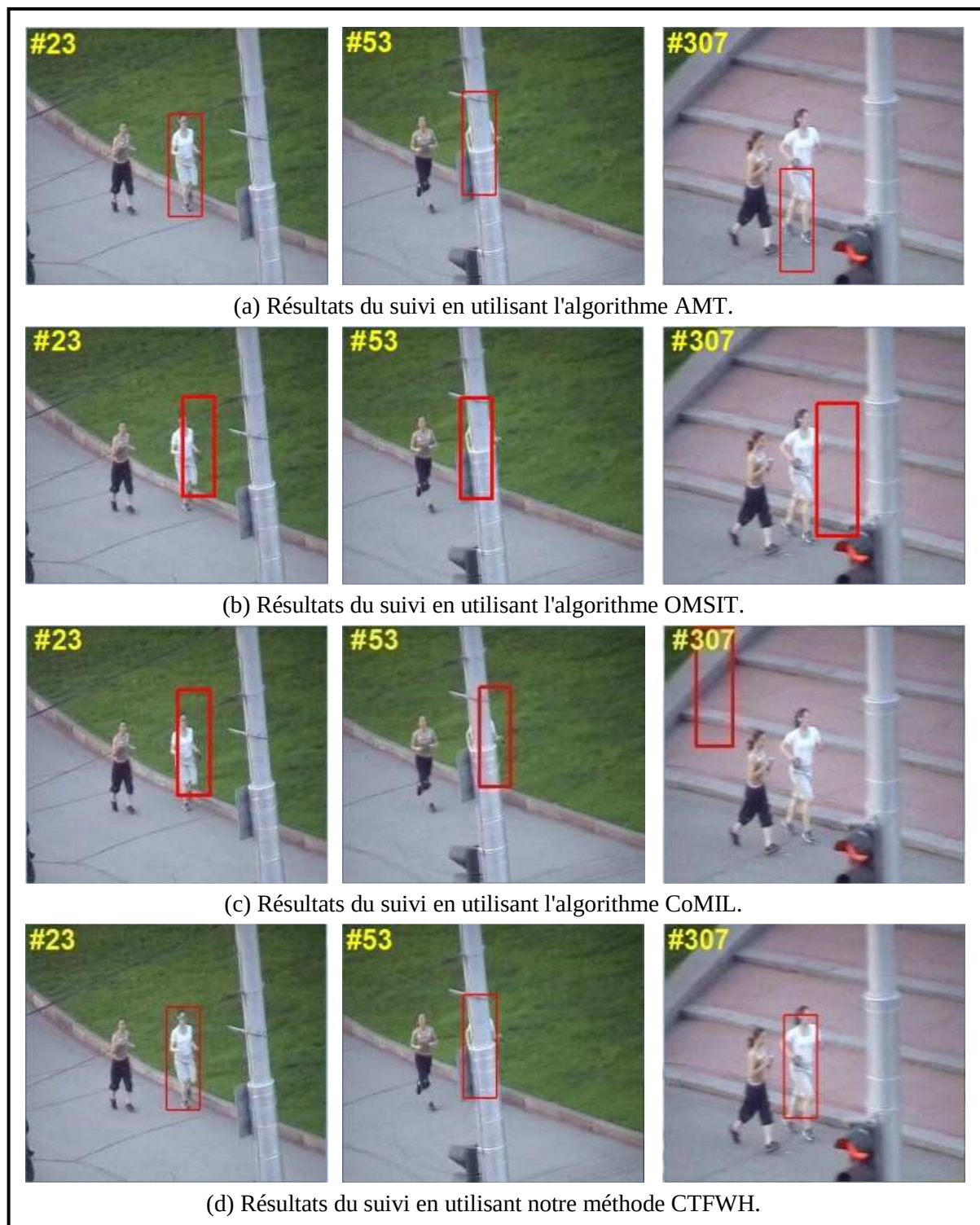


Figure 4.18 : Résultats du suivi pour la séquence d'images Jogging2 avec occultation.

- **Apparition d'éléments distracteurs** : La figure 4.19 ci-après présente les résultats du suivi de visage dans une scène modérément chargée (quatre personnes ou plus) en présence d'éléments distracteurs. Dans cette expérience, notre objectif est de tester la distinctivité des méthodes comparées par rapport à l'apparition d'éléments distracteurs. Dans la séquence Jp1, nous visons à suivre le visage d'une personne qui est notre cible en présence

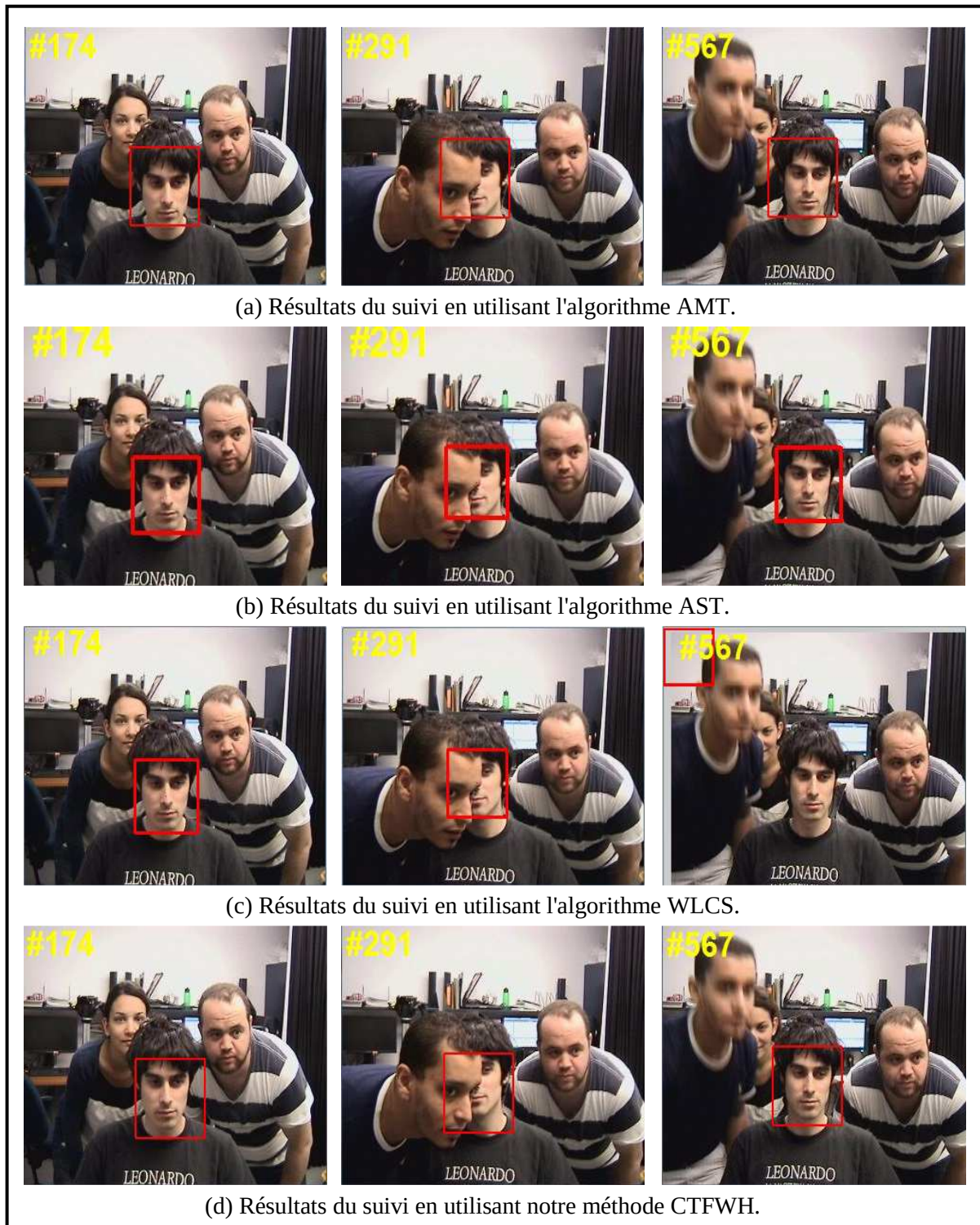


Figure 4.19 : Résultats du suivi de visage dans une scène modérément chargée pour la séquence d'images Jp1 en présence d'éléments distrayants.

d'autres visages distrayants. Ces visages se déplacent autour de la cible et ils l'occultent partiellement plusieurs fois. Certains exemples de résultats qualitatifs de suivi pour cette séquence sont présentés sur la figure 4.19 ci-avant où le visage suivi est celui de la personne qui est au centre de la scène. Concernant les résultats statistiques présentés dans

les tableaux 4.6 et 4.7 pour la séquence Jp1, ces algorithmes AST, SCMT et notre méthode CTFWH produisent le suivi le plus stable avec un taux de réussite de 100% et une erreur de localisation du centre inférieure à 4 pixels pendant presque tous les 608 images (figure 4.19 (b) et (d)). Bien que, les algorithmes BWH, AMT, FSBWH et STC réalisent les taux de réussite suivants 72.86%, 94.41%, 100% et 68,42%, respectivement, avec une erreur de localisation du centre CLE inférieure à 13 pixels, ces résultats indiquent en général de bonnes performances pour le suivi de visage. Cependant, la courbe du taux de chevauchement et celle de l'erreur de localisation du centre (figures 4.10 et 4.13) montrent que les trois premiers algorithmes dérivent la cible considérablement deux fois (d'abord entre les images #280 et #420 et la deuxième fois entre les images #450 et #570) pour suivre les visages distrayants qui occultent ou voisinent le visage cible. L'algorithme STC échoue également à l'image #400 lorsque la région du visage suivi (la personne avec une chemise noire au milieu de la scène) est souvent occultée. Nous pouvons également remarquer dans la courbe du taux de chevauchement et celle de l'erreur de la localisation du centre (figures 4.10 et 4.13), que les algorithmes CoMIL, OMSIT et WLCS n'arrivent pas à suivre la cible à l'image #291 et sont incapables de la récupérer à la fin de la séquence selon la figure 4.19 (c). Cependant, notre approche CTFWH n'est pas affectée par la présence des objets similaires autour de la cible, même si une occultation partielle se produite. Cela s'explique principalement par la distinctivité des caractéristiques couleur-texture et du premier-plan par rapport aux caractéristiques de l'arrière-plan utilisées dans les algorithmes BWH, AMT et FSBWH pour caractériser la cible.

Dans la séquence Jp2 (Figure 4.20 ci-après), nous suivons la même personne dans Jp1 qu'elle marche dans une scène modérément chargée avec quatre personnes qui se déplacent librement. Le visage cible se croise devant ou derrière des distracteurs qui peuvent l'occulter complètement pendant une courte période. Faire référence aux résultats présentés dans les tableaux 4.6 et 4.7 pour la séquence Jp2, et la courbe du taux de chevauchement et celle de l'erreur de la localisation du centre (figures 4.10 et 4.13), la plupart des algorithmes comparés FSBWH, AST, CoMIL, SCMT, OMSIT et WLCS ont confondu la cible avec un visage occulté plusieurs fois après une occultation totale, en particulier entre les images #100 et #150 ou après l'image #200 (figure 4.20 (b) et (c)). Néanmoins, STC, AMT et CTFWH sont capables de récupérer la cible correctement dès qu'une petite partie de celle-ci devient visible et ils ont obtenu respectivement le meilleur, le troisième et le deuxième meilleur résultat en termes du taux de la réussite et de l'erreur

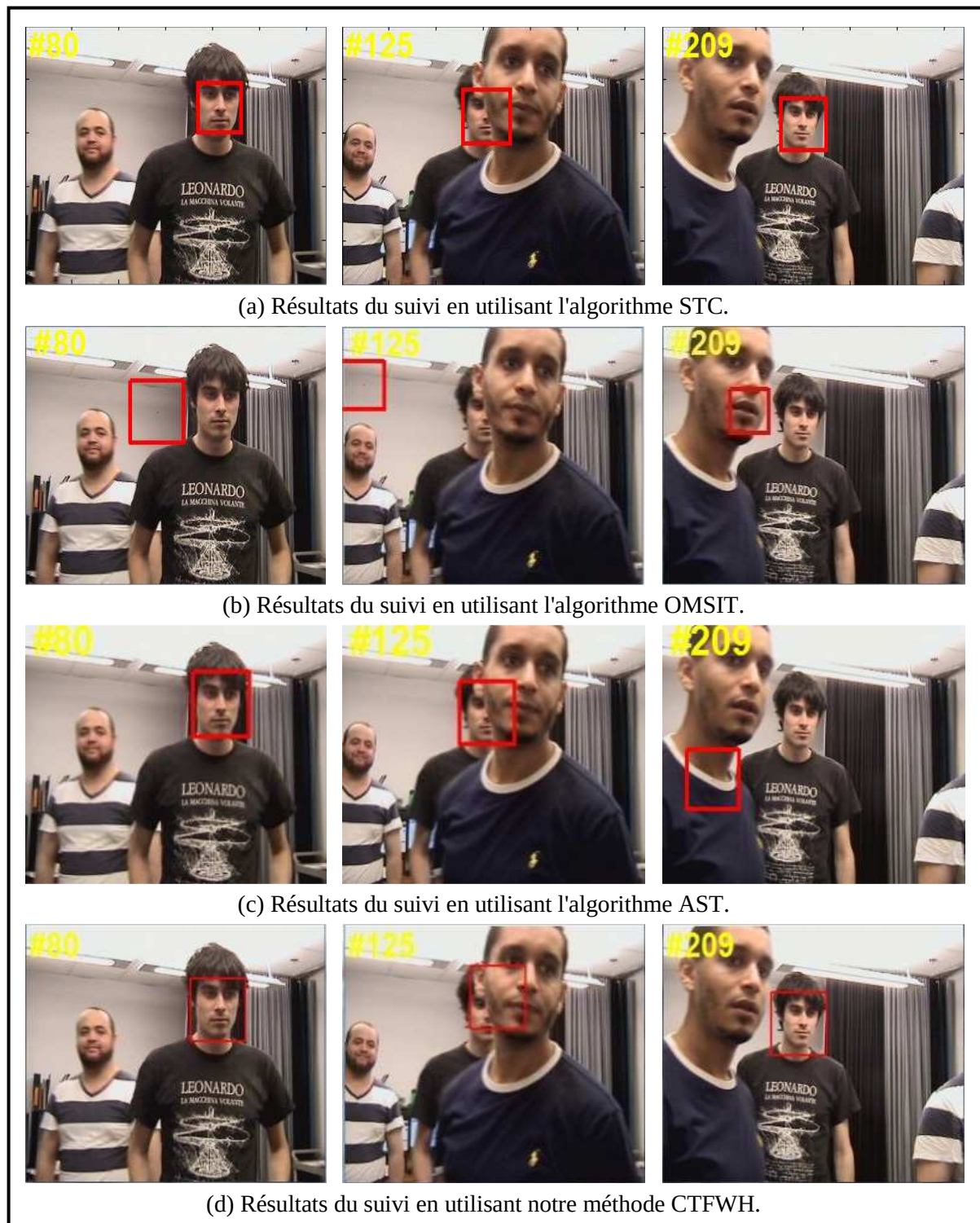


Figure 4.20 : Résultats du suivi de visage dans une scène modérément chargée pour la séquence d'images Jp2 en présence d'éléments distracteurs.

de la localisation du centre.

