

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITÉ MOULOUD MAMMÈRI, TIZI-OUZOU



FACULTE DE GENIE ELECTRIQUE ET DE L'INFORMATIQUE
DEPARTEMENT D'AUTOMATIQUE

Mémoire de fin d'études

Présenté en vue de l'obtention
du Diplôme Master Académique en Automatique des systèmes

Thème:

*Étude de méthodes de suivi
d'objets: application en temps
réel*

Proposé et dirigé par :
M^{lle} O. CHILALI.

Présenté par :

FEDDAG Idir

Soutenu le : 22/10/2012

Promotion 2011/2012

Résumé

Le travail présenté dans ce mémoire s'inscrit dans le cadre du suivi d'objets dans des vidéos, et plus particulièrement, sur l'utilisation de différents types de modèles pour le Tracking. Un système de suivi (ou de tracking) peut se baser sur modèles mathématiques dont l'algorithme tient compte de la représentation de l'objet d'intérêt (représentation de la forme ou de l'apparence). Nous avons alors, organisé notre travail de tel sorte à apprendre au lecteur le principe de plusieurs méthode de suivi d'un coté, et d'un autre, de le focaliser sur de nouvelles méthodes qui se basent sur l'apparence. En fin, nous avons choisit de pencher le lecteur sur deux axe en présentant et en évaluant les résultats de deux approches adoptées pour le tracking et leurs applications en temps réel(un modèle de contour actif et un modèle d'apparence).

Mots-clés : tracking, suivi d'objet, temps réel, contour actif, modèle d'apparence, modèles génératif, modèles discriminatifs.

Sommaire

Introduction générale

I. État de l'art sur le Tracking vidéo

I.1 Introduction	4
I.2 Taxinomie des différentes méthodes de détection de mouvement	5
I.2.1 Détection sans modélisation de l'arrière-plan	6
I.2.1.1 Dérivée temporelle	7
I.2.1.2 Flot optique	8
I.2.2 Détection par modélisation locale	9
I.2.2.1 Modélisation statistique	9
I.2.2.2 Modélisation prédictive	10
I.2.3 Détection par modélisation semi-locale	11
I.2.3.1 Caractérisation par texture	11
I.2.3.2 Régularisation à posteriori	11
I.2.4 Détection par modélisation globale.....	12
I.3 Taxinomie des méthodes de suivi selon la représentation.....	12
I.3.1 Représentation des objets	13
I.3.1.1 Représentation de la forme	13
I.3.1.2 Représentation de l'apparence	15
I.3.1.3 Paramètres de sélection d'un objet	15
I.4 Suivi selon la représentation	18
I.4.1 Suivi par point	19
I.4.2 Suivi par forme géométrique de base (boite englobant)	20

Sommaire

I.4.3 Suivi par contours actifs	20
I.5 Conclusion	21

II. Les modèles d'apparence en tracking

II.1 Introduction	23
II.2 Autres modèles d'apparences pour le tracking	23
II.2.1 Histogramme multi-échelles	23
II.2.1.1 Histogramme multi-échelles cumulé	25
II.2.2 Modèles basés sur l'apprentissage statistique	25
II.2.3 Classification en apprentissage supervisé	27
II.2.3.1 Modèles discriminatifs en classification supervisée	28
II.2.3.2 Modèles génératifs en classification supervisée	31
II.3 Conclusion	32

III Tests et évaluations

III.1 Introduction	34
III.2 Modèle de contour actif pour le tracking	34
III.2.1 Application du modèle sur une vidéo	35
III.3 Un modèle d'apparence pour le tracking	41
III.4 Expérimentation et évaluation du classifieur	46
III.4.1 Robustesse à la détection pour le tracking	46
III.4.2 Évaluation du classifieur sur différentes séquence vidéo	47

Sommaire

III.4.3 Les performances du classifieur	50
III.5 Conclusion	50

Conclusion générale

Annexes

Bibliographie

Avec la vulgarisation et la généralisation de l'utilisation d'images numériques, l'analyse du mouvement dans les vidéos s'est révélée être un outil indispensable pour des applications aussi diverses que la vidéo surveillance, la compression vidéo, l'imagerie médicale, la robotique, l'interaction homme-machine, l'analyse de séquences sportives, etc. En effet, les zones de mouvement d'une séquence d'images correspondent souvent à des événements sur lesquels un système de vision doit se focaliser.

L'analyse du mouvement est un vaste sujet qui englobe un certain nombre de problématiques parmi lesquelles nous pouvons citer :

- la détection du mouvement, qui consiste à étiqueter chaque pixel d'une image suivant s'il correspond ou non à une région en mouvement dans la scène,
- la détection des objets en mouvement, c'est-à-dire la détection d'un ensemble de régions d'intérêt en mouvement dans la scène observée,
- la segmentation basée mouvement de la scène, pour laquelle chaque région de l'image ayant un mouvement distinct des autres est détectée et segmentée.

Ces problématiques citées ci-dessus sont l'une des étapes cruciales des outils automatiques de la vision par ordinateur dont l'objectif principal est d'essayer d'adapter à la machine les hautes capacités de compréhension et d'interprétation du système de vision humain, qui parvient à analyser, segmenter, détecter des objets et suivre leurs mouvements dans des scènes complexes, par l'intermédiaire d'outils mathématiques. Ces outils peuvent, alors, avoir pour rôle, soit uniquement de détecter, soit de détecter et reconnaître, soit de détecter et suivre des objets pour, par exemple, analyser le comportement ou la

trajectoire de ces objets. Ainsi, des méthodes de détection fiables et automatiques sont indispensables pour de nombreuses applications de vision par ordinateur.

Plusieurs méthodes, opérant dans ce sens, ont été proposées, testées et évaluées. En revanche, ces méthodes diffèrent, dans leurs constructions à base de modèles mathématiques. En effet, il existe plusieurs types de modèles traitant les problèmes de la vision par ordinateur relatifs au traitement d'image et au tracking, qui est au cœur du travail présenté, parmi lesquelles nous trouvons les contours actifs pour la segmentation et le suivi d'objet, les approches se basant sur l'apparence de l'objet à suivre ainsi que d'autres basés sur l'apprentissage statistique. Cependant, « *quel type de modèle convient le plus au tracking et pouvant présenter des résultats probants quand à son application en temps réel ?* »

Notre travail consiste alors, à développer un ou plusieurs types de modèle afin de les appliquer au tracking. Pour ce faire, nous avons organisé notre travail, en trois chapitres, de la manière suivante :

Le **premier chapitre** est consacré à la présentation d'un état de l'art sur les différentes techniques, outils et modèles utilisés généralement en traitement d'image et plus particulièrement en tracking. Premièrement, nous-y avons présenté une taxinomie sur les différentes méthodes de détection de mouvement selon la modélisation/ou non de l'arrière-plan. Par la suite, une autre taxinomie des méthodes de tracking, selon la représentation de la forme ou de l'apparence de l'objet à suivre, est abordée.

Dans le **deuxième chapitre**, nous nous sommes focalisé sur l'éventuelle efficacité des modèles d'apparence pour le suivi et ce en présentant quelques uns

Introduction Générale

de ces derniers tels que, les histogrammes multi-échelles et les modèle basés sur l'apprentissage statistique.

Le **troisième chapitre** de ce mémoire porte sur le choix de deux types de modèles pour une application au suivi d'objet en mouvement à savoir : *un modèle de contour actif* et *modèle d'apparence*. Ce choix n'est point arbitraire mais plutôt justifié par la possibilité de les adapter au tracking en temps réel. Nous avons alors testé et évalué ces deux types de modèles, présentés, interprété puis conclu sur leurs efficacités.

Au terme de ce mémoire, une conclusion générale est présentée.