Dans la troisième expérience, nous avons utilisé la séquence Football1 et la cible à suivre est le casque du joueur numéro 59. Certains exemples de résultats finaux du suivi sont présentés sur la figure 4.21 ci-après. Puisque le modèle d'apparence de la cible pour les

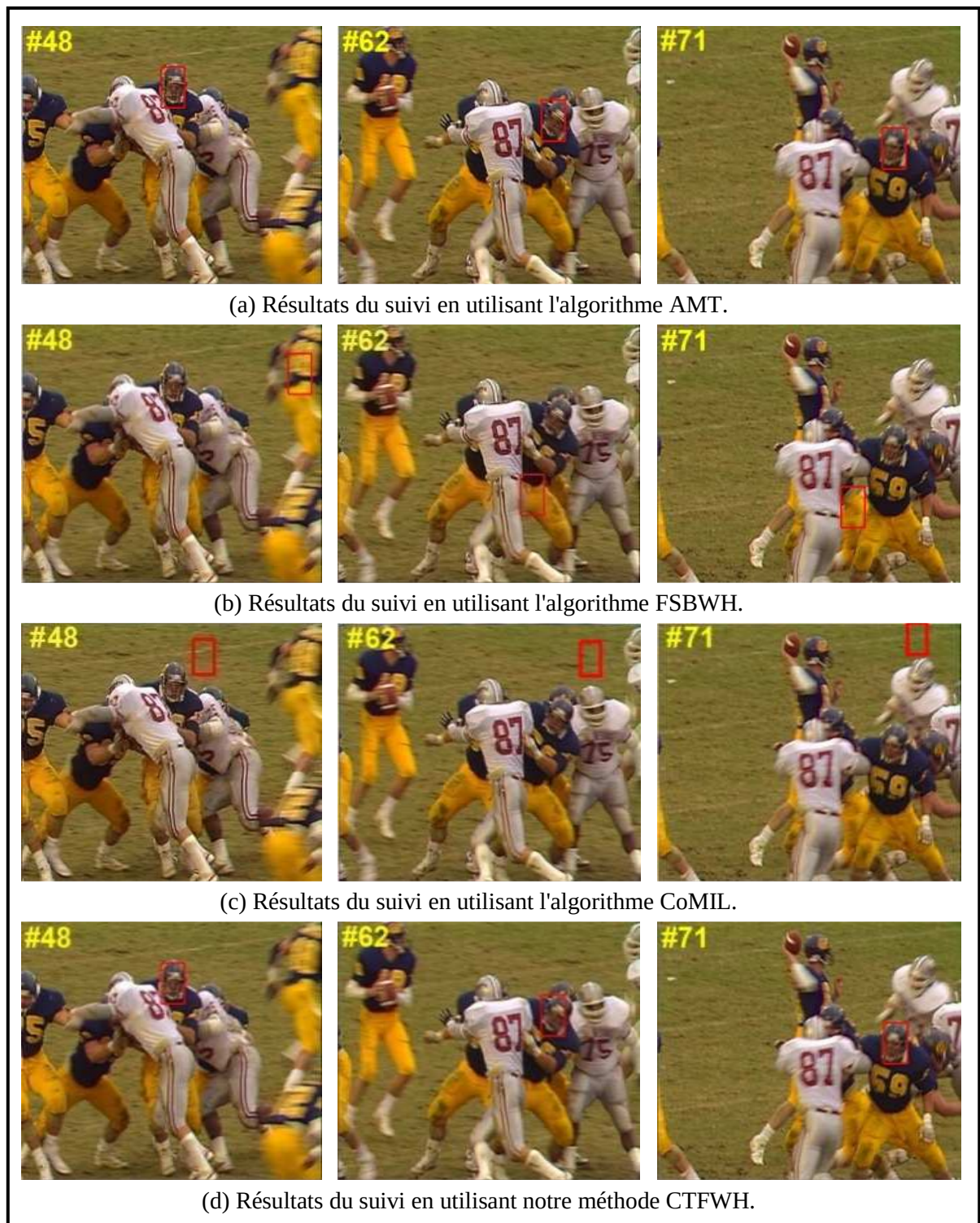


Figure 4.21 : Résultats du suivi de visage dans une scène modérément chargée pour la séquence d'images Football1 en présence d'éléments distracteurs.

algorithmes BWH et AMT exploite les caractéristiques principales de l'arrière-plan dans la localisation de la cible, ces algorithmes peuvent suivre correctement la cible dans toute la séquence selon la figure 4.21 (a) et ils ont réalisé, respectivement, le deuxième et le troisième meilleur résultat en termes du taux de la réussite et de l'erreur de la localisation

du centre (tableaux 4.6 et 4.7 pour la séquence Football1). Mais quand le joueur suivi interagit avec d'autres joueurs qui et WLCS n'arrivent pas à suivre la cible et ne peuvent pas la récupérer dans le reste de la séquence (figure 4.21 (b) et (c)). Malgré la complexité de cette expérience qui contient de nombreux joueurs distrayeurs ayant une apparence similaire à celle de la cible suivie, notre approche CTFWH est capable de suivre correctement le joueur dans toute la séquence dès qu'une petite partie de la cible devient visible (figure 4.21 (d)). De plus, elle obtient simultanément portent les mêmes tenues, la plupart des algorithmes comparés tels que FSBWH, AST, CoMIL, OMSIT, STC le meilleur taux de la réussite 86.49% et la plus petite erreur de la localisation du centre inférieur à 7 pixels. Les résultats satisfaisants obtenus par nos approches AMT et CTFWH dans ces situations sont dues à la distinctivité des caractéristiques importantes utilisées dans la représentation de la cible telles que la jointure des caractéristiques couleur-texture et celles de l'arrière-plan et du premier plan, de sorte que, la cible est effectivement discriminée par rapport à son arrière-plan, même si les valeurs aberrantes (de l'arrière-plan, des voisins ou des visages occultants) peuvent être présents dans la zone des caractéristiques.

- **Encombrement de l'arrière-plan :** La difficulté principale dans les séquences d'images Football1, Basketball et David3 est l'encombrement de l'arrière-plan (c'est-à-dire lorsque l'arrière-plan de la cible dans la scène contient de nombreux autres objets, il est appelé un environnement encombré pour la cible), dont son apparence peut perturber l'algorithme de suivi. En effet, les joueurs de l'équipe de sports sont difficiles à suivre de façon fiable parce qu'ils ont tendance à converger vers le ballon, ils changent souvent leur direction brusquement, et portent les mêmes uniformes lorsqu'ils appartiennent à la même équipe. Le seul moyen fiable pour les identifier est de lire les chiffres sur leurs chemises, mais, compte tenu de la résolution des images, cela ne peut se faire que dans des images relativement peu. Par conséquent, les séquences sportives d'une équipe posent un défi majeur et nous avons testé nos approches sur les séquences Basketball et Football1. Dans la séquence Basketball, la cible à suivre est le joueur numéro 9 (figure 4.22 (d) image #275). Certains exemples de résultats qualitatifs du suivi pour la séquence Basketball sont présentés sur la figure 4.22 ci-après. Toujours basé sur la courbe du taux de chevauchement et celle de l'erreur de la localisation du centre, nous pouvons remarquer que ces algorithmes BWH, AMT, CoMIL, SCMT, OMSIT, STC et WLCS dérivent la cible dans la plupart des images, surtout quand plusieurs joueurs sont regroupés ensemble



Figure 4.22 : Résultats du suivi pour la séquence d'images Basketball avec encombrement de l'arrière-plan.

(figure 4.22 (a) image #474). En revanche, les algorithmes FSBWH, AST et la méthode proposée CTFWH ont FSBWH, AST et la méthode proposée CTFWH ont été capable d'obtenir une meilleure performance malgré la difficulté de cette séquence selon la figure 4.22 (b) et (d). Cependant, notre approche CTFWH fonctionne légèrement mieux par

rapport aux autres approches comparées avec un taux de réussite de 90.24% et une erreur de localisation du centre inférieure à 8 pixels. Il est à noter que les algorithmes AST, FSBWH et CTFWH comprennent des caractéristiques discriminantes facilitant le suivi dans une telle condition. En fait, l'algorithme AST utilise un modèle d'apparence discriminatif basé sur les coefficients clairsemés afin de discriminer la cible et son arrière-plan, FSBWH utilise la saillance du premier-plan basé sur l'histogramme pondéré de l'arrière-plan pour représenter la cible, tandis que notre approche CTFWH utilise les informations des régions du premier-plan et de la cible comme la distribution couleur-texture pour évaluer la qualité de suivi. En effet, la cible subit des rotations importantes hors le plan de l'image et un changement drastique d'illumination (figure 4.22 image #678). L'approche proposée CTFWH n'est pas affectée par ces difficultés puisqu'elle continue à adapter le modèle d'apparence en mettant à jour le modèle du premier-plan. De même, pour la séquence Football1 présentée dans la catégorie de l'apparition d'éléments distracteurs discutée précédemment (figure 4.21 ci-avant), la méthode proposée CTFWH a obtenu une meilleure performance malgré l'encombrement de l'arrière-plan et les occultations.

Dans la séquence David3 (Figure 4.23 ci-après), la cible suivie est la personne qui marche sur la route sous de nombreux facteurs difficiles : l'occultation, la déformation, la rotation hors le plan de l'image et l'encombrement de l'arrière-plan. Certains exemples de résultats qualitatifs du suivi pour cette séquence sont présentés sur la figure 4.23 ci-après. Faire référence à la courbe du taux de chevauchement et celle de l'erreur de la localisation du centre (figures 4.11 et 4.14), notre approche CTFWH a obtenu de bons résultats par rapport aux méthodes comparées malgré le fond encombré et l'occultation fréquente durant le suivi. Tous les algorithmes comparés perdent la cible très rapidement (figure 4.23 (a), (b) et (c)), à l'exception, notre approche AMT a obtenu le deuxième meilleur résultat avec un taux de réussite de 71.43% en raison de son modèle d'apparence qui facilite la distinction entre l'objet et l'arrière-plan en utilisant la jointure des caractéristiques couleur-texture et celles de l'arrière-plan pour représenter la cible. Notez que, notre méthode présente également un aspect discriminatif entre les caractéristiques de la cible et de l'arrière-plan, alors qu'elle suit exactement la personne dans toute la séquence (figure 4.23 (d)), avec un meilleur taux de réussite de 91.32% et une erreur de localisation du centre inférieure à 13 pixels.

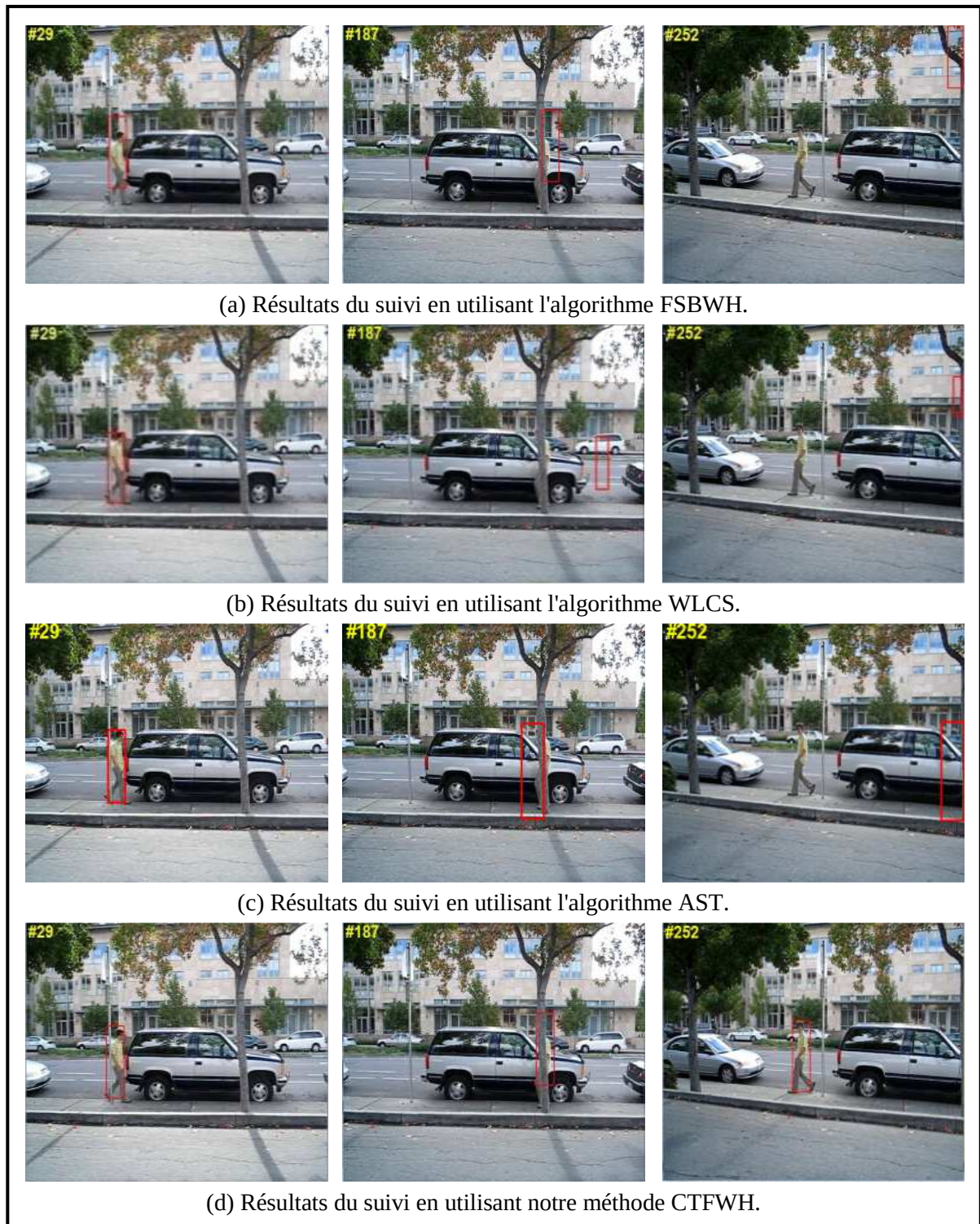


Figure 4.23 : Résultats du suivi pour la séquence d'images David3 avec encombrement de l'arrière-plan.

Dans la séquence DragonBaby, nous suivons le visage du bébé dans des conditions différentes telles que le mouvement rapide, l'occultation, la rotation, le changement d'échelle et la sortie du champ visuel. Certains exemples de résultats qualitatifs du suivi pour cette séquence sont présentés sur la figure 4.24 ci-après. Basé sur la courbe du taux



Figure 4.24 : Résultats du suivi pour la séquence d'images DragonBaby avec la sortie du champ visuel.

de chevauchement et celle de l'erreur de la localisation du centre (figures 4.12 et 4.15), nous pouvons remarquer que, seulement les algorithmes suivants BWH, FSBWH et CTFWH sont capables de suivre la cible jusqu'à la fin de la séquence (figure 4.24 (b) et (c)). En utilisant notre approche CTFWH, nous avons perdu la cible lorsque le visage a

été détourné de la caméra, mais nous avons pu récupérer le suivi rapidement dès qu'il a réapparu partiellement. En effet, l'approche proposée CTFWH permet d'obtenir le troisième meilleur résultat en termes du taux de réussite et de l'erreur de la localisation du centre suivie par les algorithmes FSBWH et BWH qui réalisent le premier et le deuxième meilleur résultat du suivi, respectivement, (figure 4.24 (a)). Ainsi, les résultats présentés dans cette catégorie montrent que notre méthode n'est pas affectée par la sortie du champ visuel; cela peut être attribué à l'utilisation des caractéristiques saillantes pour représenter la cible et qui sont capables de capturer les changements d'apparence lorsqu'une partie de la cible quitte le champ visuel.

- **Mouvement brusque et rotation** : Dans les séquences Human7, Woman, Boy, Gym, DragonBaby et Crossing la cible subit de nombreuses difficultés : le mouvement rapide, la rotation hors et dans le plan de l'image et la déformation au cours de la scène. Dans la séquence Human7, la cible à suivre est la personne qui marche sur la route. Les difficultés de la séquence sont : la capacité à prendre en compte le mouvement rapide de la cible tout autant que celle de la caméra, les changements d'échelle et d'illumination, l'occultation, la déformation et le flou du mouvement (tableau 4.5). Certains résultats finaux du suivi pour cette séquence sont présentés sur la figure 4.25 ci-après. Basé sur la courbe du taux de chevauchement et celle de l'erreur de la localisation du centre (figures 4.10 et 4.13), nous pouvons constater que tous les algorithmes comparés dérivent la cible dans la plupart des images (figure 4.25 (a), (b) et (c)). En fait, la cible subit un changement drastique de l'apparence à cause du mouvement rapide de celle-ci et malgré l'initialisation du modèle de la cible qui contient beaucoup d'informations de l'arrière-plan (figure 4.25 image #11), l'approche proposée CTFWH n'est pas affectée par ces difficultés et elle arrive à adapter le modèle d'apparence grâce à l'étape de mise à jour du modèle du premier-plan (figure 4.25 (d)). Faire référence aux résultats statistiques présentés dans les tableaux 4.6 et 4.7 pour la séquence Human7, l'approche proposée a obtenu la meilleure performance en termes du taux de réussite et de l'erreur de la localisation du centre.

Le visage de la personne cible dans la séquence Boy, expose des rotations hors et dans le plan de l'image et des mouvements brusques pendant plus de 600 images. Certains exemples de résultats qualitatifs du suivi pour cette séquence sont présentés sur la figure 4.26 ci-après. Comme nous pouvons le constater dans la courbe du taux de chevauchement et celle de l'erreur de la localisation du centre (figures 4.11 et 4.14), la plupart des algorithmes comparés n'arrivent pas à suivre correctement le visage (figure 4.26 (b) et (c)).



Figure 4.25 : Résultats du suivi pour la séquence d'images Human7 avec le flou du mouvement et le mouvement rapide.

sauf que, ces algorithmes BWH, FSBWH et CTFWH ont pu suivre le visage avec un taux de réussite plus de 90%. Cependant, notre approche CTFWH a obtenu le meilleur taux de réussite de 100% avec une erreur de localisation inférieure à 4 pixels (figure 4.26 (d))

Dans la séquence Gym, la cible subit de nombreux facteurs critiques tels que les



Figure 4.26 : Résultats du suivi pour la séquence d'images Boy avec rotation hors et dans le plan de l'image, le flou du mouvement et le mouvement rapide.

changements de l'échelle, les rotations hors et dans le plan de l'image et les déformations. Certains exemples de résultats du suivi pour cette séquence sont présentés dans la figure 4.27 ci-après. Selon les courbes du taux de réussite et de l'erreur de la localisation du centre (figures 4.12 et 4.15), la plupart des algorithmes comparés perdent la cible

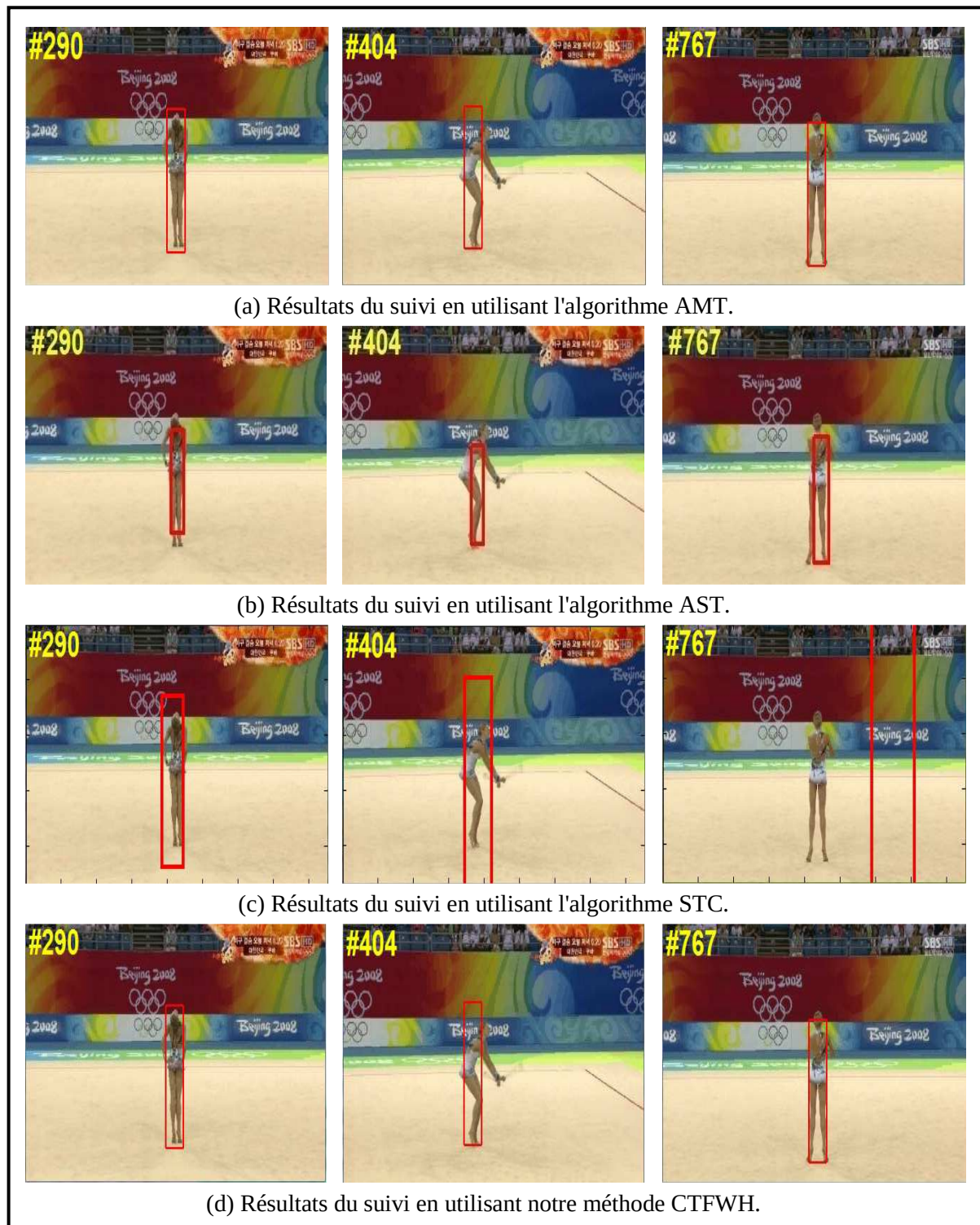


Figure 4.27 : Résultats du suivi pour la séquence d'images Gym avec rotation hors ou dans le plan de l'image.

rapidement (figure 4.27 (c)). Alors que, notre méthode CTFWH a réalisé le taux de la réussite le plus élevé de 75.76% et l'erreur de la localisation la plus basse inférieure à 8 pixels, ce qui signifie que l'approche proposée CTFWH est significativement plus robuste et précise que la plupart des méthodes comparées (figure 4.27 (d)). De même, pour les

séquences Football1, Woman, DragonBaby et Crossing, où la cible suivie subit de nombreux facteurs critiques, l'approche proposée CTFWH permet d'obtenir les meilleures performances en termes du taux de réussite et de l'erreur de la localisation du centre (tableaux 4.5 et 4.6).

- **Changement d'illumination** : Parmi les principales difficultés dans les séquences Human7, Woman et Basketball sont les changements brusques de l'illumination pendant que la cible se déplace dans la scène. Dans la séquence Human7 présentée sur la figure 4.25 ci-avant, le contraste entre la cible et l'arrière-plan est élevé, cela a provoqué des dérives plusieurs fois pour tous les algorithmes comparés (voir la courbe du taux de chevauchement et celle de l'erreur de la localisation du centre figures 4.10 et 4.13). Cependant, l'algorithme SCMT et notre méthode CTFWH arrivent à bien suivre la femme qui marche, tandis que, les autres méthodes comparées la dérivent lorsqu'une variation drastique de l'illumination se produite. Pour la séquence Woman, la cible a connu un changement important de l'apparence en raison des variations importantes de l'illumination et de l'échelle (figure 4.16 (d) l'image #93). De même, pour la séquence Basketball, la cible subit un grand changement d'illumination (figure 4.22 (d) l'image #678). Alors que, quelques algorithmes sont capables de suivre la cible jusqu'à la fin de la séquence et l'algorithme proposé CTFWH permet d'obtenir une erreur de la localisation du centre minimale et un taux de réussite le plus élevé de 90.24%. D'où, les résultats présentés montrent que notre méthode CTFWH n'est pas influencée par les changements importants de l'illumination, cela est attribué à l'utilisation des caractéristiques saillantes pour représenter la cible qui sont capables de capturer les changements d'apparence à cause des changements importants de l'illumination.
- **Coût de calcul** : Nos approches proposées AMT et CTFWH sont conçus pour maintenir une complexité de calcul raisonnable. En fait, nous avons extrait les points caractéristiques significatifs de la cible dans une région de l'image limitée et initialisée dans la première image de la séquence en utilisant la position de la vérité du terrain. Tandis que, pour calculer les caractéristiques du premier-plan ou de l'arrière-plan, nous avons utilisé une zone égale à deux fois la région de la cible, afin d'exploiter de nombreuses caractéristiques de cette dernière. Pour la prise en compte des contraintes du temps de calcul, pour permettre le suivi en temps réel, nous avons également limité le nombre d'itérations maximum de l'algorithme Mean Shift à $N_{max} = 20$ [Comaniciu 03c]. À chaque itération, le vecteur Mean Shift est calculé et la similarité entre l'histogramme du modèle de la cible

et les cibles candidates est augmentée. Ce processus est répété jusqu'à ce que la convergence soit réalisée, dans la pratique, le nombre moyen d'itérations est beaucoup plus petit, d'environ 4, alors l'estimation de l'état de l'objet s'effectue ainsi dans un nombre limité d'itérations. En pratique, le temps de calcul de nos approches AMT et CTFWH est étroitement lié au nombre des caractéristiques détectées qui votent pour la position de l'objet, et qui dépend principalement de la taille et de la texture de l'objet. À titre d'exemple, pour la séquence Football1 avec une petite taille de la cible, notre algorithme CTFWH exige moins d'une seconde par image 0.72 et suit à peu près trois images par seconde. D'autre part, lorsque la taille de l'objet est plus grande, comme dans la séquence Woman, notre algorithme CTFWH nécessite jusqu'à deux secondes pour trouver la cible dans certaines images et suit 0.51 images par seconde.

Le tableau 4.8 ci-après, expose une évaluation quantitative en termes de temps moyen d'exécution par image en secondes TPF (Time per Frame), la moyenne des images suivies par seconde FPS (Frame per Second) et le nombre moyen d'itérations NoI (Number of Iterations). Nous présentons le nombre moyen d'itérations (NoI) uniquement pour nos méthodes AMT et CTFWH et les algorithmes comparés qui utilisent un modèle d'apparence basé sur la technique Mean Shift [Comaniciu 03c] à savoir les algorithmes BWH et FSBWH. Dans le tableau 4.8 ci-dessous, la police rouge gras indique le meilleur résultat, la police bleue gras indique le deuxième meilleur résultat et la police verte gras indique le troisième meilleur résultat. Les méthodes proposées AMT et CTFWH ont été comparé avec huit algorithmes récents de la littérature sur quinze séquences des images différentes. Selon les mesures effectuées, notre approche CTFWH a besoin en moyenne de 0.46 s pour traiter une image, qui est le deuxième meilleur temps d'exécution. De plus, elle suit 3.76 images par seconde qui représente le troisième meilleur résultat. C'est principalement parce que la méthode proposée CTFWH ne suit que les points caractéristiques de la région candidate, mais elle a un temps de calcul supplémentaire pour calculer les caractéristiques de texture. Le nombre moyen d'itérations par image (NoI) est également présenté dans le tableau 4.8 environ 3.74, ce qui signifie que la méthode proposée CTFWH converge plus rapidement que les autres algorithmes référencés : BWH, AMT et FSBWH. Par conséquent, le suivi par l'algorithme Mean Shift en utilisant nos représentations proposées pour la cible focalise sur les principales caractéristiques pendant le suivi, ce qui est très utile pour accélérer la convergence de l'algorithme Mean Shift. En général, la méthode proposée CTFWH répond à l'exigence du temps réel. Nous notons que, l'algorithme STC a obtenu le temps d'exécution le plus court, en

	BWH	FSBWH	AST	CoMIL	SCMT	OMSIT	STC	WLCS	AMT	CTFWH
TPF	0.77	0.9	0.80	547.75	57.82	163.04	0.15	0.67	1.53	0.46
FPS	3.04	3.10	1.31	0.05	0.93	0.12	21.06	5.02	0.97	3.76
NoI	4.96	3.81	-	-	-	-	-	-	5.45	3.74

Tableau 4.8 : Comparaison en termes du temps moyen d'exécution (TPF), la moyenne des images suivies par seconde (FPS) et le nombre moyen d'itérations (NoI).

traitant une image en 0.15 s et il suit 21.06 images par seconde, et l'algorithme WLCS a réalisé le troisième meilleur temps d'exécution. Cependant, l'algorithme CoMIL est réalisé le dernier résultat.

4.3.5 Discussion

D'après les expérimentations qualitatives et quantitatives effectuées dans la section précédente, nous pouvons dire que nos contributions AMT et CTFWH donnent des résultats satisfaisants par rapport aux méthodes avec lesquelles nous avons fait les comparaisons. En faisant une comparaison entre les deux méthodes que nous avons proposées dans cette section, nous pouvons conclure que :

- La méthode AMT utilise la combinaison des caractéristiques couleur-texture et l'histogramme pondéré de l'arrière-plan pour la représentation de la cible.
- La méthode CTFWH utilise la jointure des caractéristiques couleur-texture qui est une partie commune avec la première méthode et l'histogramme pondéré du premier-plan pour la représentation de la cible, mais son point fort par rapport à la première approche réside dans l'utilisation des caractéristiques plus robustes par rapport à la similarité d'apparence de l'objet avec l'arrière-plan de l'image, ainsi qu'à la présence d'éléments distracteurs.

En guise de conclusion, les résultats expérimentaux montrent que la méthode proposée CTFWH permet de suivre correctement la cible dans des scénarios difficiles de suivi à savoir, l'encombrement de l'arrière-plan, les changements d'illumination et d'échelle, les rotations, le flou du mouvement, l'apparition d'éléments distracteurs, le mouvement rapide de l'objet, les variations de l'arrière-plan, la déformation, la similarité d'apparence entre l'objet et son arrière-plan, l'occultation et le mouvement de la caméra. En effet, les meilleures performances obtenues par notre méthode CTFWH en termes du taux de réussite et de l'erreur de la localisation du centre sont dues à l'exploitation efficace de l'information spatiale de

texture avec les caractéristiques principales du premier-plan pour la localisation de la cible, ce qui confirme la pertinence des idées développées dans ce travail, en démontrant que nous avons dépassé une variété des algorithmes récents de la littérature. Nous apportons le succès de l'approche proposée CTFWH à plusieurs raisons : Premièrement, nous avons exploité l'algorithme Mean Shift comme cadre de suivi, ce qui a renforcé la stabilité de suivi des personnes dans une séquence d'images et nous avons construit un modèle d'apparence plus précis pour ajuster à la variation de l'apparence, qui est un facteur très important pour le suivi à long terme dans des situations difficiles. Deuxièmement, l'information de texture basée sur un modèle d'apparence a été validée plus efficacement pour les approches de suivi, puis nous avons introduit ce concept dans la représentation de la cible afin de construire un modèle d'apparence capable de différencier les objets ayant des distributions de couleurs identiques. Troisièmement, nous avons proposé un modèle de représentation du premier-plan à savoir l'histogramme pondéré du premier-plan FWH qui exploite les caractéristiques principales du premier-plan. Le modèle d'apparence final permet de discriminer entre la cible et l'arrière-plan, ce qui est très important pour supprimer les caractéristiques de la cible qui se trouvent dans l'arrière-plan. Basé sur ce modèle discriminatif, l'algorithme de suivi améliore les performances dans la localisation de la cible. De plus, la mise à jour du modèle du premier-plan rend notre méthode plus robuste aux changements d'apparence pendant le processus du suivi.

Néanmoins, l'approche AMT présente plusieurs avantages. Tout d'abord, la jointure des caractéristiques couleur-texture vise à éliminer les régions lisses de l'arrière-plan, ainsi qu'elle permet de réduire le bruit dans le processus du suivi. Elle sert à exploiter efficacement les caractéristiques principales de l'arrière-plan de sorte que le suivi ne pourra être confondu avec des objets stables de l'arrière-plan. Cependant, les résultats expérimentaux de notre approche AMT montrent aussi qu'elle peut suivre correctement la cible dans des scènes peu complexes, comme, le mouvement de la caméra, l'occultation partielle et le changement de l'arrière-plan, cela est dû aux caractéristiques utilisées telles que couleur-texture et l'histogramme pondéré de l'arrière-plan afin d'avoir une représentation plus effective de la cible. En effet, ces caractéristiques sont moins robustes par rapport à la similarité d'apparence de l'objet avec le fond de l'image, ainsi qu'à la présence d'éléments distracteurs qui est notre objectif visé dans l'approche CTFWH.

4.4 Conclusion

Dans ce chapitre, nous avons exposé les résultats expérimentaux, les comparaisons et les discussions pour les trois contributions SOBWH, AMT et CTFWH qui sont proposées dans le chapitre 3 et qui sont nos contributions aux méthodes de suivi des personnes dans une séquence d'images basées sur les histogrammes. De nombreuses expérimentations qualitatives et quantitatives en comparaison avec d'autres algorithmes récents de la littérature et sur plusieurs types différentes de séquences d'images, montrent les résultats satisfaisants obtenus en termes du taux de réussite, l'erreur de la localisation du centre et le temps d'exécution.

Conclusion générale

Au cours de cette thèse, nos travaux de recherche ont porté sur le suivi de personnes dans une séquence d'images par modèle d'apparence afin de produire de nouvelles méthodes précises, robustes et rapides dans ce domaine.

Le premier chapitre a été consacré aux modèles et principes de suivi d'objets en vision par ordinateur. À cet égard, nous avons défini le principe de suivi d'objets, puis nous avons exposé les représentations des objets basées sur leurs formes et leurs apparences, ainsi que les caractéristiques les plus utilisées pour le suivi d'objets. Les différentes méthodes de suivi existantes dans la littérature ont été classées et présentées selon plusieurs catégories. Toutes ces différentes techniques ont leurs propres avantages et leurs limites. Finalement, nous avons présenté quelques techniques pour l'évaluation des algorithmes de suivi.