I.1 Introduction

L'analyse des séquences d'images permet d'accéder à des informations importantes sur les objets contenus dans les scènes observées que ce soit à des fins de détection, d'estimation, de suivi ou de reconnaissance. Le suivi d'objet ou « Tracking » (en Anglais), dans ce contexte est, depuis ces dernières décennies, un thème de recherche très actif en vision par ordinateur. Le suivi correspond à l'estimation de la localisation de l'objet dans chacune des images d'une séquence vidéo, la caméra et/ou l'objet pouvant être simultanément en mouvement. Le procédé de localisation se fonde sur la reconnaissance de l'objet d'intérêt à partir d'un ensemble de caractéristiques visuelles telles que la couleur, la forme, la texture, la vitesse, etc.

Pour réaliser ce tracking, il faudrait donc, au préalable, qu'il y ait une détection d'objet appartenant ou non à la scène, puis vient la détection de mouvement par une méthode qui a pour objet de déterminer, avec rigueur, en quels points de l'image, d'une séquence vidéo, un mouvement a eu lieu.

L'algorithme régissant cet objectif fera l'objet de ce chapitre qui sera subdivisé en deux états de l'art complémentaires portant, respectivement, sur :

- *Classification des méthodes de détection de mouvement*
- *Classification des méthodes de suivi ou « Tracking »*

L'organigramme de la figure (I.1) décrit d'une manière globale l'algorithme comportant les différentes étapes qui régissent le tracking :

- **Détection de mouvement** : celle-ci doit fournir, pour tout pixel de chaque image d'entrée, une variable **quantitative** (quantité de mouvement) ou **qualitative** (booléenne), en sortie.
- **Estimation du mouvement** : ici, le résultat de la détection doit être quantitatif comme par exemple l'estimation du flot optique dont le résultat de l'estimation serait un champ de vecteur qui représente

la projection sur le plan de l'image du mouvement réel ayant lieu dans la scène.

- **Modélisation de l'arrière-plan** : ceci consiste en la mise en évidence d'un modèle de l'arrière-plan de la scène filmée sans objet mobile.
- **Soustraction de l'arrière-plan** : c'est l'opération qui suit l'étape de modélisation de l'arrière-plan afin d'obtenir une éventuelle détection de mouvement.
- **Segmentation de mouvement** : cette étape va au-delà de la détection de mouvement car, il s'agit là de segmenter chaque image en régions qui présentent une homogénéité du mouvement apparent.

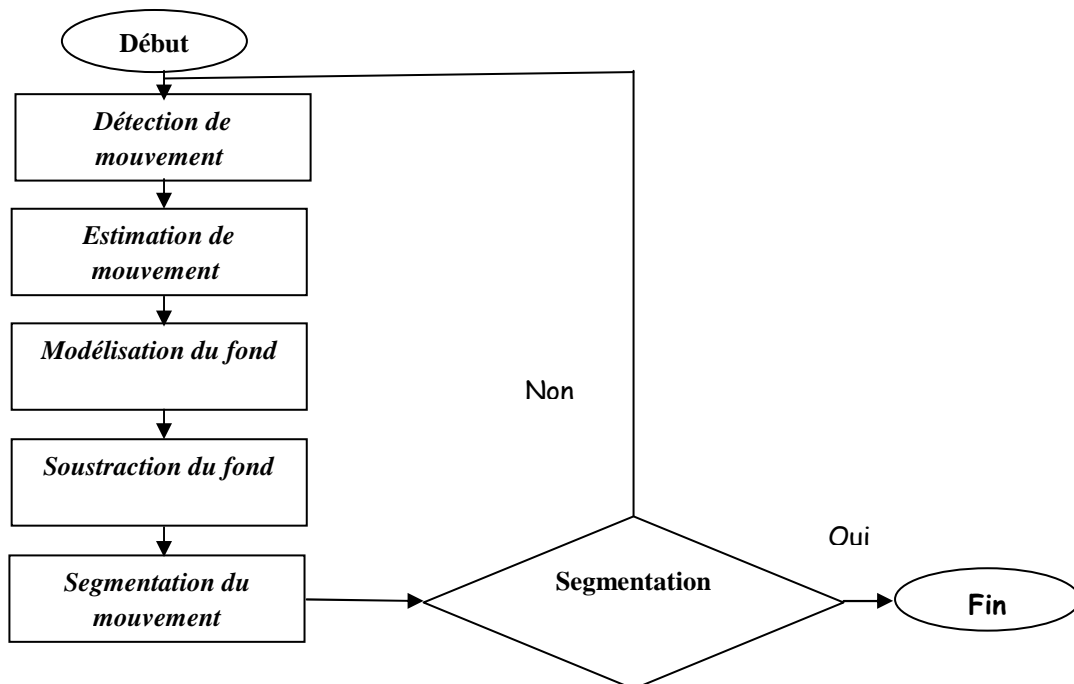


Figure I.1 Organigramme du tracking.

I.2 Taxinomie des différentes méthodes de détection de mouvement

Toutes les méthodes de détection du mouvement ont deux principales contraintes qui sont l'éclairage de la scène et le mouvement du capteur (une

caméra par exemple). En effet, la situation est différente selon que le capteur est fixe, qu'il bouge lentement ou qu'il soit mobile avec un mouvement compensé. De même, un éclairage constant, des variations d'éclairage à basse fréquence ou des changements brusques d'intensité lumineuse seront gérés différemment par un système de détection du mouvement ; rajoutant à cela les bruits éventuels dus à l'acquisition et à la numérisation de l'image.

La classification des méthodes de détection du mouvement peut se faire en fonction de la modélisation de l'arrière-plan [6], cela nous offre, hiérarchiquement, quatre grandes familles de détection qui sont représentées dans la figure I.2.

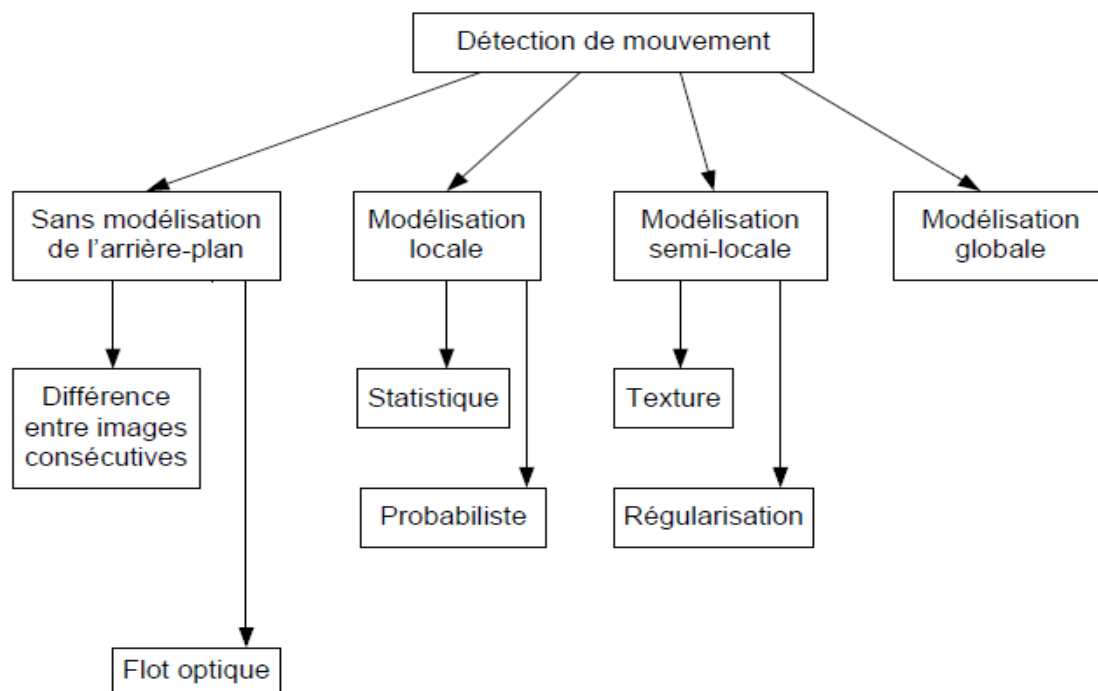


Figure I.2 Taxinomie des méthodes de détection de mouvement.