Dans le deuxième chapitre, nous avons présenté notre première approche pour le suivi des personnes dans une séquence d'images basée sur les points d'intérêt couleurs. La méthode proposée utilise le modèle de mélange gaussien GMM (Gaussian Mixture Model) pour la segmentation des personnes du premier-plan. Ensuite, à partir du premier-plan détecté, nous avons déterminé des règles qui permettent de définir les régions de la peau pour une bonne détection des personnes. Les points d'intérêt couleurs sont détectés sur les régions de la peau trouvées par l'algorithme de Harris couleur. L'utilisation de l'ensemble des points d'intérêt couleurs nous a permis de suivre les personnes, en faisant des correspondances entre ces points par la mesure de corrélation ZNCC (Zero mean Normalised Cross Correlation). La méthode proposée a été validée sur différentes séquences d'images et nous a permis d'obtenir de très bons résultats même dans le cas où deux personnes se déplacent dans la même scène. Le point faible de cette contribution est l'utilisation d'une caméra statique, ce qui n'est pas toujours le cas dans la réalité. Pour cette raison, nous avons proposé d'autres approches qui s'adaptent avec le mouvement de la caméra.

Dans le troisième chapitre, nous avons proposé trois approches pour le suivi des personnes basées sur les histogrammes et sous le cadre de l'algorithme Mean Shift. Dans la première approche, nous avons développé deux axes principaux, d'une part, les descripteurs de moments de poids de l'image ont été utilisés pour estimer la position, la taille et l'orientation de la cible. D'autre part, l'histogramme pondéré et corrigé de l'arrière-plan a été incorporé dans la représentation de la cible afin de dériver une représentation simple des caractéristiques

de l'arrière-plan et réduire l'interférence de ces caractéristiques dans la localisation de la cible. La deuxième approche s'appuie sur les caractéristiques de texture qui sont intégrées dans l'histogramme de couleur et l'histogramme pondéré de l'arrière-plan. Le point faible de cette contribution est l'utilisation des caractéristiques moins robustes par rapport aux scènes similaires à l'arrière-plan, ainsi qu'à la présence d'éléments distracteurs qui sont des facteurs critiques pour les méthodes de suivi. Pour cette raison, nous avons proposé une troisième approche qui utilise la jointure des caractéristiques couleur-texture et l'histogramme pondéré du premier-plan. L'avantage principal de cette approche étant l'exploitation des caractéristiques importantes du premier-plan et de leur utilisation pour sélectionner uniquement les caractéristiques principales du modèle de la cible.

Dans le quatrième chapitre, nous avons présenté les résultats expérimentaux, les comparaisons et les discussions pour les trois approches présentées dans le chapitre 3. Les résultats expérimentaux qualitatifs et quantitatifs montrent la validité des approches proposées et leurs robustesses, en particulier, à la large variation de l'arrière-plan, la présence d'éléments distracteurs, la déformation, le changement d'échelle et d'orientation, l'occultation, la scène similaire à l'arrière-plan et au mouvement de la caméra.

De nombreux travaux ont été effectués, auparavant, afin d'améliorer le suivi d'objets en vision par ordinateur. Malgré les performances des méthodes déployées, cet axe mérite des réflexions plus approfondies et les perspectives demeurent nombreuses :

- ✓ Proposer de nouvelles approches pour le suivi de plusieurs personnes dans des scénarios du monde réel.
- ✓ Exploiter les résultats de suivi des personnes qui représentent un niveau intermédiaire, dans le processus de l'analyse du mouvement humain pour la reconnaissance et l'interprétation des mouvements du corps humain qui représentent un haut niveau. Il existe multiples formes des mouvements du corps humain, par exemple, les mouvements associés à la manière de marcher, de communiquer ou d'effectuer des activités.
- ✓ Réduire le temps d'exécution des approches proposées pour que le suivi soit en temps réel reste un défi de taille dans tous les algorithmes de suivi.

Bibliographie

- [Adam 06] A. Adam, E. Rivlin and I. Shimshoni. *Robust fragments-based tracking using the integral histogram*. In Proc of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. 798–805, USA, June 2006.
- [Ali 01] A. Ali and J. K. Aggarwal. *Segmentation and recognition of continuous human activity*. IEEE Workshop on detection and recognition of events in video, Vol.1 pp. 28-35, 2001.
- [Allen 04] J. G. Allen, R. Y. Xu and J. S. Jin. *Object tracking using camshift algorithm and multiple quantized feature spaces*. In Proc of the Pan-Sydney area workshop on Visual information processing, pp. 3-7, Australian, June 2004.
- [Avidan 04] S. Avidan. *Support vector tracking*. IEEE transactions on pattern analysis and machine intelligence, Vol. 26, No. 8, pp. 1064-1072, 2004.
- [Argyros 04] A. A. Argyros and M. I. Lourakis. *Real-time tracking of multiple skin-colored objects with a possibly moving camera*. In European Conference on Computer Vision, pp. 368-379, Springer, Berlin, Heidelberg, 2004.
- [Babenko 09] B. Babenko, M. H. Yang and S. Belongie. *Visual tracking with online multiple instance learning*. In Proc of Computer Vision and Pattern Recognition (CVPR), pp. 983–990, USA, June 2009.
- [Bach 96] J. Bach, C. Fuler, A. Gupta, A. Hampapur, B. Horowitz, R. Humphrey, R. Jain and C. Shu. *The Virage image search engine: An open framework for image management*. In proc of Storage and Retrieval for Image and Video Databases SPIE, pp. 76–87, March1996.
- [Bashir 06] F. Bashir and F. Porikli. *Performance evaluation of object detection and tracking systems*. In Proceedings 9th IEEE International Workshop on PETS, pp. 7-14, 2006.
- [Ballard 82] D. H. Ballard and C. M. Brown. *Computer Vision*. Vol. 2, Englewood Cliffs (NJ), Prentice-Hall, 1982.
- [Bar-Shalom 90] Y. Bar-Shalom, T. E. Fortmann and P. G. Cable. *Tracking and data association*. The Journal of the Acoustical Society of America, Vol. 87, No. 2, pp. 918-919, 1990.
- [Baumberg 94] A. M. Baumberg and D. C. Hogg. *An efficient method for contour tracking using active shape models*. In Proc of the IEEE Workshop on Motion of Non-Rigid and Articulated Objects, pp. 194-199, USA,

- November 1994.
- [Bay 06] H. Bay, T. Tuytelaars and L. Van Gool. *SURF: Speeded up robust features*. In Proc of the European Conference on Computer Vision (ECCV), pp. 404–417, 2006.
- [Beaudet 78] P. R. Beaudet. *Rotationally invariant image operators*. In Proc of the 4th International Joint Conference on Pattern Recognition. Tokyo, pp. 579–583, 1978.
- [Benchmark 17] The Visual Tracker Benchmark database, 2017.
<http://www.visual-tracking.net/>
- [Bernardin 07] K. Bernardin, F. Van De Camp and R. Stiefelhagen. *Automatic person detection and tracking using fuzzy controlled active cameras*. In Proc of the IEEE Conference on Computer Vision and Pattern Recognition CVPR’07, pp. 1–8, 2007.
- [Bertalmio 00] M. Bertalmio, G. Sapiro and G. Randall. *Morphing active contours*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 7, pp. 733–737, 2000.
- [Beymer 99] D. Beymer and K. Konolige. *Real-time tracking of multiple people using continuous detection*. In IEEE International Conference on Computer Vision (ICCV) Frame-Rate Workshop, pp. 1-8, 1999.
- [Birchfield 98a] S. Birchfield. *Elliptical head tracking using intensity gradients and color histograms*. In Proc of the IEEE Conference on Computer Vision and Pattern Recognition CVPR’98, Washington, USA, 1998.
- [Birchfield 05b] S. Birchfield and S. Rangarajan. *Spatiograms vs. histograms for region based tracking*. Proc In IEEE Conf Comput Vision Pattern Recognit (CVPR), Vol. 2, pp. 1158–1163, 2005
- [Birgé 02] L. Birgé and Y. Rozenholc. *How many bins should be put in a regular histogram*. Technical report, Laboratoire Probabilités et Modèles Aléatoires, Université Pierre et Marie Curie, Paris, France, PMA-721. 2002.
- [Black 96a] M. J. Black and P. Anandan. *The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields*. Computer vision and image understanding, Vol. 63, No. 1, pp. 75-104, 1996.
- [Black 03b] J. Black, T. Ellis and P. Rosin. *A novel method for video tracking performance evaluation*. Proceedings of the IEEE International Workshop on Visual Surveillance and Performance (VS-PETS), Nice, pp. 125-132, 2003.
- [Black 98c] M. J. Black and A. D. Jepson. *EigenTracking: robust matching and tracking of articulated objects using a view-based representation*. International Journal of Computer Vision, Vol. 26, No. 1, pp. 63-84,

- 1998.
- [Boltz 09] S. Boltz, E. Debreuve and M. Barlaud. *High-dimensional statistical measure for region-of-interest tracking*. IEEE Transactions on Image Processing, Vol. 18, No. 6, pp. 1266-1283, 2009.
- [Boomgard 92] V. D. Boomgard and R. Van Balen. *Methods for Fast Morphological Image Transforms Using Bitmapped Images*. CVGIP: Graphical Models and Image Processing, Vol.54, No.3, pp. 254-258, 1992.
- [Bouachir 15] W. Bouachir and G. A Bilodeau. *Collaborative part-based tracking using salient local predictors*. Computer Vision and Image Understanding Vol. 137, pp. 88-101, 2015.
- [Bousetouane 13] F. Bousetouane, L. Dib and H. Snoussi. *Improved mean shift integrating texture and color features for robust real time object tracking*. The Visual Computer, Vol. 29, No. 3, pp. 155–170, 2013.
- [Bouwman 08] T. Bouwmans, F. El Baf and B. Vachon. *Background modeling using mixture of gaussians for foreground detection-a survey*. Recent Patents on Computer Science, Vol. 1, No. 3, pp. 219-237, 2008.
- [Bowyer 01] K. Bowyer, C. Kranenburg and S. Dougherty. *Edge detector evaluation using empirical roc curve*. Comput. Vision Image Understand, Vol. 10, pp. 77–103, 2001.
- [Bradski 98] G. Bradski. Computer vision face tracking for use in a perceptual user interface. Intel. Technol. J., Vol. 2, No. 2, pp. 1–15, 1998.
- [Broida 86] T. J. Broida and R. Chellappa. *Estimation of object motion parameters from noisy images*. IEEE transactions on pattern analysis and machine intelligence, No. 1, pp. 90–99, 1986.
- [Brown 05] L. M. Brown, A. W. Senior, Y. L. Tian, J. Connell, A. Hampapur, C. F. Shu and M. Lu. *Performance evaluation of surveillance systems under varying conditions*. In Proceedings of IEEE PETS Workshop, pp. 1-8, 2005.
- [Bugeau 11] Aurélie Bugeau. *Détection et suivi d'objets en mouvement dans des scènes complexes, application a la surveillance des conducteurs*. Thèse de l'École Doctorale Mathématiques, Informatique, Signal, Électronique et Télécommunications. Université de Rennes 1. Janvier 2011.
- [Cadoz 94] C. Cadoz. *Le geste canal de communication homme/machine: la communication instrumentale*. Technique et science informatique, Vol. 13, No. 1, pp. 31-61, 1994.
- [Canny 86] J. Canny. *A computational approach to edge detection*. IEEE transactions on pattern analysis and machine intelligence, Vol. 8, No. 6, pp. 679–698, 1986.

- [CAVIAR] Context Aware Vision using Image-based Active Recognition (2017).
<http://homepages.inf.ed.ac.uk/rbf/CAVIAR/>.
- [Chambon 11] S. Chambon and A. Crouzil. *Similarity measures for image matching despite occlusions in stereo vision*. Pattern Recognit., Vol. 44, No. 9, pp. 2063–2075, 2011.
- [Château 02] T. Château, F. Jurie, M. Dhome and X. Clady. *Real time tracking using wavelet representation*. Proc of the 24th DAGM Symposium on Pattern Recognition, 2002.
- [Checchin 08] P. Checchin, S. Gidel and C. Blanc. *Système de détection de piétons a bord de véhicules : approche par télémétrie laser*. Proc In Nat. Res. Univ. Inst. Technol. (CNRIUT), 2008.
- [Chen 01] Y. Chen, Y. Rui and T. Huang. *JPDF based HMM for real-time contour tracking*. Proc In Computer Vision and Pattern Recognition (CVPR), Vol. 1, pp. 543–550, 2001.
- [Cheng 95] Y. Cheng. *Mean shift, mode seeking, and clustering*. IEEE transactions on pattern analysis and machine intelligence, Vol. 17, No. 8, pp. 790-799, 1995.
- [Chetverikov 00] D. Chetverikov, M. Nagy and J. Verestoy. *Comparison of Tracking Techniques Applied to Digital PIV*. Proc of the International Conference on Pattern Recognition, pp. 4619-4622, 2000.
- [Collins 03a] R. T. Collins. *Mean-shift blob tracking through scale space*. Proc In IEEE Conf. Computer Vision and Pattern Recognition, Vol. 2, pp.234–240, 2003.
- [Collins 01b] R. T. Collins, A. J. Lipton, H. Fujiyosh and T. Kanade. *Algorithms for cooperative multisensor surveillance*. Proc of the IEEE, Vol. 89, No. 10, pp. 1456-1477, 2001.
- [Collins 05c] R. T. Collins, Y. Liu and M. Leordeanu. *Online Selection of Discriminative Tracking Features*. IEEE Transactions on pattern analysis and machine intelligence, Vol. 27, No. 10, pp. 1631-1643, 2005.
- [Comaniciu 99a] D. Comaniciu and P. Meer. *Mean shift analysis and applications*. Proc In IEEE International Conference on Computer Vision, pp. 1197–1203, Kerkyra, Greece, 1999.
- [Comaniciu 02b] D. Comaniciu and P. Meer. *Mean shift: A robust approach toward feature space analysis*. IEEE Transactions Pattern Analysis Machine Intelligence, Vol. 24, No. 5, pp. 603–619, 2002.
- [Comaniciu 03c] D. Comaniciu, V. Ramesh and P. Meer. *Kernel-based object tracking*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 25, No. 5, pp. 564–577, 2003.

- [Comaniciu 00d] D. Comaniciu, V. Ramesh and P. Meer. *Real-time tracking of non-rigid objects using mean shift*. Proc of the IEEE Conference Computer Vision and Pattern Recognition, Vol. 2, pp. 142–149, 2000.
- [Conaire 07] C. O. Conaire, N. E. O. Connor and A. F. Smeaton. *An improved spatiogram similarity measure for robust object localization*. Proc In International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2007
- [Cootes 01] T. Cootes, G. Edwards and C. Taylor. *Robust real-time periodic motion detection, analysis, and applications*. IEEE Trans. Patt. Analy. Mach. Intell. Vol. 23, No. 6, pp. 681–685, 2001.
- [Cremers 02] D. Cremers, T. Kohlberger and C. Schnörr. *Nonlinear shape statistics in mumford—shah based segmentation*. In Computer Vision—ECCV 2002, pp. 93–108, 2002.
- [Cucchiara 03] R. Cucchiara, C. Grana, M. Piccardi and A. Prati. *Detecting moving objects, ghosts, and shadows in video streams*. IEEE Trans. On Pattern Analysis and Machine Application, Vol. 25, No. 10, pp. 1337-1342, 2003.
- [Cutler 00] R. Cutler and L. S. Davis. *Robust real-time periodic motion detection, analysis, and applications*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 8, pp. 781-796, 2000.
- [Edwards 98] G. Edwards, C. Taylor and T. Cootes. *Interpreting face images using active appearance models*. International Conference on Face and Gesture Recognition, pp. 300–305, 1998.
- [El Fkihi 08] Sanaa El Fkihi. *Modèles probabilistes indexés par les arbres : Application à la détection de la peau dans les images couleur*. Thèse, Université des Sciences et Technologies de Lille, Université Mohammed V Agdal Rabat, 2008.
- [Elgammal 00a] A. Elgammal, D. Harwood and L.S. Davis. *Non-parametric model for background subtraction*. European Conference on Computer Vision, pp. 751-767, 2000.
- [Elgammal 03b] A. Elgammal, R. Duraiswami and L.S. Davis. *Probabilistic tracking in joint feature-spatial spaces*. In IEEE International Conference on Computer Vision and Pattern Recognition, pp. 781–788, 2003.
- [Elgammal 02c] A. Elgammal, R. Duraiswami, D. Harwood and L. S. Davis. *Background and foreground modeling using nonparametric kernel density estimation for visual surveillance*. Proceedings of the IEEE, Vol. 90, No. 7, pp. 1151-1163, 2002.
- [FEI 14] The FEI face database.
<http://fei.edu.br/~cet/facedatabase.html>, 2014.