I.2.1 Détection sans modélisation de l'arrière-plan

Les méthodes de cette classe consistent à détecter en tout point de l'image une quantité mathématique qui est fonction de l'intensité ou de la couleur

de l'ensemble des pixels et qui est censé refléter l'importance du mouvement visible dans la scène.

Soit $I(t)$, l'image d'entrée à l'instant t , $I(x,y,t)$, le pixel de coordonnées $X=(x,y)$ de la même image au même instant. $I_x(x,y,t)$, $I_y(x,y,t)$ et $I_t(x,y,t)$ sont respectivement, Les dérivées spatiale et temporelle de l'image $I(t)$. Ainsi, le gradient de cette image au point X au temps t , sera noté par l'équation (I.1) :

$$\nabla I(x, y, t) = [I_x(x, y, t), I_y(x, y, t)]^T \quad (\text{I.1})$$

I.2.1.1 Dérivée temporelle

Cette méthode de détection, qui est dite aussi gradient temporel, consiste à mesurer la vraisemblance du mouvement ou bien le changement d'apparence des pixels entre deux trames consécutives, ceci revient donc à calculer la dérivée temporelle en tout point. L'estimation de la dérivée temporelle instantanée à un instant t peut être approximée par la relation suivante :

$$\forall (x, y) \in \mathbb{E} \quad \forall t > 0 \quad I_t(x, y, t) \approx |I(x, y, t) - I(x, y, t - 1)| \quad (\text{I.2})$$

Les méthodes basées sur le gradient temporel s'adaptent au changement de l'environnement mais s'avère, tout de même, peu robuste par rapport à des phénomènes tels que, la vitesse du mouvement, la taille des objets mobiles, les arrêts brusques des objets en mouvement, d'où la nécessité d'introduire à ces modèles des filtres spatio-temporels tel que le lisseur temporel.

I.2.1.2 Flot optique

La dérivée temporelle nous renseigne quantitativement sur l'aspect (vitesse et position) de tout pixel pris individuellement, le flot optique quand à lui est un champ de vecteurs à deux dimensions représentant la projection sur le plan image du mouvement réel observée (tridimensionnel). Le but étant de détecter un éventuel mouvement, ceci revient rechercher une correspondance entre deux images prise à deux temps différents afin de construire un champ de vitesse $V(x,y)$ d'une séquence d'image. La figure suivante représente quelques exemples de champ de mouvement :

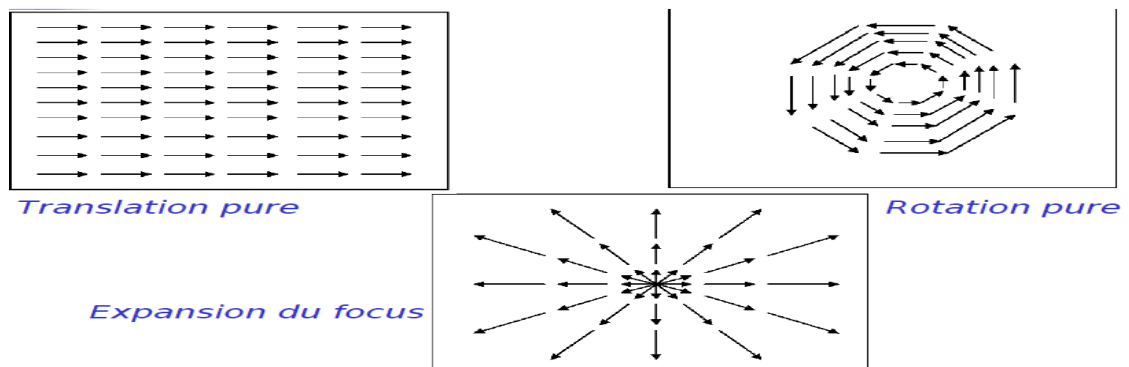


Figure I.3 Exemples de champ de mouvement.

L'un des algorithmes les plus rapides et les plus populaires est celui de Lucas et Kanade. Il s'agit de calculer au temps t , le déplacement $d=(dx; dy)^T$ du point $x = (x; y)^T$. L'hypothèse sur laquelle se base la méthode est celle de la conservation de l'intensité lumineuse régie par l'équation suivante :

$$\forall x \in E \quad \forall t > 0 \quad I(x + d, t + 1) - I(x, t) = 0 \quad (\text{I.3})$$

L'estimation du flot optique pourrait ainsi être formulée comme étant un problème d'optimisation au voisinage de $N(x)$ qui minimise le critère quadratique ε suivant :

$$\xi(\mathbf{d}) = \sum_{\mathbf{x}' \in \mathcal{N}(\mathbf{x})} [I(\mathbf{x}' + \mathbf{d}, t + 1) - I(\mathbf{x}', t)]^2 \quad (\text{I.4})$$

I.2.2 Détection par modélisation locale

Le modèle d'arrière-plan le plus simple serait une image représentant la scène dépourvue d'objets. Cette méthode présente l'avantage de détecter aussi bien les mouvements lents que les mouvements rapides. Par ailleurs, même les objets momentanément immobiles sont détectés. Cependant, ces modèles sont sensibles aux changements d'intensité lumineuse, il serait donc nécessaire d'effectuer une mise à jour de cette image d'arrière-plan. Le principe de ces méthodes se résume par une soustraction de fonds entre une trame courante et l'image constituant le fond de référence selon la relation suivante :

$$|\Delta \mathbf{I}_{t,t-1}(\mathbf{x}, \mathbf{y})| = |\mathbf{I}_t(\mathbf{x}, \mathbf{y}) - \mathbf{I}_{t-1}(\mathbf{x}, \mathbf{y})| \quad (\text{I.5})$$

Théoriquement, en l'absence du mouvement, la relation vaudra zéro.

I.2.2.1 Modélisation statistique

Le problème de la modélisation de l'arrière-plan peut être exprimé d'un point de vue statistique. Il s'agit, pour chaque pixel, d'estimer la probabilité d'y observer telle ou telle couleur (ou niveau de gris) en se basant sur un modèle censé représenter l'arrière-plan de la scène. Le modèle consiste en un ensemble de fonctions de densité de probabilité, une fonction 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 avant-plan.

Des auteurs proposent de modéliser l'intensité des points de l'arrière-plan par une distribution gaussienne [1]. 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.

L'algorithme des Mean-Shift peut être utilisé pour l'estimation de l'arrière-plan, mais il présente un problème relatif au temps de calcul puisque il en est gourmand.

I.2.2.2 Modélisation prédictive

Cette approche consiste à utiliser un filtre de Kalman ou de Wiener pour segmenter le mouvement. Pour estimer l'amplitude de ce dernier, on utilise l'écart entre l'observation et la prédiction. Ces méthodes permettent de gérer les problèmes liés aux arrière-plans non statiques [2].