- [Fieguth 97] P. Fieguth and D. Terzopoulos. *Color-based tracking of heads and other mobile objects at video frame rates*. Proc of the IEEE Computer Society Conference In Computer Vision and Pattern Recognition, pp. 21-27, 1997.
- [Fischler 81] M. A. Fischler and R. C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Graphics and Image Processing, June 1981.
- [Fleck 96] M. M. Fleck, D. A. Forsyth and C. Bregler. *Finding naked people*. European Conference on Computer Vision, Springer-Verlag, Berlin, Germany, pp. 592-602, 1996.
- [Fukunaga 75] K. Fukunaga and L. Hostetler. *The estimation of the gradient of a density function, with applications in pattern recognition*. IEEE Transactions on information theory, Vol. 21, No. 1, pp. 32-40, 1975.
- [Gagalowicz 83] A. Gagalowicz. *Vers un modèle de textures*. Thèse doctorale, université Pierre et Marie Curie, Paris V, 1983.
- [Georgescu 04] B. Georgescu and P. Meer. *Point matching under large image deformations and illumination changes*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 26, No. 6, pp. 674-688, 2004.
- [Geusebroek 01] J. M. Geusebroek, R. Van den Boomgaard, A. W. M. Smeulders and H. Geerts. *Color invariance*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 23, No. 12, pp. 1338-1350, 2001.
- [Girondel 02] V. Girondel. *Détection de peau, suivi de tête et de mains pour des applications multimédia*. DEA Signal Image Parole Télécom, Juillet 2002.
- [Gotlieb 90] C. C. Gotlieb and H. E. Kreyszig. *Texture descriptors based on co-occurrence matrices*. Computer Vision, Graphics, and Image Processing, Vol. 51, No. 1, pp. 70-86, 1990.
- [Gouet 02a] V. Gouet and N. Boujemaa. About optimal use of color points of interest for content-based image retrieval. Internal Report, INRIA Rocquencourt, 2002.
- [Gouet 04b] V. Gouet and B. Lameyre. *SAP: A robust approach to track objects in video streams with snakes and points*. British Machine Vision Conference. pp 737–746, 2004.
- [Grabner 07a] M. Grabner, H. Grabner and H. Bischof. *Learning features for tracking*. Proc of the IEEE Conference In Computer Vision and Pattern Recognition, pp. 1–8, IEEE, 2007.
- [Gros 97] P. Gros, G. Mclean, R. Delon, R. Mohr, C. Schmid and G. Mistler. *Utilisation de la couleur pour l'appariement et l'indexation d'images*.

- Technical report, INRIA, 1997.
- [Guo 14] Y. Guo, Y. Chen, F. Tang, A. Li, W. Luo and M. Liu. *Object tracking using learned feature manifolds*. Computer Vision and Image Understanding, Vol. 118, pp. 128–139, 2014.
- [Hager 04] G. D. Hager, M. Dewan and C. V. Stewart. *Multiple kernel tracking with SSD*. Proc of the IEEE Computer Society Conference In Computer Vision and Pattern Recognition, CVPR, Vol. 1, pp. I–790, 2004.
- [Haralick 73] R. M. Haralick, K. Shanmugam and I. Dinstein. *Textural features for image classification*. IEEE Transactions on Systems, Man, and Cybernetics, SMC, Vol. 3, No. 6, pp. 610–621, 1973.
- [Hare 12] S. Hare, A. Saffari and P. H. Torr. *Efficient online structured output learning for keypoint-based object tracking*. Proc of the IEEE Conference In Computer Vision and Pattern Recognition (CVPR), pp. 1894–1901, 2012.
- [Haritaoglu 00a] I. Haritaoglu, D. Harwood and L. S. Davis. *W⁴: Real-time surveillance of people and their activities*. IEEE Transactions on pattern analysis and machine intelligence, Vol. 22, No. 8, pp. 809–830, 2000.
- [Haritaoglu 01b] I. Haritaoglu and M. Flickner. *Detection and tracking of shopping groups in stores*. IEEE Conf Comput Vision Pattern Recognit (CVPR) 1:431–438, 2001.
- [Harris 88] C. Harris and M. Stephens. *A combined corner and edge detector*. Proc of the 4th Alvey Vision Conference. pp. 147–151, 1988.
- [Hausdorff 62] F. Hausdorff. *Set theory*. American Mathematical Soc. Vol. 119, 1962.
- [He 13] L. He, Y. Xu, Y. Chen and J. Wen. *Recent advance on mean shift tracking: A survey*. International Journal of Image and Graphics, Vol. 13, No. 03, 2013.
- [Heikkilä 06] M. Heikkilä and M. Pietikäinen. *A texture-based method for modeling the background and detecting moving objects*. IEEE Trans. Patt. Anal. Mach. Intell. Vol. 28, No. 4, pp. 657–662, 2006.
- [Hou 11] Y. L. Hou and G. K. Pang. *People counting and human detection in a challenging situation*. IEEE Transactions on Systems, Man, and Cybernetics-part a: systems and humans, Vol. 41, No. 1, pp. 24–33, 2011.
- [Horn 91a] R. A. Horn and C. R. Johnson. *Topics in matrix analysis*. Cambridge University Press, U.K. 1991
- [Horn 81b] B. K. Horn and B. G. Schunck. *Determining optical flow*. Artificial intelligence, Vol. 17, No. 1-3, pp. 185–203, 1981.
- [Hsu 02] R. L. Hsu, M. Abdel-Mottaleb and A. K. Jain. *Face detection in color images*. IEEE Transactions on Pattern Analysis and Machine

- Intelligence, Vol. 24, No. 5, pp. 696-706, May 2002.
- [Hu 62a] M. Hu. *Visual pattern recognition by moment invariants*. IRE Transactions on Information Theory 8, pp. 179-187, 1962.
- [Hu 04b] W. Hu, T. Tan, L. Wang and S. Maybank. A survey on visual surveillance of object motion and behaviors. Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 34(3):334–352, 2004
- [Huttenlocher 93] D. P. Huttenlocher, J. J. Noh and W. J. Rucklidge. *Tracking non-rigid objects in complex scenes*. Proc of the Fourth International IEEE Conference in Computer Vision, pp. 93–101, 1993.
- [Huynh 10] Huu-Hung Huynh. *Vidéosurveillance pour appartements intelligents: application à la détection de prise de médicaments*. Thèse de l'École Doctorale de Mathématiques et Informatique. Université de la Méditerranée Aix Marseille 2, Décembre 2010.
- [Irani 98] M. Irani and P. Anandan. *Video indexing based on mosaic representations*. IEEE Trans. Patt. Analy. Mach. Intell., Vo. 20, No. 6, pp. 577–589, 1998.
- [Isard 98] M. Isard and A. Blake. *Condensation – Conditional Density Propagation for Visual Tracking*. International Journal of Computer Vision 29, pp. 5–28, 1998.
- [Islam 09] M. Z. Islam, C.-M. Oh and C.-W. Lee. *Video based moving object tracking by particle filter*. International Journal of Signal Processing, Image Processing and Pattern, Vol. 2, No. 1, 2009.
- [Ito 91] Y. Ito. *Representation of functions by superpositions of a step or sigmoid function and their applications to neural network theory*. Neural Networks, Vol. 4, No. 3, pp.385-394, 1991.
- [Jeyakar 08] J. Jeyakar, R.V. Babu and K. R. Ramakrishnan. *Robust object tracking with background-weighted local kernels*. Comput Vision and Image Understanding, Vol. 112, No. 3, pp. 296–309, 2008.
- [Jepson 03] A. D. Jepson, D. J. Fleet and T. F. El-Maraghi. *Robust online appearance models for visual tracking*. IEEE transactions on pattern analysis and machine intelligence, Vol. 25, No. 10, pp. 1296-1311, 2003.
- [Ji 04] Q. Ji, Z. Zhu and P. Lan. *Real-time non intrusive monitoring and prediction of driver fatigue*. IEEE Transaction on Vehicular Technology, Vol. 53, No. 4, pp. 1052-1068, 2004.
- [Jia 12] X. Jia, H. Lu and M.-H. Yang. *Visual tracking via adaptive structural local sparse appearance model*. Proc of the IEEE Conf Comput Vision Pattern Recognit (CVPR), pp. 1822–1829, 2012.

- [Joenes 99] M. Joenes and J. Rehg. *Statistical Color Models with Application to Skin Detection*. Proc of the IEEE Conference Computer Vision and Pattern Recognition, CVPR'99, pp. 274-280, 1999.
- [Kailath 67] T. Kailath. *The divergence and Bhattacharyya distance measures in signal selection*. IEEE transactions on communication technology, Vol. 15, No.1, pp. 52-60, 1967.
- [Kang 04] J. Kang, I. Cohen and G. Medioni. *Object reacquisition using invariant appearance model*. Proc of the 17th International Conference In Pattern Recognition ICPR, Vol. 4, pp. 759–762, 2004.
- [Kalal 12] Z. Kalal, K. Mikolajczyk and J. Matas. *Tracking-learning-detection*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 34, No. 7, pp. 1409–1422, 2012.
- [Kettnaker 99] V. Kettnaker and R. Zabih. *Bayesian multi-camera surveillance*. Proc of the IEEE Conf Comput Vision Pattern Recognit CVPR, Vol. 2, pp. 2544–2550, 1999.
- [Kim 10] I. S. Kim, H. S. Choi, K. M. Yi, J. Y. Choi and S. G. Kong. *Intelligent visual surveillance: A survey*. International Journal of Control, Automation and Systems, Vol. 8, No. 5, pp. 926–939, 2010.
- [Kitagawa 87] G. Kitagawa. *Non-gaussian state—space modeling of nonstationary time series*. Journal of the American statistical association, Vol. 82, No. 400, pp. 1032–1041, 1987.
- [Koller 93] D. Koller, K. Danilidis and H. Nagel. *Model-based object tracking in monocular image sequences of road traffic scenes*. International Journal of Computer Vision, Vol. 10, No. 3, pp. 257-281, 1993.
- [Kovac 03] J. Kovac, P. Peer and F. Solina. *2D versus 3D color space face detection*. Proc of the 4th EURASIP Conference on Video/Image Processing and Multimedia Communications, Croatia, pp. 449-454, 2003.
- [Kullback 59a] S. Kullback. *Information theory and statistics*. John Wiley and Sons, 1959.
- [Kullback 51b] S. Kullback and R. A. Leibler. *On information and sufficiency*. The Annals of Mathematical Statistics, Vol. 22, No. 1, pp. 79–86, 1951.
- [Laaroussi 16a] K. Laaroussi, A. Saaidi, M. Masrar and K. Satori. *Human tracking based on appearance model*. Proc of the Mediterranean Conference on Information and Communication Technologies, Springer, Vol.1, pp. 297-305, 2016.
- [Laaroussi 17b] K. Laaroussi, A. Saaidi, M. Masrar and K. Satori. *Human Tracking using joint color-texture features and foreground-weighted histogram*. Multimedia Tools and Applications. pp. 1-35, 2017.

- [Laaroussi 16c] K. Laaroussi, A. Saaidi, M. Masrar and K. Satori. *Scale and orientation-based background weighted histogram for human tracking*. 3D Research Journal, Vol. 7, No. 3, pp. 1-17, 2016.
- [Laaroussi 14d] K. Laaroussi, A. Saaidi, M. Masrar and K. Satori. *Video-surveillance system for tracking moving people using color interest points*. World Applied Sciences Journal, Vol. 32, No. 2, pp. 289-301, 2014.
- [Laaroussi 14e] K. Laaroussi, A. Saaidi and K. Satori. *People tracking using color control points and skin color*. Journal of emerging technologies in web intelligence, Vol. 6, No. 1, pp. 94-100, 2014.
- [Lee 05a] D. S. Lee. *Effective Gaussian Mixture learning for video background subtraction*. Pattern Analysis and Machine Intelligence, Vol. 27, pp. 827-832, 2005.
- [Lee 01b] Y. Lee, B. You and S. Lee. *A real time color based object tracking robust to irregular illumination variations*. IEEE International Conference on Robotics and Automation, pp. 1659–1664, 2001.
- [Leichter 10] I. Leichter, M. Lindenbaum and E. Rivlin. *Mean shift tracking with multiple reference color histograms*. Computer Vision and Image Understanding, Vol. 114, No. 3, pp. 400–408, 2010.
- [Li 10a] S. Li, H. Chang and C. Zhu. *Adaptive pyramid mean shift for global real time visual tracking*. Image Vis. Comput., Vol. 28, No. 3, pp. 424–437, 2010.
- [Li 01b] B. Li, R. Chellappa, Q. Zheng and S. Z. Der. *Model-based temporal object verification using video*. Image Processing, IEEE Transactions on, vol. 10, no. 6, pp. 897–908, 2001.
- [Li 13c] X. Li, W. Hu, C. Shen, Z. Zhang, A. Dick and A. V. D. Hengel. *A survey of appearance models in visual object tracking*. ACM Trans. Intel. Systems and Technology. Vol. 4, No. 4, 2013.
- [Liang 07] D. Liang, Q. Huang, S. Jiang, H. Yao and W. Gao. *Mean shift blob tracking with adaptive feature selection and scale adaptation*. Proc. International Conference Image Processing, Vol. 3, pp. III-369, 2007.
- [Lim 11] H. Lim and D. Kang. *Object tracking system using a VSW algorithm based on color and point features*. EURASIP Journal on Advances in Signal Processing, September 2011.
- [LITIV 17] The Litiv Datasets, 2017.
<http://www.polymtl.ca/litiv/en/vid/>
- [Liu 06] H. Liu, J. Yang and Z. Wei. *Moving Object Tracking and Vision Navigation Based on Selective Attention Mechanism*. Proc of the IEEE International Conference on Robotics and Biomimetics, pp. 1500-1505,

- 2006.
- [Lowe 04] D. G. Lowe. *Distinctive image features from scale-invariant keypoints*. International journal of computer vision, Vol. 60, No. 2, pp. 91–110, 2004.
- [Lu 11] H. Lu, Q. Zhou, D. Wang and R. Xiang. *A co-training framework for visual tracking with multiple instance learning*. IEEE Conf Automatic Face and Gesture Recognition, pp. 539-544, 2011.
- [Lucas 81] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In IJCAI, Vol. 81, No. 1, pp. 674-679, 1981.
- [MacCormick 00] J. MacCormick and A. Blake. *A probabilistic exclusion principle for tracking multiple objects*. International Journal of Computer Vision, Vol. 39, No. 1, pp. 57–71, 2000.
- [Mansouri 02] A. Mansouri. *Region tracking via level set PDEs without motion computation*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 7, pp. 947–961, 2002.
- [McKenna 00] S. J. McKenna, S. Jabri, Z. Duric, H. Wechsler and A. Rosenfeld. *Tracking groups of people*. Computer Vision and Image Understanding, Vol. 80, No. 1, pp. 42-56, 2000.
- [Mennesson 10] J. Mennesson, C. Saint-Jean and L. Mascarilla. *De nouveaux descripteurs de Fourier géométriques pour l'analyse d'images couleur*. In Reconnaissance des Formes et Intelligence Artificielle, pp. 599-606, 2010.
- [Mikolajczyk 05a] K. Mikolajczyk and C. Schmid. *A performance evaluation of local descriptors*. IEEE transactions on pattern analysis and machine intelligence, Vol. 27, No. 10, pp. 1615-1630, 2005.
- [Mikolajczyk 01b] K. Mikolajczyk and C. Schmid. *Indexing based on scale invariant interest points*. Proc In International Conference on Computer Vision, Vancouver, Canada, July 2001.
- [Mikram 08] Mounia Mikram. *Suivi d'objets dans une séquence d'images par modèle d'apparence: conception et évaluation*. Université de Bordeaux I, Décembre 2008.
- [Montesinos 98] P. Montesinos, V. Gouet and R. Deriche. *Differential invariants for color images*. International Conference on Pattern Recognition. pp 838–840, 1998.
- [Mora 03] P. Mora and S. Héas. *Du joueur de jeux vidéo à l'e-sportif: vers un professionnalisme florissant de l'élite*. La pratique du jeu vidéo: réalité ou virtualité, pp. 131-146, 2003.