D'autres auteurs utilisent un filtre de Wiener pour la segmentation locale du mouvement effectuée à l'aide d'un filtre qui permet de construire une valeur estimée \hat{Z}_t de la valeur qui devrait être observée à l'instant t , à partir d'un échantillon de N mesures $\{Z_i\}_{i=1}^N$

$$\hat{Z}_t = \sum_{i=1}^N \omega_i Z_{t-i} \quad (\text{I.6})$$

Où les ω_i sont des coefficients pondérateurs calculés de manière à minimiser l'erreur quadratique moyenne d'estimation $\xi(t)$ tel que :

$$\xi(t) = E \left[|Z_t - \hat{Z}_t|^2 \right] \quad (I.7)$$

I.2.3 Détection par modélisation semi-locale

Etant donnée la nature progressive des mouvements observés dans les séquences à analyser, et compte tenu des imprécisions dues aux caméras utilisées, certains auteurs [3] préconisent de prendre en compte l'ensemble des pixels d'un voisinage au lieu de chercher à détecter les mouvements en un point donné sans se préoccuper des pixels alentour.

I.2.3.1 Caractérisation par texture

Il consiste en le calcul en tout point, d'un vecteur caractérisant la texture à cet endroit, et en utilisant l'ensemble des vecteurs calculés comme espace de représentation des données. Pour caractériser la texture des pixels de l'image, plusieurs travaux ont été effectués tels que le codage LBP (Local Binary Patterns) [4].

I.2.3.2 Régularisation à posteriori

L'objectif de ces méthodes est l'exploitation des corrélations spatiales et temporelles entre pixels voisin afin d'obtenir des résultats plus réguliers. Après avoir segmenté l'arrière-plan à l'aide d'une modélisation locale, des auteurs considèrent que parmi les points détectés, ceux où la dérivée temporelle est importante appartiennent nécessairement à des objets mobiles [5]. Dans ce cas de figure, les filtres morphologiques sont largement utilisés pour la régularisation des zones mobiles d'intérêt.

I.2.4 Détection par modélisation globale

Pour prendre en compte la totalité de la scène dans le processus de segmentation entre avant-plan et arrière-plan, des auteurs décident de construire et de garder en mémoire k modèles de l'arrière-plan [5]. La détection de mouvement au niveau local est effectuée avec chacun des modèles, et celui qui détecte le **moins de pixels en mouvement** est retenu pour la décision finale.

Une autre façon de prendre en compte la globalité de l'espace image pour détecter les mouvements est de considérer les pixels comme des dimensions d'un espace de représentation, et les images successives comme des individus dans cet espace. Les méthodes d'analyse de données permettent alors de considérer tous les pixels de l'image dans une approche globale pour définir de nouvelles caractéristiques que l'on pourra appliquer en tout point pour y détecter d'éventuels mouvements.

La vision par ordinateur conduit à de nombreuses applications dans des domaines très divers allant de la télévision à la météorologie, de la robotique à la surveillance ou encore au médical. Ceci motive le développement de plusieurs approches de suivi d'objet.

I.3 Taxinomie des méthodes de suivi selon la représentation

La plupart des algorithmes de suivi d'objets prenant en entrée les images fournies par une camera fixe effectuent une première étape de détection de mouvement afin de déterminer parmi les pixels de l'image courante lesquels appartiennent à l'arrière-plan de la scène, et lesquels représentent des objets mobiles.

Plus formellement, le but d'une méthode de suivi d'objet est d'estimer, dans chaque image de la séquence, les paramètres de l'objet ou des objets présents dans le champ de vision de la caméra.

I.3.1 Représentation des objets

Le choix d'un modèle est un des problèmes centraux du suivi d'objet visuel. Il consiste à choisir une représentation pertinente de l'objet, donnant naissance à des primitives visuelles significatives et fiables qui décrivent le contenu visuel de l'objet tout en permettant de reconnaître cet objet au cours du temps, malgré des changements des conditions environnementales et des variations de l'apparence de l'objet le long de la séquence vidéo. Les primitives visuelles sont souvent basées sur la couleur, la forme, la texture, le mouvement ou une combinaison de ces attributs. L'un des principaux facteurs qui limite la performance des algorithmes de suivi visuel est la capacité du modèle à rester valide lorsque l'objet en question change de forme ou d'apparence. Par conséquent, deux types de représentation s'imposent : **représentation de la forme et de l'apparence**.

I.3.1.1 Représentation de la forme

Le but est de trouver une forme qui peut représenter l'objet à segmenter, en tenant compte de notre application et des informations qu'on veut avoir. Par exemple : un point suffirait pour représenter un avion dans l'air, par contre un contour est nécessaire pour la représentation d'un corps humain en mouvement.

- **Points** : L'objet est représenté par un point, qui est le centre de gravité de la surface de l'objet, ou par un ensemble de points (Figure I.4 (a), (b)). Cette représentation est utilisée, généralement, lorsqu'on a affaire avec un objet qui occupe une taille réduite dans l'image, dont ses caractéristiques physiques sont peu déterministes.
- **Formes géométriques de base** : Un rectangle ou une ellipse, peuvent être affectés à la forme de l'objet (Figure I.4 (c), (d)). Cette

représentation est souvent utilisée lorsqu'il s'agit d'un mouvement de translation et lorsque l'on a affaire à un objet rigide.

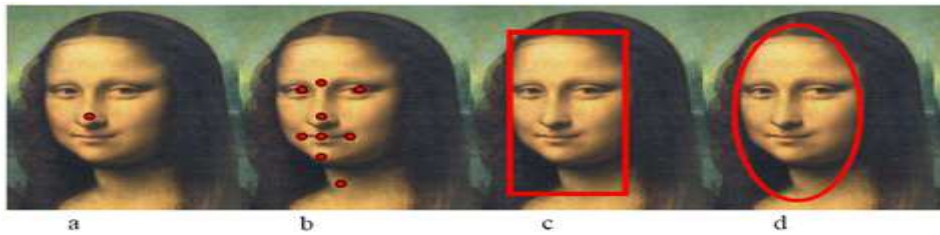


Figure I.4 Représentation d'objet par point et forme géométrique de base

- **Modèle articulé** : Un objet comportant plusieurs membres interconnectés, est un objet articulé. On représente par une forme géométrique de base (ellipse ou cylindre) chacun de ces membres. Ainsi, on obtient un modèle de représentation articulé. Le meilleur exemple pour le modèle articulé est le corps humain (Figure I.5 (e)).
- **Ossature** : Obtenu par représentation de l'axe médian de la silhouette de l'objet à suivre (Figure I.5(c), (d)). Ce modèle peut être appliqué soit pour un objet rigide ou non rigide.
- **Silhouette et contour** : Mise en évidence des contours de l'objet (Figure I.5 (g) et I.5.(h)) qui permettra de retrouver la silhouette de l'objet (figure I.1 (i)). C'est une représentation très appréciée dans le cas du tracking d'objets complexes et non rigides.

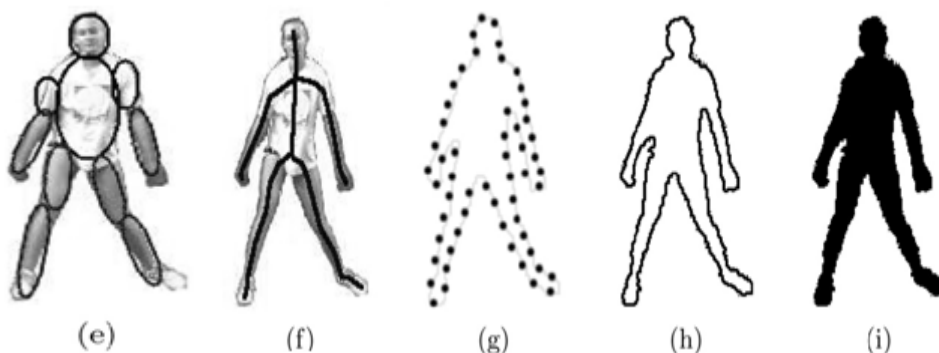


Figure I.5 Représentation d'objet par ossature (f), contour (g) et (h), et silhouette (i).