- [Moravec 77] H. Moravec. *Towards automatic visual obstacle avoidance*. Proc of the 5th International Joint Conference on Artificial Intelligence, pp. 584, August 1977.
- [Mukundan 98] R. Mukundan and K. R. Ramakrishnan. *Moment functions in image analysis: theory and applications*. Singapore: World Scientific, vol.100, 1998.
- [Mughadam 97] B. Mughadam and A. Pentland. *Probabilistic visual learning for object representation*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, pp. 696–710, 1997.
- [Murray 94] D. Murray and A. Basu. *Motion tracking with an active camera*. IEEE Trans. Pattern Anal. Mach. Intell., Vol.16, No.5, pp.449-459, 1994.
- [Ning 10a] J. Ning, L. Zhang, D. Zhang and C. Wu. *Robust mean shift tracking with corrected background-weighted histogram*. IET Computer Vision, Vol. 6, No. 1, pp. 62-69, 2010.
- [Ning 09b] J. Ning, L. Zhang, D. Zhang and C. Wu. *Robust object tracking using joint color-texture histogram*. Int J Pattern Recog Artif Intell Vol. 23, No. 7, pp. 1245–1263, 2009.
- [Ning 12c] J. Ning, L. Zhang, D. Zhang and C. Wu. *Scale and orientation adaptive mean shift tracking*. Computer Vision, IET, Vol. 6, No. 1, pp. 52–61, 2012.
- [Nummiaro 03] K. Nummiaro, E. Koller-Meier and L. Van Gool. *An adaptive color-based particle filter*. Image and vision computing, Vol. 21, No. 1, pp. 99–110, 2003.
- [Ojala 96a] T. Ojala, M. Pietikäinen and D. Harwood. *A comparative study of texture measures with classification based on feature distributions*. Pattern Recognition, Vol. 29, No. 1, pp. 51-59, 1996.
- [Ojala 02b] T. Ojala, M. Pietikäinen and T. Maenpaa. *Multiresolution gray-scale and rotation invariant texture classification with local binary patterns*. IEEE Trans Patt Anal Mach Intell, Vol. 24, No. 7, pp. 971–987, 2002.
- [Oliver 00] N. M. Oliver, B. Rosario and A. P. Pentland. *A bayesian computer vision system for modeling human interactions*. Pattern Analysis and Machine Intelligence, Vol. 22, pp. 831-843, 2000.
- [Pan 16] Z. Pan, S. Liu and W. Fu. *A review of visual moving target tracking*. Multimedia Tools and Applications, pp. 1-30, 2016.
- [Paragios 99a] N. Paragios and R. Deriche. *Geodesic active regions for motion estimation and tracking*. IEEE International Conference on Computer Vision, Vol.1, pp. 688-694, 1999.
- [Paragios 00b] N. Paragios and R. Deriche. *Geodesic active contour and level sets for*

- the detection and tracking of moving objects*. IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 22, No. 3, pp. 266–280, 2000.
- [Paragios 02c] N. Paragios and R. Deriche. *Geodesic active regions and level set methods for supervised texture segmentation*. International Journal of Computer Vision, Vol. 46, No. 3, pp. 223-247, 2002.
- [Parameswaran 06] V. Parameswaran, V. Ramesh and I. Zoghlami. *Tunable kernels for tracking*. In: Proc. IEEE Comput. Soc. Conf. on Computer Vision and Pattern Recognition, pp. 2179–2186, 2006.
- [Park 04] S. Park and J. K. Aggarwal. *A hierarchical bayesian network for event recognition of human actions and interactions*. Multimed. Syst, Vol. 10, No. 2, pp. 164–179, 2004.
- [Patella 99] M. Patella. *Similarity Search in Multimedia Databases*. PhD thesis, University of Bologna, Italy. 1999
- [Peng 05] N. S. Peng, J. Yang, Z. Liu and F.C. Zhang. *Automatic selection of kernel-bandwidth for mean shift object tracking*. Journal of Software, Vol. 16, No. 9, pp. 1542–1550, 2005.
- [Peterfreund 00] N. Peterfreund. *Robust tracking of position and velocity with Kalman snakes*. IEEE Trans. Pattern Anal. Machine Intell., Vol. 22, pp. 564–569, 2000.
- [Persoon 77] E. Persoon and K. Fu. *Shape discrimination using fourier descriptors*. IEEE Transactions on systems, man, and cybernetics, Vol. 7, No. 3, pp. 170-179, 1977.
- [Pérez 02] P. Pérez, C. Hue, J. Vermaak and M. Gangnet. *Color-based probabilistic tracking*. Computer vision-ECCV, pp. 661-675, 2002.
- [Phung 05] S. L. Phung, A. Bouzerdoum and D. Chai. *Skin segmentation using color pixel classification: analysis and comparison*. IEEE transactions on pattern analysis and machine intelligence, Vol. 27, No. 1, pp. 148-154, 2005.
- [Pietikäinen 11] M. Pietikäinen, A. Hadid, G. Zhao and T. Ahonen. *Computer vision using local binary patterns*, 40, Springer Science & Business Media, 2011.
- [Polat 06] E. Polat and M. Ozden. *A nonparametric adaptive tracking algorithm based on multiple feature distributions*. IEEE Trans Multimedia Vol. 8, No. 6, pp.1156-1163, 2006.
- [Prokop 92] R. J. Prokop and A. P. Reeves. *A survey of moment based techniques for unoccluded object representation and recognition*. Graphical Models and Image Processing, Vol. 54, No. 5, pp. 438-460, 1992.
- [Qingchang 07] G. Qingchang, C. Xiaojuan and C. Hongxia. *Mean-shift of variable window based on the Epanechnikov Kernel*. Proc of the International

- Conference on Mechatronics and Automation, pp. 2314– 2319, 2007.
- [Rangarajan 91] K. Rangarajan and M. Shah. *Establishing motion correspondence*. Conference Vision Graphics Image Process, Vol. 54, No. 1, pp. 56–73, 1991.
- [Ronfard 94] R. Ronfard. *Region-based strategies for active contour models*. International Journal of Computer Vision, Vol. 13, No. 2, pp. 229–251, 1994.
- [Rosales 99] R. Rosales and S. Sclaroff. *3D trajectory recovery for tracking multiple objects and trajectory guided recognition of actions*. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Vol. 2, pp. 117–123, 1999.
- [Ross 08] D. A. Ross, J. Lim, R. S. Lin and M. H. Yang. *Incremental learning for robust visual tracking*. International Journal of Computer Vision, Vol. 77, No. 1-3, pp. 125–141, 2008.
- [Rosten 10] E. Rosten, R. Porter and T. Drummond. *Faster and better: A machine learning approach to corner detection*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, Vol. 32, No. 1, pp. 105–119, 2010.
- [Sato 04] K. Sato and J. Aggarwal. *Temporal spatio-velocity transform and its application to tracking and interaction*. Computer Vision and Image Understanding, Vol. 96, No. 2, pp. 100–128, 2004.
- [Schmid 96a] C. Schmid. *Appariement d’images par invariants locaux de niveaux de gris*. Doctoral thesis, Institut National Polytechnique de Grenoble, GRAVIR, 1996.
- [Schmid 97b] C. Schmid and R. Mohr. *Local grayvalue invariants for image retrieval*. IEEE Transactions Pattern Analysis Machine Intelligence, Vol. 19, No. 5, pp. 530–534, 1997.
- [Schmid 98c] C. Schmid, R. Mohr and C. Bauckhage. *Comparing and evaluating interest points*. In IEEE International Conference on Computer Vision. pp. 230–235, January 4-7, 1998.
- [Schmid 00d] C. Schmid, R. Mohr and C. Bauckhage. *Evaluation of interest point detectors*. IEEE Transactions Pattern Analysis Machine Intelligence, 37(2):151–172, 2000.
- [Schneiders 05] S. Muller-Schneiders, T. Jager, H. S. Loos and W. Niem. *Performance evaluation of a real time video surveillance systems*. In Visual Surveillance and Performance Evaluation of Tracking and Surveillance. 2nd Joint IEEE International Workshop on, pp. 137-143, 2005.
- [Schweitzer 02] H. Schweitzer, J. W. Bell and F. Wu. *Very fast template matching*. In European Conference on Computer Vision, pp. 358–372, London, UK, 2002.

- [Senior 06] A. Senior, A. Hampapur, Y. Tian, L. Brown, S. Pankanti and R. Bolle. *Appearance models for occlusion handling*. Image and Vision Computing, Vol. 24, No. 11, pp. 1233-1243, 2006.
- [Serby 04] D. Serby, E. Koller-Meier and L. Van Gool. *Probabilistic object tracking using multiple features*. In International Conference on Pattern Recognition, pp. 184–187, 2004.
- [Sethi 87] I. K. Sethi and R. C. Jain. *Finding trajectories of feature points in a monocular image sequence*. IEEE Transactions Pattern Analysis Machine Intelligence, Vol. 9, No. 1, pp. 56–73, 1987.
- [Shafique 05] K. Shafique and M. Shah. *A non-iterative greedy algorithm for multi-frame point correspondence*. IEEE transactions on pattern analysis and machine intelligence, Vol. 27, No. 1, pp. 51-65, 2005.
- [Shi 94] J. Shi and C. Tomasi. *Good features to track*. In IEEE Conference on Computer Vision and Pattern Recognition CVPR, pp. 593–600, 1994.
- [Shih 04] F. Y. Shih and Y. T. Wu. *Fast Euclidean distance transformation in two scans using a 3*3 neighborhood*. Computer Vision and Image Understanding, Vol. 93, No. 2, pp. 195–205, 2004.
- [Sidibe 07] Dro Désiré SIDIBE. *Une technique de relaxation pour la mise en correspondance d'images : Application à la reconnaissance d'objets et au suivi du visage*. Thèse, Université Montpellier II, 2007.
- [Singh 03] SK. Singh, DS. Chauhan, M. Vatsa and R. Singh. *A Robust Skin Color Based Face Detection Algorithm*. Tamkang Journal of Science and Engineering. Vol. 6. No. 4. pp. 227-234, 2003.
- [Smith 97a] S. M. Smith and J. M. Brady. *A New Approach to Low Level Image Processing*. The International Journal of Computer Vision, Vol. 23, No. 1, pp. 45-78, 1997.
- [Smith 05b] K. Smith, D. G. Perez and J. M. Odobez. *Using particles to track varying numbers of interacting people*. Proc of the IEEE Conference on Computer Vision and Pattern Recognition CVPR, 2005.
- [Stauffer 99] C. Stauffer and W. E. L. Grimson. *Adaptive background mixture models for real-time tracking*. Proc of the IEEE Conference on Computer Vision and Pattern Recognition CVPR, pp. 246-252, 1999.
- [Stefano 05] L. Di Stefano, S. Mattoccia and F. Tombari. *ZNCC-based template matching using bounded partial correlation*. Pattern Recognit. Letters 26. pp. 2129–2134, 2005.
- [Stricker 95] M. Stricker and M. Orengo. *Similarity of color images*. In Proceedings of the SPIE Storage and Retrieval for Image and Video Databases, 1995.
- [Swain 91] M. Swain and D. Ballard. *Color indexing*. International Journal of

- Computer Vision 7, pp. 11-32, 1991.
- [Suhr 11] J. K. Suhr, H.G. Jung, G. Li, S-In Noh and J. Kim. *Background compensation for Pan-Tilt-Zoom cameras using 1-D feature matching and outlier rejection*. IEEE Trans. On circuit and systems for video technology, Vol. 21, No. 3, 2011.
- [Szeliski 97] R. Szeliski and J. Coughlan. *Spline-based image registration*. International Journal of Computer Vision, Vol. 22, No. 3, pp. 199-218, 1997.
- [Takala 07] V. Takala and M. Pietikainen. *Multi-object tracking using color, texture and motion*. IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-7, 2007.
- [Tamura 78] H. Tamura, S. Mori and T. Yamawaki. *Textural features corresponding to visual perception*. IEEE Transactions on Systems, Man, and Cybernetics, Vol. 8, No. 6, pp. 460-473, 1978.
- [Tao 02] H. Tao, H. S. Sawhney and R. Kumar. *Object tracking with bayesian estimation of dynamic layer representations*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, Vol. 24, No. 1, pp. 75–89, 2002.
- [Terzopoulos 93] D. Terzopoulos and R. Szeliski. *Tracking with kalman snakes*. In Active vision, pp. 3-20, 1993.
- [Trichet 07] R. Trichet and B. Mérialdo. *Probabilistic matching algorithm for keypoint based object tracking using a delaunay triangulation*. Proc of the International Workshop on Image Analysis for Multimedia Interactive Services, Santorini, Greece, June 2007.
- [Tsekeridou 98] S. Tsekeridou and I. Pitas. *Facial feature extraction in frontal views using biometric analogies*. Proc of the IX European Signal Processing Conference, Vol. I, pp. 315-318, 1998.
- [Varcheie 11] P. Varcheie and G.-A. Bilodeau. *People tracking using a network-based PTZ camera*. Machine Vision and Applications, Vol. 22, No. 4, pp. 671–690, 2011.
- [Veenman 01] C. Veenman, M. Reinders and E. Backer. *Resolving motion correspondence for densely moving points*. IEEE Pattern Analysis and Machine Intelligence, Vol. 23, No. 1, pp. 54–72, 2001.
- [Vezhnevets 03] V. Vezhnevets, V. Sazonov and A. Andreeva. *A Survey on Pixel-Based Skin Color Detection Techniques*. In Proc. Graphicon, Vol. 3, pp. 85-92, 2003.
- [Vidal 04] R. Vidal and Y. Ma. *A unified algebraic approach to 2-D and 3-D motion segmentation*. In European Conference on Computer Vision (ECCV), pp. 1-15, 2004.

- [VOIR 17] VOIR - Vision and Observation In Robotics (2017), <http://liris.cnrs.fr/voir/>
- [Wang 14a] D. Wang, K. Chen, W. Sun, S. Yu and H. Wang. *Efficient Mean Shift tracking using an improved weighted histogram scheme*. KSII Transactions on Internet and Information Systems, Vol. 8, No. 6, pp. 1964-1981, 2014.
- [Wang 15b] D. Wang, H. Lu and C. Bo. *Visual tracking via weighted local cosine similarity*. IEEE transactions on cybernetics, Vol. 45, No. 9, pp. 1838-1850, 2015.
- [Wang 16c] D. Wang, Y. Shi and W. Sun. *A novel background-weighted histogram scheme based on foreground saliency for mean-shift tracking*. Multimedia Tools and Applications, Vol. 75, No. 17, pp. 10271–10289, 2016.
- [Wang 08d] J. Q. Wang and Y. S. Yagi. *Integrating color and shape-texture features for adaptive real-time object tracking*. IEEE Trans. on Image Processing, Vol. 17, No. 2, pp. 235-240, 2008.
- [Weng 06] S. K. Weng, C. M. Kuo and S. K. Tu. *Video object tracking using adaptive Kalman filter*. Journal of Visual Communication and Image Representation, Vol. 17, No. 6, pp. 1190-1208, 2006.
- [Williams 05] O. Williams, A. Blake and R. Cipolla. *Sparse bayesian learning for efficient visual tracking*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, pp. 1292–1304, Aug 2005.
- [Wolsztynski 06] E. Wolsztynski. *Critère d'entropie pour l'estimation semi paramétrique*. PhD thesis, Université de Nice - Sophia Antipolis, 2006.
- [Wren 97] C. R. Wren, A. Azarbayejani, T. Darrell and A. Pentland. *Pfinder: real-time tracking of the human body*. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, pp. 780-785, 1997.
- [Wu 15] Y. Wu, J. Lim, M.-H. Yang. *Object tracking benchmark*. IEEE Trans Pattern Anal Mach Intell 37(9):1834-1848, 2015.
- [Xiang 08] S. Xiang, F. Nie, Y. Song and C. Zhang. *Contour graph based human tracking and action sequence recognition*. Pattern Recognition, Vol. 41, No. 12, pp. 3653-3664, 2008.
- [Xiaorong 12] P. Xiaorong and Z. Zhihu. *A more robust Mean Shift tracker on joint color-CLTP histogram*. Internat Journal of Image, Graphics and Signal Processing, Vol. 4, No. 12, pp. 34-42, 2012.
- [Xu 02] N. Xu and N. Ahuja. *Object contour tracking using graph cuts based active contours*. Proc of the International Conference on Image Processing, Vol. 3, pp. III–277, 2002.