I.3.1.2 Représentation de l'apparence

L'apparence d'un objet en mouvement joue un rôle crucial, lorsqu'il s'agit de le distinguer des autres objets qui sont comme lui en mouvement d'un côté et de son arrière plan de l'autre. Par exemple, la couleur de l'objet est un détail très exploitable lorsqu'on a à utiliser l'histogramme de l'image. Par contre, la lisière de l'objet est plus intéressante lorsqu'on se base sur la méthode des contours. Il est à noter que l'apparence et la forme peuvent être combinées pour représenter l'objet, dans le contexte du suivi d'objets.

Les méthodes fondées sur des modèles d'apparence cherchent à déterminer les localisations successives des objets suivis en **maximisant** diverses mesures de **similarité** fondées sur des critères de couleur, de forme et/ou de texture.

I.3.1.3 Paramètres de sélection d'un objet

La sélection du bon paramètre est cruciale dans le suivi d'objets. En général, on désire avoir les propriétés visuelles d'un objet uniques, afin de le distinguer facilement de l'ensemble des autres objets. Les paramètres de sélection sont fortement rattachés aux représentations des objets. Par exemple : la couleur est un paramètre de sélection utilisé pour la représentation des apparences basées sur les histogrammes, les bordures de l'objet sont des paramètres de sélection pour les représentations d'apparences basées sur les contours, mais aussi la combinaison de plus d'un paramètre de sélection. Quelques paramètres seront présentés dans ce qui suit.

- **Couleur** : L'apparence couleur d'un objet est influencée par deux paramètres physiques :
 - La distribution spectrale de la luminance.
 - Les propriétés de réflexion de la surface de l'objet.

Dans le traitement d'images, le RGB (Red, Green, Blue) est souvent utilisé comme espace de représentation de la couleur, mais il a le désavantage de ne pas être une distribution uniforme de la couleur. Les espaces $L^*u^*v^*$ et $L^*a^*b^*$ et HSV (Hue, Saturation, Value) sont des espaces de distribution uniforme de la couleur, mais ils ont l'inconvénient d'être sensibles aux bruits. Un exemple de représentation d'une image au niveau de gris et couleur est représenté dans la figure I.6.

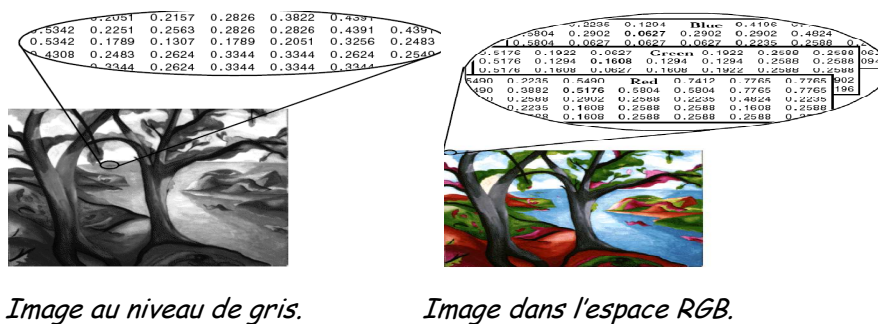
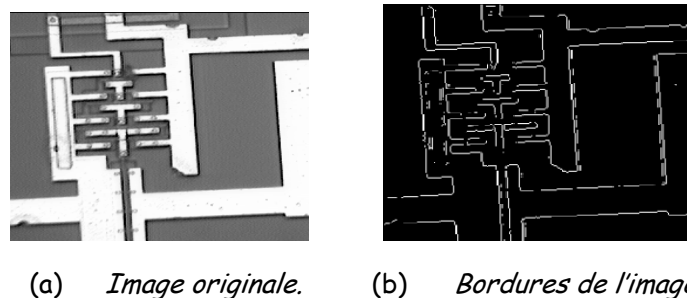


Figure I.6 : Représentation des couleurs.

- **Bordures** : Souvent, les grands changements d'intensités d'une image se trouvent sur les bordures (figure I.7) des objets, qui ont l'avantage d'être moins sensibles aux changements de luminance. Les algorithmes qui se basent sur la détection et le suivi des contours, utilisent souvent les bordures des objets comme paramètre de sélection.

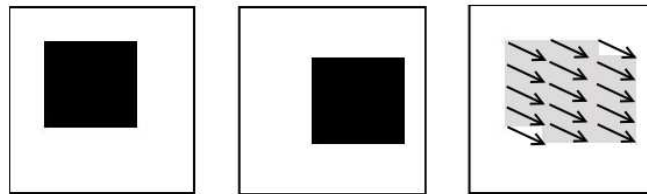


(a) *Image originale.*

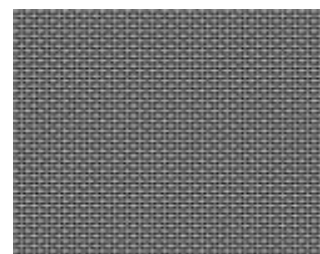
(b) *Bordures de l'image.*

Figure I.7 : Représentation des couleurs.

- **Flot optique** : C'est un champ de vecteurs qui représentent le déplacement de chaque pixel d'un objet dans une région (figure I.8). Il suppose l'intensité des pixels invariante. Le flot optique est, largement, utilisé dans la segmentation pour les applications du suivi d'objets. Pour représenter l'apparence, il donne des informations spatiales de l'objet d'intérêt.

**Figure I.8** : Exemple de flot optique.

- **Texture** : Une texture est utilisée pour la représentation d'une surface d'un objet caractérisé par une certaine homogénéité. La texture est une mesure qualitative qui peut décrire l'information caractérisant l'objet. Cette information peut avoir les aspects suivants : grossière, lisse, tachetée, granuleuse, fine, etc. La mise en œuvre de la texture d'une région nécessite une étape supplémentaire pour le calcul des descripteurs (texels), qui sont les plus petits éléments représentatifs d'une texture. Les images de la figure (I.9) représentent deux textures différentes.

(a) *Texture quelconque.*(b) *Texture uniforme.***Figure I.9** : Exemples de textures.

Notons que l'étape de la détection de mouvement, nécessaire pour les algorithmes de suivi d'objet, et l'établissement de la correspondance entre les positions de l'objet en mouvement à travers les images de la séquence (Tracking), peuvent se faire soit séparément, où séquentiellement. Dans les deux cas, l'objet à suivre est représenté par sa forme ou/et son apparence. Souvent, les algorithmes de suivi d'objets sont contraints par le type de la représentation choisie pour l'objet. Par exemple, si l'objet est représenté par un point, la translation et le seul mouvement à suivre par l'algorithme.

I.4 Suivi selon la représentation

La figure ci-dessous (Figure I.10) est une classification des différentes méthodes de suivi selon la représentation.

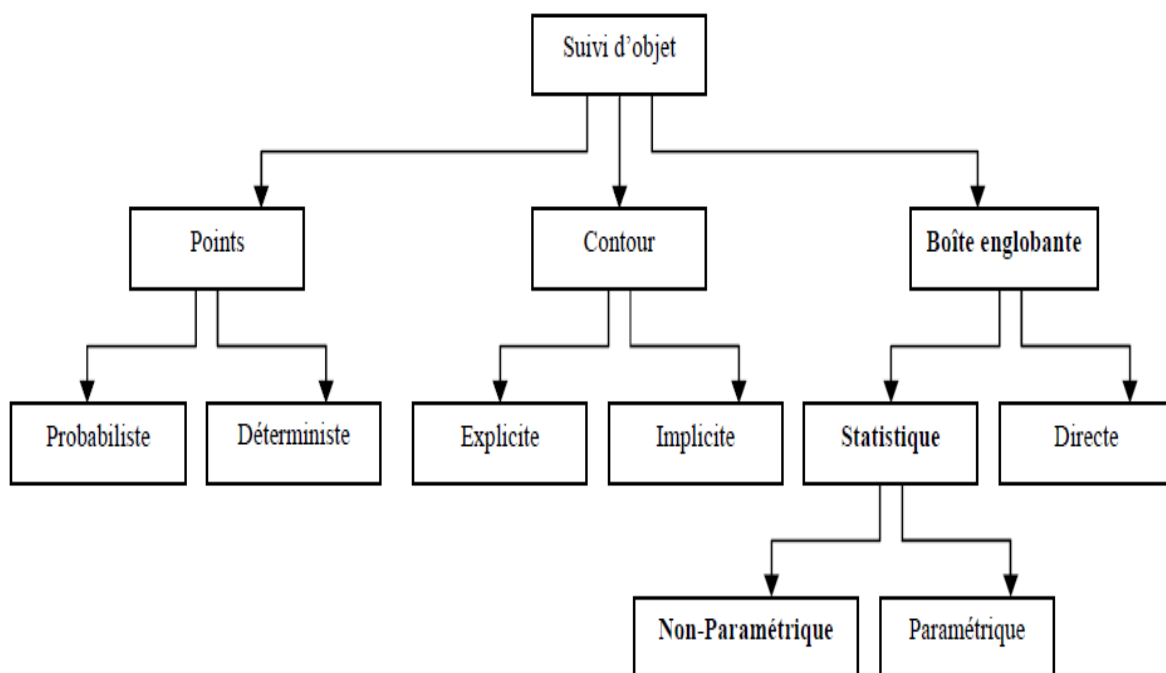


Figure I.10 Taxonomie des méthodes de suivi.

I.4.1 Suivi par point

Les objets détectés dans des images consécutives sont représentés par des points (figure I.11). La correspondance entre les points est faite par la connaissance de leurs états dans les images précédentes de la scène (que ce soit sa position ou son mouvement). Ces approches nécessitent un mécanisme de détection dans chaque image de la séquence.

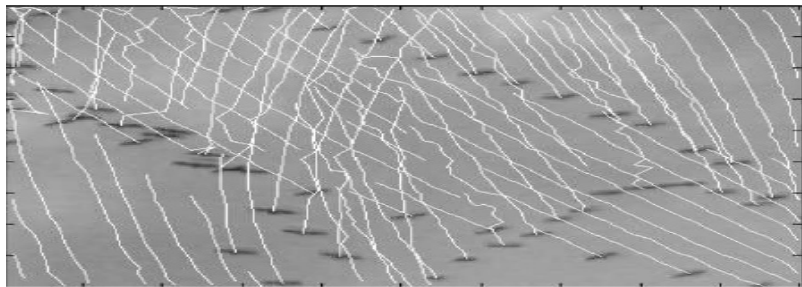


Figure I.11 : *Suivi des oiseaux par des points.*

Les méthodes basées sur le suivi de point peuvent être classées en deux catégories qui sont :

- **Méthodes statistiques**: Les méthodes statistiques de suivi de points sont, relativement, robustes par rapport aux bruits et aux perturbations, car dans le processus d'estimation de mouvement, elles tiennent en compte des erreurs de mesures et des incertitudes. Pour modéliser le mouvement, elles utilisent une information spatiale qui peut être : la position, la vitesse ou l'accélération de l'objet à suivre.
- **Méthodes déterministes** : A la place des fonctions de la densité de probabilité utilisées pour l'estimation du mouvement dans les méthodes statistiques, ces méthodes, dites déterministes, utilisent des mesures quantitatives sur des contraintes bien déterminées, comme la régularité et la rigidité. L'estimation du mouvement se fait par l'optimisation de ces mesures, où à chaque objet mesuré, dans une image à l'instant t , correspond un seul objet de la même mesure à l'instant $(t+1)$. L'avantage

de ces méthodes est qu'on peut rajouter des contraintes additionnelles pour gérer le mouvement, ce qui pourrait aussi augmenter le temps de calculs.

Les domaines d'application des méthodes de suivi d'objets par un point, sont très limités vu le peu d'informations qu'on peut extraire de l'objet et son mouvement.

Les difficultés persistantes pour le suivi par point sont en principe :

- l'occlusion : le chevauchement des objets ayant les mêmes caractéristiques.
- la mauvaise détection : détection d'un bruit en tant que point d'intérêt, ou bien, considération des points d'intérêt en tant que bruit.
- changement de topologie : le changement de topologie n'est pas géré automatiquement.

I.4.2 Suivi par forme géométrique de base (boite englobant)

L'idée est de suivre le mouvement d'un objet représenté par une forme géométrique de base (ellipse, rectangle, cercle, etc.) d'une image à une autre. Selon la représentation de l'apparence utilisée, nous pouvons distinguer, les modèles se basant sur la densité de probabilité d'un critère de sélection à l'intérieur d'une forme de base donnée [6]. Ces méthodes sont, généralement, utilisées pour le suivi des objets rigides.

I.4.3 suivi par contours actifs

Le principe du contour actif est de le faire évoluer vers les bords de l'objet à suivre. Cette évolution est gouvernée par la minimisation d'une fonctionnelle d'énergie, qui est formée de descripteurs (basés contour, ou basés région)

caractérisant l'objet intéressant. D'une manière générale, cette énergie prend la forme suivante :

$$E(\Gamma) = \int_0^1 E_{int}(\mathbf{v}) + E_{im}(\mathbf{v}) + E_{contexte}(\mathbf{v}) ds \quad (\text{I.8})$$

Avec :

- \mathbf{v} : la position d'un point.
- E_{int} : c'est l'énergie interne. Elle comprend souvent un terme qui gère la courbure du contour, et qui gère sa régularité.
- $E_{contexte}$: comprend des contraintes additionnelles.
- E_{im} : peut être vue comme l'énergie potentielle du contour.

Principalement, deux approches s'opposent pour la modélisation des contours actifs. La première, qualifiée de formulation Lagrangienne représente les contours actifs de manière **explicite** (ces contours sont souvent paramétriques). La seconde, qualifiée de formulation Eulérienne, représente les contours actifs de manière **implicite** (voir l'annexe).

I.5 Conclusion

La conclusion importante qu'on peut tirer de cette étude est que ça serait une gageure de prétendre faire un état de l'art exhaustif sur l'analyse de mouvement par ordinateur compte tenu des multitudes des travaux publiés.

En effet, selon l'application et le type d'objet qui nous intéresse, et en tenant compte des contraintes qui peuvent surgir, on peut choisir un modèle de représentation bien spécifié. Par exemple, si on est appelé à suivre un objet de taille minuscule, il suffirait de représenter sa forme par un point. Par contre, si l'objet était de taille plus importante, rigide ou pas, sa représentation de forme

pourrait être une forme géométrique de base, ou une représentation de silhouette, combinée avec une représentation d'apparence qui peut être aussi l'intensité, ou la couleur.

Aussi, l'algorithme utilisé pour le suivi d'objet est fortement lié à la représentation choisie. En effet, si le mouvement était une translation, la représentation par forme géométrique de base est bien adaptée. D'autre part, si on a affaire à un mouvement complexe, la représentation et le suivi par contours actifs est recommandé.

Le chapitre suivant sera consacré à la présentation d'autres modèles d'apparence pour le tracking et leurs performances.

II.1 Introduction

Nous rappelons que les méthodes basées sur un modèle d'apparence cherchent à localiser l'objet d'intérêt d'une manière successive et ce en maximisant diverses mesures de similarité fondées sur les attributs de couleur, de forme et/ou de texture. Par conséquent, la qualité d'un système de suivi est fortement conditionnée par le choix d'un modèle d'apparence adapté pour extraire de l'information qui permet de définir un critère objectif efficace pour déterminer la position réelle de l'objet suivi dans l'image.

En effet, un bon modèle d'apparence pour le suivi doit répondre à certaines exigences relatives à la précision de localisation, la complexité calculatoire, l'efficacité en temps réel, etc.