- [Yang 99a] M. H Yang and N. Ahuja. *Gaussian mixture model for human skin color and its applications in image and video databases*. Proc of International Society for Optics and Photonics In *Electronic Imaging'99*, pp. 458-466, 1999.
- [Yang 05b] C. Yang, R. Duraiswami and L. Davis. *Efficient mean-shift tracking via a new similarity measure*. Proc of IEEE Computer Society Conference In Computer Vision and Pattern Recognition CVPR, Vol. 1, pp. 176–183, 2005.
- [Yang 04c] T. Yang, S. Z. Li, Q. Pan and J. Li. *Real-time and accurate segmentation of moving objects in dynamic scene*. Proc of the ACM 2nd international workshop on Video surveillance & sensor networks, pp. 136-143, 2004.
- [Yang 13d] F. Yang, H. Lu and M.-H. Yang. *Learning structured visual dictionary for object tracking*. Image and Vision Computing, Vol. 31, No. 12, pp. 992–999, 2013.
- [Yang 11e] H. Yang, L. Shao, F. Zheng, L. Wang and Z. Song. *Recent advances and trends in visual tracking: A review*. Neurocomputing, Vol. 74, No. 18, pp. 3823-3831, 2011.
- [Yilmaz 06a] A. Yilmaz, O. Javed and M. Shah. *Object tracking: a survey*. ACM Comput. Surveys Vol. 38, No. 4, 2006.
- [Yilmaz 04b] A. Yilmaz, X. Li and M. Shah. *Contour-based object tracking with occlusion handling in video acquired using mobile cameras*. IEEE Trans. Patt. Analy. Mach. Intell. Vol. 26, No. 11, pp. 1531–1536, 2004.
- [Yinghong 11] M. Yinghong, M. Zhi-Hong, J. Wenyan, L. Chengliu, Y. Jiawei and S. Mingui. *Magnetic hand tracking for human-computer interface*. IEEE Trans Magn, Vol. 47, No. 5, pp. 970–973, 2011.
- [Yu 15] W. Yu, X. Tian, Z. Hou, Y. Zha and Y. Yang. *Multi-scale mean shift tracking*. Computer Vision, IET, Vol. 9, No. 1, pp. 110-123, 2015.
- [Zhao 06a] Q. Zhao, J. Kang, H. Tao and W. Hua. *Part-based human tracking in a multiple cues fusion framework*. International Conference on Pattern Recognition, Vol. 1, pp. 450-455, 2006.
- [Zhao 08b] C. Zhao, A. Knight and I. Reid. *Target tracking using mean shift and affine structure*. Proc of the 19th International Conference In Pattern Recognition, pp. 1–5, 2008.
- [Zhao 05c] Q. Zhao and H. Tao. *Object tracking using color correlogram*. Proc of IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance, pp. 263-270, 2005.
- [Zhao 10d] Q. Zhao, Z. Yang and H. Tao. *Differential earth mover's distance with its applications to visual tracking*. Pattern Analysis and Machine

- Intelligence, IEEE Transactions on, Vol. 32, No. 2, pp. 274–287, 2010.
- [Zhang 10a] X. Zhang, Y. M. Dai, Z. W. Chen and H. X. Zhang. *An improved Mean Shift tracking algorithm based on color and texture feature*. IEEE Conf Wavelet Analysis and Pattern Recognition ICWAPR, pp. 38–43, 2010.
- [Zhang 13b] K. Zhang, L. Zhang, M. H. Yang and D. Zhang. *Fast tracking via spatio-temporal context learning*. arXiv preprint arXiv 1311.1939, 2013.
- [Zhong 14] W. Zhong, H. Lu and M. H. Yang. *Robust object tracking via sparse collaborative appearance model*. IEEE Transactions on Image Processing, Vol. 23, No. 5, pp. 2356–2368, 2014.
- [Zhou 01a] Q. Zhou and J. Aggarwal. *Tracking and classifying moving objects from video*. Performance Evaluation of Tracking Systems Workshop, 2001.
- [Zhou 11b] Q.-H. Zhou, H. Lu H and M.-H. Yang. *Online multiple support instance tracking*. IEEE Conf Automatic Face and Gesture Recognition, pp. 545–552, 2011.
- [Zhou 09c] H. Zhou, Y. Yuan and C. Shi. *Object tracking using SIFT features and mean shift*. Computer vision and image understanding, Vol. 113, No. 3, pp. 345–352, 2009.
- [Zhu 96] S. C. Zhu and A. Yuille. *Region competition: Unifying snakes, region growing, and Bayes/MDL for multiband image segmentation*. IEEE transactions on pattern analysis and machine intelligence, Vol. 18, No. 9, pp. 884–900, 1996.
- [Zivkovic 04a] Z. Zivkovic and B. Kröse. *An EM-like algorithm for color histogram based object tracking*. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, vol. 1, pp. 798–803, 2004.
- [Zivkovic 06b] Z. Zivkovic and F. van der Heijden. *Efficient adaptive density estimation per image pixel for the task of background subtraction*. Pattern recognition letters, Vol. 27, pp. 773–780, 2006.

Annexe A

Algorithmes de Harris Couleur et de RANSAC et la mesure de corrélation ZNCC

Dans cette annexe, nous présentons les outils fondamentaux utilisés dans notre première contribution [Laaroussi 14d, Laaroussi 14e], à savoir l'algorithme de Harris couleur [Harris 88] pour la détection des points d'intérêt (section A.1), la mesure de corrélation *ZNCC* [Chambon 11, Stefano 05] (section A.2) et l'algorithme *RANSAC* [Fischler 81] (section A.3).

A.1 Détecteur de Harris Couleur

Les points d'intérêt sont les points les plus identifiables que les autres pixels de l'image où des changements significatifs se produisent. Plusieurs détecteurs ont été développés, à titre d'exemple Schmid et Mohr [Schmid 00d] ont comparé les performances de plusieurs d'entre eux. Le détecteur le plus populaire est le détecteur de Harris [Montesinos 98, Harris 88] avec ses adaptations [Montesinos 98, Schmid 97b, Mikolajczyk 01b, Lowe 04]. Ce détecteur étant seulement applicable aux images en niveau de gris, Montesinos et al. [Montesinos 98] l'ont généralisé pour les images couleurs. Ce détecteur est basé sur une fonction d'auto-corrélation. Selon les comparaisons faites par Gouet et Boujemaa [Gouet 02a], le détecteur de Harris apparaît comme étant le plus sûr parmi les détecteurs des points d'intérêt couleurs connus, et ce par rapport aux changements d'illumination, aux bruits, aux rotations et aux changements de point de vue. Les principales étapes de l'approche de Harris couleur, sont présentées dans l'algorithme suivant :

Algorithme : Détection des points d'intérêt couleurs par l'algorithme de Harris.
Début Entrée : une image I - Déterminer les gradients I_u et I_v de l'image de la façon suivante : $I_u = I \otimes (-1, 0, 1) \text{ et } I_v = I \otimes (-1, 0, 1)^T \text{ avec } \otimes \text{ signe de convolution}$ - Déterminer les gradients $I_u I_u$ et $I_v I_v$ de l'image, tel que :

$$I_u I_u = I_u^2 \text{ et } I_v I_v = I_v^2$$

3. Déterminer les coins de Harris pour chaque pixel de l'image de la façon suivante :

- Calculer la matrice de Harris couleur $M_{RGB}(x, y)$ de taille 2×2 pour un pixel (x, y) de l'image I :

$$M_{RGB}(x, y) = \begin{bmatrix} A & C \\ C & B \end{bmatrix}$$

Avec $A = R_x^2 + G_x^2 + B_x^2$, $B = R_y^2 + G_y^2 + B_y^2$ et $C = R_x R_y + G_x G_y + B_x B_y$

R , G et B sont les trois composantes Rouge, Vert et Bleu.

- Calculer la trace et le déterminant de la matrice de Harris $M_{RGB}(x, y)$:

$$trace(M_{RGB}) = A + B - C^2$$

$$det(M_{RGB}) = A + B$$

- Calculer la réponse F du détecteur de Harris :

$$F = det(M_{RGB}) - k' * trace^2(M_{RGB})$$

Avec $k' \in [0.04 \ 0.06]$ (est une valeur fixée par Harris).

4. Extraire les maxima locaux positifs dans un voisinage 3×3 de la réponse F (c'est-à-dire mettre à 0 tous les points négatifs ou dont la valeur est inférieure à celle des huit voisins).

5. Extraire les n' meilleurs points de Harris, avec n' est un paramètre à choisir.

Sorties : les points d'intérêt couleurs.

Fin

A.2 Mesure de corrélation ZNCC

Après la détection des points d'intérêt, on fait appel à l'une des méthodes de mise en correspondance qui cherchent les pixels qui se ressemblent le plus en utilisant des mesures de corrélation. La recherche du correspondant est locale et s'effectue dans une zone de recherche. De très nombreuses mesures ont été proposées dans la littérature pour prendre en compte les différentes difficultés rencontrées lors de cette étape comme les bruits et les occultations. La robustesse de ces mesures est classée selon leurs capacités d'appariement dans les cas des grands déplacements et des changements de luminosité.

Dans ce travail, nous avons choisi la mesure de corrélation ZNCC (Zero mean Normalized Cross Correlation) [Stefano 05, Chambon 11] qui est une mesure de corrélation croisée

centrée normalisée. Sa formule correspond à une sommation des produits entre les éléments des fenêtres de corrélation, normalisée par le produit des moyennes quadratiques calculées à partir de chacune de ces fenêtres. Cette fonction fournit une mesure dans l'intervalle $[-1,1]$ et la ressemblance entre les deux points X et Y est élevée quand $ZNCC(X,Y)$ tend vers à 1 [Stefano 05, Chambon 11].

Soient $X = (x_i)_{i=1\dots n}$ et $Y = (y_i)_{i=1\dots n}$ les coordonnées des points X et Y . La fonction $ZNCC$ utilise la formule suivante pour mesurer la corrélation entre les points des images :

$$ZNCC(X,Y) = \frac{\sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{X})^2 (y_i - \bar{Y})^2}} \quad (A.1)$$

Avec $\bar{X}$ et $\bar{Y}$ sont les moyennes de X et de Y respectivement tel que :

$$\bar{X} = \frac{\sum_{i=1}^n x_i}{n}, \bar{Y} = \frac{\sum_{i=1}^n y_i}{n} \quad (A.2)$$

Dans la pratique, on fixe un seuil s (en général $s = 0.5$) [Stefano 05, Chambon 11] tel que les couples (X,Y) sont considérés des vrais correspondants si la contrainte suivante est vérifiée :

$$ZNCC(X,Y) > s \quad (A.3)$$

Mais le problème qui se pose est la production des faux appariements lorsque on applique la contrainte A.3, alors dans ce cas on doit réaliser l'étape de régularisation afin d'éliminer les faux appariements.

A.3 Algorithme de RANSAC

RANSAC (RANdom SAmple Consensus) [Fischler 81] est une méthode itérative pour estimer les paramètres d'un modèle mathématique d'une manière robuste à partir d'un ensemble de données observées et face aux faux appariements dont l'erreur par rapport à l'entité trouvée est supérieure à un seuil. Soit un ensemble E de n appariements, soit X l'entité cherchée et paramétrée par m paramètres. La technique de *RANSAC* peut être décrite par l'algorithme suivant :

Algorithme : RANSAC**Début**

$ner \leftarrow 1$ (Représente le nombre des essais restants pour trouver les vrais appariements)

$ne \leftarrow 0$ (Compteur du nombre des essais)

$d \leftarrow 0.001$ (Distance minimale entre un point est l'entité X)

$dim \leftarrow 0$ (Dimension de l'ensemble des vrais appariements)

$p \leftarrow 0.99$ (Probabilité d'avoir choisi au moins un échantillon de m points de données)

Tant que $ner > ne$ faire

 Début

 1- Choisir aléatoirement un sous ensemble E_1 de E qui contient m appariements

 2- Calculer l'entité X en utilisant les m appariements de l'ensemble E_1

 3- Déterminer le sous ensemble $\bar{E}_1$ de E dont l'erreur par rapport à X est inférieur à un seuil d

 4- Si $card(\bar{E}_1) > dim$ alors

 Début

$dim \leftarrow card(\bar{E}_1)$

$Y \leftarrow X$

$VA \leftarrow \bar{E}_1$

$ner \leftarrow \frac{\log(1-p)}{\log(1-(\frac{dim}{n})^m)}$

 Fin

$ne \leftarrow ne + 1$

 Fin

$X \leftarrow Y$

$\bar{E}_1 \leftarrow VA$

Fin

Cet algorithme retourne l'entité estimée X et l'ensemble de vrais appariements $\bar{E}_1$.

A.4 Conclusion

Dans cette annexe, nous avons présenté l'algorithme de Harris couleur et celui de RANSAC, ainsi que, nous avons analysé la mesure de corrélation ZNCC pour calculer la ressemblance entre les points des images traitées.

Annexe B

Algorithme Mean Shift pour le suivi d'objets

B.1 Introduction

La procédure Mean Shift est une approche largement utilisée pour le suivi d'objets par forme géométrique et caractéristiques globales [Comaniciu 03c]. Elle utilise la distribution de probabilité des couleurs ou des niveaux de gris à l'intérieur de la forme géométrique sous forme des histogrammes afin de représenter l'apparence de l'objet [Comaniciu 03c]. Cette approche a été initialement développée par Fukunaga et al. [Fukunaga 75] pour le regroupement des données, et plus tard Cheng [Cheng 95] l'introduit dans le domaine de la vision par ordinateur. Bradski [Bradski 98] modifie la méthode Mean Shift et propose l'algorithme CAMSHIFT (Continuously Adaptive Mean SHIFT) pour le suivi de visage. Cette théorie est devenue populaire dans la communauté de la vision par ordinateur après sa réussite dans les applications de segmentation d'images [Comaniciu 02b] et le suivi d'objets en temps réel en utilisant une fonction de densité de probabilité de l'objet [Comaniciu 03c]. Les auteurs de [Comaniciu 03c] utilisent un histogramme pondéré calculé sur une région elliptique pour représenter l'apparence de l'objet suivi. Au lieu de réaliser une recherche exhaustive pour localiser l'objet d'intérêt, ils utilisent la procédure Mean Shift. Dans cette annexe nous allons présenter l'algorithme Mean Shift pour le suivi d'objets.

B.2 Procédure Mean Shift pour le suivi d'objets

La technique Mean Shift consiste à rechercher la position de la cible dans l'image courante à partir de sa distribution de couleur qui est généralement approximée par un histogramme normalisé. D'un point de vue fonctionnel, la procédure Mean Shift proposée dans [Comaniciu 03c] repose sur deux parties principales, à savoir la représentation et la localisation de la cible.