II.2 Modèles d'apparences pour le tracking

En plus des approches citées dans le chapitre précédent, d'autres techniques se basant sur un modèle d'apparence sont conçus pour le tracking .

II.2.1 Histogramme multi-échelles

Le principe de cette méthode est d'utiliser un descripteur qui introduit de l'information spatiale dans la représentation d'un objet pour le suivi. Cette nouvelle représentation permet, grâce à ce descripteur, d'effectuer la discrimination entre des objets différents qui ont pourtant le même histogramme brut. Ce qui n'est possible de faire en utilisant la représentation habituelle par histogramme brut (HB) dans l'algorithme Mean-shift [9].

L'histogramme multi-échelle est une famille d'histogrammes p élaborés à différentes échelles t d'une image. Pour obtenir un histogramme multi-échelles de l'image I , on effectue la convolution de cette image avec des filtres gaussiens $G(\tau)$ [10,11].

$$G(\tau) = \frac{1}{2\pi\tau^2} \exp\left(-\frac{X^2+Y^2}{2\tau^2}\right) \quad (\text{II.1})$$

- Où τ est l'écart type2 du filtre. L'histogramme de l'image convoluée est noté $P^\tau(\mathbf{I})=p(\mathbf{I}*\mathcal{G}_\tau)$ où $*$ représente l'opérateur de convolution.
- L'histogramme multi-échelles est le vecteur défini par :

$$P= [P^0, \dots, P^\tau, \dots, P^{N-1}] \quad (\text{II.2})$$

- Où P^τ , est un vecteur correspondant à l'histogramme à l'échelle τ et N est le nombre d'échelles considérées. Cette représentation inclut implicitement l'information spatiale de l'image (Figure II.1) :

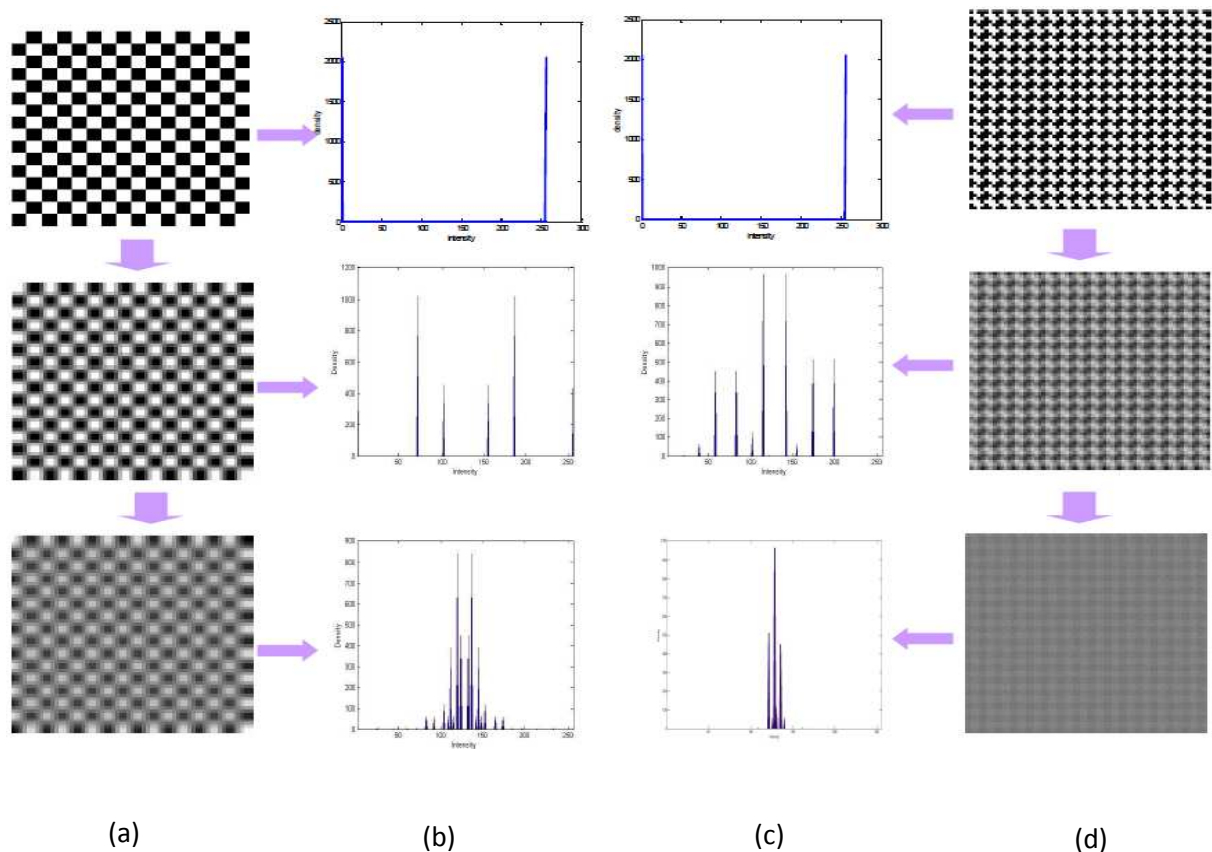


Figure II.1 : Exemples de deux histogrammes multi-échelles. Les colonnes (a) et (d) illustrent la décomposition multi-échelles de deux images. La ligne supérieure illustre les images originales. Les colonnes (b) et (c) illustrent leurs histogrammes multi-échelles respectifs. Les histogrammes originaux sont identiques mais les histogrammes multi-échelles sont distincts.

II.2.1.1 Histogramme multi-échelles cumulé

L'histogramme multi-échelles cumulé (HMC) est un dérivé de l'histogramme multi-échelles, c'est la moyenne d'une famille d'histogrammes à plusieurs échelles afin de créer un nouveau descripteur. La moyenne est définie selon la relation suivante :

$$P^{\text{cum}} = \frac{1}{N} \sum_{\tau=0}^{N-1} P^{\tau} \quad (\text{II.3})$$

- Où P^{τ} est un vecteur correspondant à l'histogramme à l'échelle τ et N est le nombre d'échelles considérées.

Le suivi par Mean-Shift via l'histogramme multi-échelles s'avère plus prometteur en termes de robustesse et de précision de localisation [9].

II.2.2 Modèles basés sur l'apprentissage statistique

L'apprentissage *statistique*, aussi appelé (*machine learning*) a pour but la résolution automatique de problèmes complexes à partir d'exemples. Les méthodes d'apprentissage statistique analysent des *entrées* (images, sons, mesures, symptômes, *etc.*) pour en déduire des sorties (catégorie d'objet, phrase, probabilité de défaillance, diagnostic, *etc.*).

Nous parlons d'apprentissage supervisé lorsque les sorties sont observées, en revanche, nous sommes dans un apprentissage non-supervisé si les sorties ne sont observées et *semi-supervisé* si elles ne sont que partiellement observées. Ces méthodes peuvent être classées en deux catégories :

- **Approches discriminatives** : elles consistent, exclusivement, à modéliser les relations entrées-sorties du système, avec un minimum d'hypothèses sur la structure des données d'entrée. Ces méthodes ne sont possibles que dans le cadre supervisé, ou semi-supervisé.

- **Approches génératives** : elles modélisent d'abord la structure du système avant de donner une réponse, déduite de ce modèle, relative à l'objectif de l'utilisateur. Ces approches modélisent en plus les liens qui existent au sein des variables d'entrée, contrairement aux approches discriminatives qui ne s'intéressent qu'aux relations d'entrées-sorties, pour avoir ainsi une modélisation plus importante.

La vision par ordinateur est l'un des domaines d'application l'apprentissage statistique connue d'ailleurs sous le nom de reconnaissance de formes. Ces techniques interviennent en télédétection, en imagerie médicale, en suivi et analyse du mouvement etc.