B.2.1 Représentation de la cible

Pour caractériser la cible, il faut d'abord choisir un espace de caractéristique. Le modèle de référence de la cible est représenté par sa fonction de densité de probabilité FDP $\hat{q}$ dans l'espace de caractéristique. Par exemple, pour Comaniciu et al. [Comaniciu 03c], le modèle de

référence de la cible est représenté par un histogramme de couleurs calculé dans une région elliptique. Le modèle de référence de la cible peut être considéré comme centré à la position 0. Dans l'image suivante, le modèle de la cible candidate est défini à la position y , et se caractérise par la fonction de densité de probabilité FDP $\hat{p}(y)$. Les deux densités de probabilités $\hat{q}$ et $\hat{p}(y)$ doivent être estimées à partir des données. Pour satisfaire le faible coût du calcul imposé par le traitement en temps réel les densités discrètes, c'est-à-dire, les histogrammes de m -classe devraient être utilisés. Le modèle de la cible est représenté par une région elliptique. Soit $\{x_i^*\}_{i=1 \dots n}$ l'ensemble des coordonnées normalisées de n pixels du modèle de la cible centré à 0. Le modèle de la cible $\hat{q}$ et la probabilité $\hat{q}_u$ de la caractéristique, $u = 1 \dots m$, dans ce modèle sont calculées par :

$$\begin{cases} \hat{q} = \{\hat{q}_u\}_{u=1 \dots m} \\ \hat{q}_u = C \sum_{i=1}^n k(\|x_i^*\|^2) \beta[l(x_i^*) - u] \end{cases} \quad (\text{B.1})$$

Où m est la taille de l'histogramme (c-à-d le nombre de classes utilisées), β est la fonction Kronecker où la somme de ces fonctions pour, $u = 1 \dots m$, est égal à 1 [Comanciu 03c]. La fonction $l : R^2 \rightarrow \{1, \dots, m\}$ associe à chaque pixel à la position x_i^* l'indice $l(x_i^*)$ de sa classe dans le m -histogramme. $k(x)$ ¹ est une fonction de profil convexe, monotone et décroissante, attribuant un poids plus faible aux coordonnées éloignées du centre de la région. L'utilisation de ces poids augmente la robustesse de l'estimation de la densité de probabilité puisque les pixels périphériques sont les moins fiables, car souvent affectés par des occultations ou l'interférence de l'arrière-plan. $\|x_i^*\|$ est la distance entre le pixel x_i^* et le centre de la région. La constante de normalisation C est obtenue en imposant $\sum_{u=1}^m \hat{q}_u = 1$ [Comanciu 03a], d'où :

$$C = \frac{1}{\sum_{i=1}^n k(\|x_i^*\|^2)} \quad (\text{B.2})$$

De même, soit $\{x_i\}_{i=1 \dots n_h}$ l'ensemble des coordonnées normalisées de n_h pixels du modèle de la cible candidate centré à y dans l'image courante. En utilisant le même profil de noyau $k(x)$ mais avec un rayon h . Le modèle de la cible candidate $\hat{p}(y)$ et la probabilité $\hat{p}_u(y)$ de la caractéristique, $u = 1 \dots m$, dans ce modèle sont calculées par :

¹. Profil du noyau K est défini par la fonction $k: [0 \quad +\infty[\rightarrow R$ sachant que $K(x) = k(\|x\|^2)$ [Comanciu 03c].

$$\begin{cases} \hat{p}(y) = \{\hat{p}_u(y)\}_{u=1 \dots m} \\ \hat{p}_u(y) = C_h \sum_{i=1}^{n_h} k\left(\left\|\frac{y - x_i}{h}\right\|^2\right) \beta[l(x_i) - u] \end{cases} \quad (\text{B.3})$$

Où C_h est une constante de normalisation définie par :

$$C_h = \frac{1}{\sum_{i=1}^{n_h} k\left(\left\|\frac{y - x_i}{h}\right\|^2\right)} \quad (\text{B.4})$$

La similarité entre l'histogramme du modèle de la cible et celui de la cible candidate est mesurée en utilisant une métrique dérivée du coefficient de Bhattacharyya [Kailath 67] comme suit :

$$\rho[\hat{p}(y), \hat{q}] = \sum_{u=1}^m \sqrt{\hat{p}_u(y) \hat{q}_u} \quad (\text{B.5})$$

Où $\hat{q}_u$ et $\hat{p}_u(y)$ représentent, respectivement, les distributions du modèle de la cible et celle de la cible candidate. La distance de Bhattacharyya [Kailath 67] est alors définie par :

$$d[\hat{p}(y), \hat{q}] = \sqrt{1 - \rho[\hat{p}(y), \hat{q}]} \quad (\text{B.6})$$

B.2.2 Localisation de la cible

Pour trouver la position correspondante à la cible dans l'image courante, la distance définie par l'équation (B.6) ci-dessus doit être minimisée comme une fonction de y , ce qui est équivalent à maximiser le coefficient de Bhattacharyya défini par l'équation (B.5). Cette maximisation est effectuée efficacement en utilisant les itérations de l'algorithme Mean Shift comme illustré sur la figure B.1 ci-dessous. La recherche de la nouvelle position de la cible dans l'image courante commence à la position $\hat{y}_0$ de la cible dans l'image précédente. Ainsi, les probabilités $\{\hat{p}_u(\hat{y}_0)\}_{u=1 \dots m}$ de la cible candidate à la position $\hat{y}_0$ dans l'image courante doivent être calculées en premier. En utilisant le développement de Taylor autour de la valeur $\hat{p}_u(\hat{y}_0)$, l'approximation linéaire du coefficient de Bhattacharyya (B.5) est obtenue par :

$$\rho[\hat{p}(y), \hat{q}] \approx \frac{1}{2} \sum_{u=1}^m \sqrt{\hat{p}_u(y_0) \hat{q}_u} + \frac{1}{2} C_h \sum_{i=1}^{n_h} w_i k\left(\left\|\frac{y - x_i}{h}\right\|^2\right) \quad (\text{B.7})$$

Où les poids w_i sont définis par la formule suivante :

$$\omega_i = \sum_{u=1}^m \sqrt{\frac{\hat{q}_u}{\hat{p}_u(\hat{y}_0)}} \beta[l(x_i) - u] \quad (\text{B.8})$$

Pour minimiser la distance définie dans l'équation (B.6), il suffit de maximiser le deuxième terme (B.7), le premier terme étant indépendant de y . Dans le processus itératif de l'algorithme Mean Shift, la cible se déplace de la position courante $\hat{y}_0$ vers la nouvelle position $\hat{y}_1$ selon la relation suivante :

$$\hat{y}_1 = \frac{\sum_{i=1}^{n_h} x_i \omega_i g\left(\left\|\frac{\hat{y}_0 - x_i}{h}\right\|^2\right)}{\sum_{i=1}^{n_h} \omega_i g\left(\left\|\frac{\hat{y}_0 - x_i}{h}\right\|^2\right)} \quad (\text{B.9})$$

Quand on choisit le noyau g avec le profil Epanechnikov [Qingchang 07], l'équation (B.9) est réduite à la formule suivante :

$$\hat{y}_1 = \frac{\sum_{i=1}^{n_h} x_i \omega_i}{\sum_{i=1}^{n_h} \omega_i} \quad (\text{B.10})$$

Où $g(x) = -\dot{k}(x)$ [Comanciu 03c]. En utilisant l'équation (B.10), l'algorithme Mean Shift cherche dans l'image courante la position de la région la plus semblable à l'objet avec un nombre restreint d'itérations et la nouvelle position $\hat{y}_1$ est le centre de la région cible dans l'image courante.

B.3 Algorithme Mean Shift

L'algorithme Mean Shift selon Comanciu et al. [Comanciu 03c] fonctionne comme suit : pour chaque image, il utilise la position de la cible estimée à l'image précédente comme initialisation. L'étape de l'initialisation de l'objet à suivre s'effectue manuellement. Puis, il calcule la distribution de couleur pondérée par un noyau spatial qui donne des poids plus élevés aux pixels plus près du centre de la région. Cette distribution initiale est référencée en tant que modèle de référence, et est ensuite comparée à celle des cibles candidates pour déterminer la position la plus probable de la cible dans l'image courante. Ensuite, il évalue la similarité entre l'histogramme du modèle de référence et les histogrammes des régions candidates (sélectionnés autour de la dernière position connue de l'objet à suivre). À chaque itération, le vecteur Mean Shift est calculé et la similarité entre l'histogramme du modèle de la cible et celui de la cible candidate est augmentée. Ce processus est répété jusqu'à ce que la

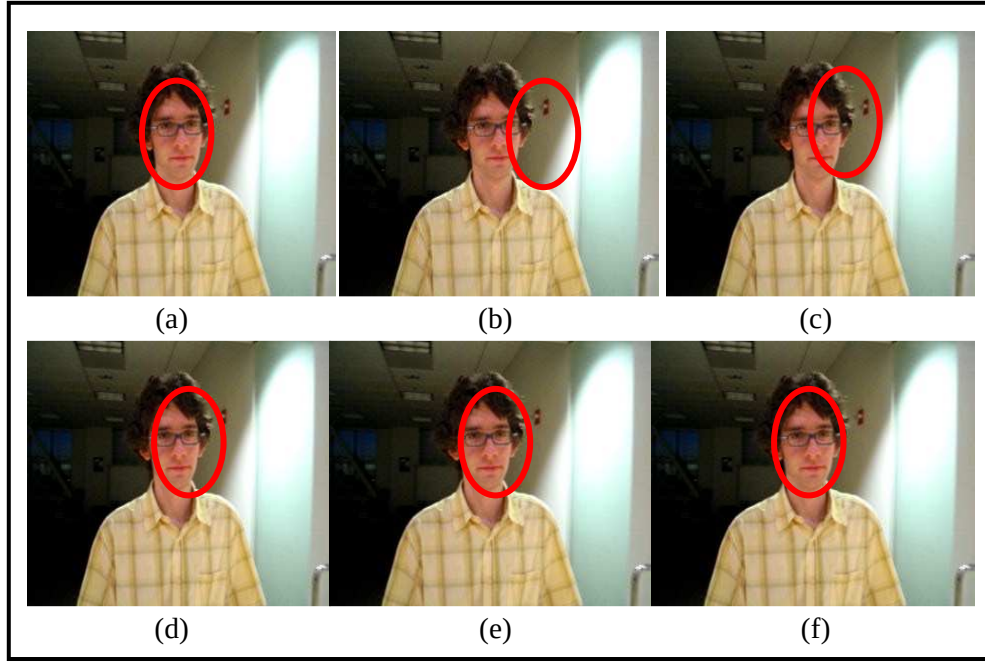


Figure B.1 : Itérations du suivi par l'algorithme Mean Shift. (B.1a) la position estimée de l'objet à l'instant $t - 1$, (B.1b) l'image à l'instant t avec une estimation initiale de position correspondant à l'estimation précédente, (B.1c), (B.1d), (B.1e) mise à jour de la position estimée avec le calcul itératif du vecteur Mean Shift, (B.1f) estimation finale de la position de l'objet à l'image t . Les images sont prises de la séquence vidéo David [Benchmark 17].

convergence soit réalisée. Comme illustré sur la figure B.1 ci-dessous, ce processus est répété généralement de quatre à six fois avant d'atteindre la convergence définie par un déplacement minimum.

En utilisant toutes les analyses et les notations précédentes, les étapes principales de l'algorithme Mean Shift pour le suivi d'objets sont comme suit :

Algorithme : Mean Shift standard pour le suivi d'objets

Nous disposons de la distribution de probabilité du modèle de la cible $\{\hat{q}_u\}_{u=1\dots m}$ selon l'équation (B.1), ainsi que l'estimation de sa position initiale $\hat{y}_0$ dans la première image.

Début

1. Initialiser la position de la cible candidate $\hat{p}(\hat{y}_0)$ dans l'image courante avec $\hat{y}_0$ et définir le nombre d'itérations $k = 0$, puis calculer la distribution $\{\hat{p}_u(\hat{y}_0)\}_{u=1\dots m}$ selon (B.3) dans l'image courante, ensuite évaluer le coefficient de Bhattacharyya $\rho[\hat{p}(\hat{y}_0), \hat{q}] = \sum_{u=1}^m \sqrt{\hat{p}_u(\hat{y}_0) \hat{q}_u}$.
2. Calculer les poids $\{\omega_i\}_{i=1\dots n_h}$ selon l'équation (B.8).

3. Trouver la nouvelle position $\hat{y}_1$ de la cible candidate selon l'équation (B.10).
4. Calculer la distribution de probabilité $\hat{p}_u(\hat{y}_1)_{u=1\dots m}$, puis évaluer le coefficient de Bhattacharyya $\rho[\hat{p}(\hat{y}_1), \hat{q}] = \sum_{u=1}^m \sqrt{\hat{p}_u(\hat{y}_1) \hat{q}_u}$.
5. Tant que $\rho[\hat{p}(\hat{y}_1), \hat{q}] < \rho[\hat{p}(\hat{y}_0), \hat{q}]$

Faire $\hat{y}_1 \leftarrow \frac{1}{2}(\hat{y}_0 + \hat{y}_1)$

Évaluer $\rho[\hat{p}(\hat{y}_1), \hat{q}]$
6. Soit $d \leftarrow \|\hat{y}_1 - \hat{y}_0\|$

Si $d < \varepsilon$ ou $k \geq N_{max}$

Arrêter l'itération.

Sinon

$\hat{y}_0 \leftarrow \hat{y}_1, k \leftarrow k + 1$ et aller à l'étape 2.

Fin

La valeur par défaut pour le critère d'arrêt ε utilisé dans l'étape 6 est $\varepsilon = 0.1$ [Comanciu 03c]. Le nombre d'itérations maximum a été fixé à $N_{max} = 20$. Dans la pratique, le nombre moyen d'itérations est beaucoup plus petit, d'environ quatre itérations et nous itérons seulement par le calcul des poids à l'étape 2, puis nous dérivons la nouvelle position à l'étape 3, et après nous testons la taille de déplacement du noyau à l'étape 6. Le coefficient de Bhattacharyya est calculé après l'achèvement de l'algorithme pour évaluer la similarité entre le modèle de la cible et celui de la cible candidate. La complexité de l'algorithme Mean Shift est très faible car elle nécessite des opérations très simples. En effet, la complexité du temps de celui-ci est donnée par $O(Tn^2)$ où T est le nombre d'itérations et n est le nombre des points dans l'ensemble de données.

B.4 Conclusion

Dans cette annexe, nous avons présenté l'algorithme Mean Shift pour le suivi d'objets. Un avantage évident de cet algorithme par rapport à la mise en correspondance standard est la suppression d'une recherche exhaustive. L'estimation de l'état de l'objet s'effectue donc dans un nombre restreint d'itérations. L'algorithme Mean Shift s'exécute en temps réel, et les résultats présentés dans les travaux de [Comanciu 03c] montrent l'intérêt de l'utilisation de cet algorithme pour le suivi d'objets en temps réel.

Publications

Articles publiés dans des journaux internationaux

- 1) Khadija Laaroussi, Abderrahim Saaidi, Mohamed Masrar, Khalid Satori. Human tracking using joint color-texture features and foreground weighted histogram. **Multimedia Tools and Applications (Springer)**, pp. 1-35, 1 August 2017. doi: 10.1007/s11042-017-5000-7
- 2) Khadija Laaroussi, Abderrahim Saaidi, Mohamed Masrar, Khalid Satori. Scale and orientation-based background weighted histogram for human tracking. **3D Research Journal (Springer)**, Vol. 7, No. 3, pp. 1-17, 15 July 2016. doi:10.1007/s13319-016-0097-4
- 3) Khadija Laaroussi, Abderrahim Saaidi, Mohamed Masrar, Khalid Satori. Human tracking based on appearance Model. **Lecture Notes in Electrical Engineering LNEE (Springer)**, Vol.1, pp. 297-305, 06 April 2016. doi:10.1007/978-3-319-30301-7_31
- 4) Khadija Laaroussi, Abderrahim Saaidi, Mohamed Masrar, Khalid Satori. Video-Surveillance System for Tracking Moving People using Color interest points. **World Applied Sciences Journal**, Vol. 32, No. 2, pp. 289-301, 2014. doi:10.5829/idosi.wasj.2014.32.02.343
- 5) Khadija Laaroussi, Abderrahim Saaidi, Khalid Satori. People Tracking using Color Control Points and Skin Color. **Journal of Emerging Technologies in Web Intelligence**, Vol. 6, No. 1, pp. 94-100, 2014. doi:10.4304/jetwi.6.1.94-100

Communications

- 1) Khadija Laaroussi, Abderrahim Saaidi, Mohamed Masrar, Khalid Satori. Scale and orientation adaptive human tracking. **2^{ème} Workshop en Imagerie, Systèmes et Applications**. 30 et 31 Octobre 2015. Faculté Polydisciplinaire de Taza.
- 2) Khadija Laaroussi, Abderrahim Saaidi, Mohamed Masrar, Khalid Satori. Human tracking based on appearance model. **The first International Conference on Mediterranean Conference on Information & Communication Technologies (MEDICT)**. 7-9 Mai 2015, Saïdia, Morocco.

- 3) Khadija Laaroussi, Abderrahim Saaidi, Mohamed Masrar, Khalid Satori. The state-of-the-Art in Visual People Tracking. **The International Conference on Wireless Technologies embedded and intelligent Systems**. 29-30 Avril 2015. Ecole nationale des Sciences appliquées de Fès en collaboration avec la Faculté des Sciences et Techniques de Fès.
- 4) Khadija Laaroussi, Abderrahim Saaidi, Mohamed Masrar, Khalid Satori. Suivi des personnes en utilisant les points de contrôle couleurs et la couleur de la peau. **The Fifth Workshop on Information Technologies and Communication**. 26-27 Décembre 2013. Faculté des Sciences Dhar El Mahraz de Fès.
- 5) Khadija Laaroussi, Abderrahim Saaidi, Mohamed Masrar, Khalid Satori. People tracking using color control points and skin color. **International Conference on Intelligent Information and Network Technology**. 13-14 Novembre 2013. Faculté des Sciences Hassan 1^{er} Settati.
- 6) Khadija Laaroussi, Abderrahim Saaidi, Mohamed Masrar, Khalid Satori. Suivi des personnes en utilisant les points de Harris couleur. **1^{er} Workshop en Imagerie, Systèmes et Applications : WoISA'13**. Faculté Polydisciplinaire de Taza, 23-24 Mai 2013.
- 7) Khadija Laaroussi, Abderrahim Saaidi, Mohamed Masrar, Khalid Satori. Méthodes de détection des personnes: État de l'art. **Journées Doctoriales en Systèmes d'Information, Réseaux et Télécommunication**. 14-15 Mai 2012. ENSIAS-Rabat.
- 8) Khadija Laaroussi, Abderrahim Saaidi, Mohamed Masrar, Khalid Satori. Méthodes numériques pour la détection des personnes. **Journée sur la Modélisation et Analyse en Mécanique Energétique**. 8 Mai 2012. Faculté Polydisciplinaire de Taza.