Comme nous l'avons cité précédemment, nous distinguons deux types d'apprentissages, supervisé et non-supervisé. Dans ce qui suit, nous n'allons nous intéresser qu'à des modèles qui font intervenir l'apprentissage supervisé. La figure (II.2) illustre le principe de l'apprentissage supervisé.

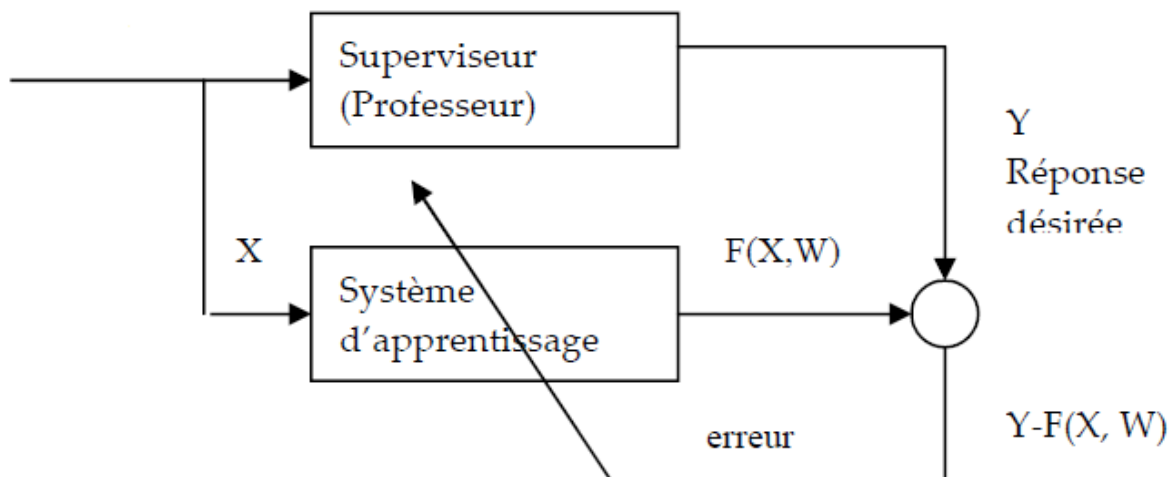


Figure II.2 : Approche supervisée.

X : représente les exemples d'entrée ; Y : représente la sortie désirée tracée par le superviseur ; $F(X, W)$: représente la réponse du système en fonction de X et les paramètres W ; erreur : représente l'erreur d'apprentissage est égale à $Y - F(X, W)$

II.2.3 Classification en apprentissage supervisé

Nous rappelons que les méthodes supervisées ont un rôle prédictif du fait qu'elles permettent d'évaluer une quantité donnée en sortie (la taille d'un individu) sans la calculer directement mais plutôt en se basant sur des valeurs qui lui sont liées en entrée (le poids de la personne).

En effet, l'apprentissage supervisé fait intervenir deux types de variables :

- les variables d'entrée, notées X .
- les variables de sortie, notées Y , qui doivent être prédites à partir de la valeur de X associé.

Nous cherchons donc à modéliser le lien existant entre X et Y , ce qui donne lieu à la classification supervisée. Pour résoudre un problème d'apprentissage supervisé, deux types d'approche sont possibles :

- l'approche *discriminative* : modélise directement la règle de classification $P(Y|X)$,
- l'approche *générative* : cherche à modéliser la distribution jointe $P(X, Y)$ des entrées et des sorties. On en déduit ensuite la règle de classification par application de la loi de Bayes.

La figure (II.3) regroupe les principales méthodes de classification génératives et discriminatives [8,12].

Nous n'allons pas détailler toutes les méthodes de cette taxonomie mais plutôt expliquer, d'une manière générale, les deux grandes familles de modèles, génératives et discriminatives et les modèles intermédiaires et nous nous focaliseront sur leur application ainsi que sur leur apport en tracking. Nous nous attarderont plus sur les méthodes qui feront l'objet du chapitre suivant.

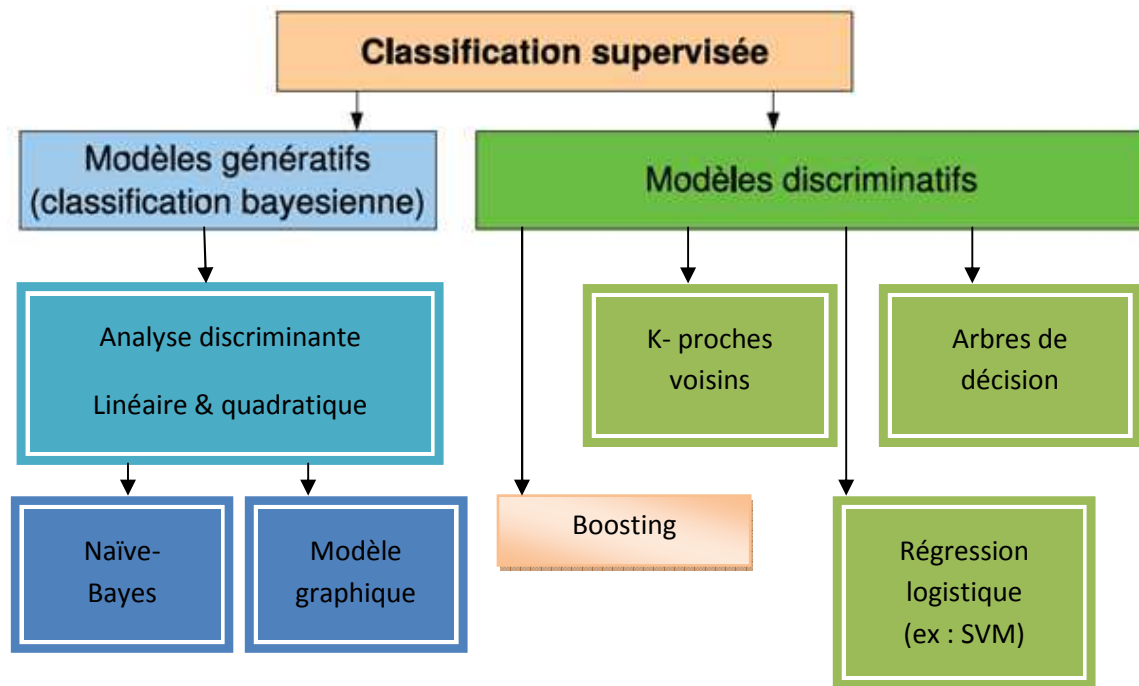


Figure II.3 : principales méthodes de classification, génératives et discriminatives. Les méthodes indiquées par la couleur jaune sont des méthodes discriminatives qui peuvent être liées à un modèle génératif.

II.2.3.1 Modèles discriminatifs en classification supervisée

Les méthodes discriminatives proposent un modèle pour $P(Y/X)$ dont les paramètres sont estimée en minimisant le cout de la classification.

Application au tracking

En tracking, les algorithmes de cette catégorie posent le problème comme étant une classification binaire afin de trouver la limite de la décision pour la séparation de l'objet cible du fond. Dans cette optique, une multitude de modèles de classification existent, en particulier *la méthode des k plus proches voisins*, *les arbres de décision* [13] qui ont démontré que pour prédire la classe d'un nouvel échantillon dans une image, l'algorithme cherche les k plus proches voisins de ce même échantillon et prédit la réponse la plus fréquente de ces k plus

proches voisins. Ces derniers peuvent être stabilisés par *boosting* pour offrir une règle de décision particulièrement simple et efficace.

- **Machines à Support Vecteurs (SVM)** : une méthode non paramétrique récente [14] qui consiste à résoudre un problème de classification binaire en plaçant un hyperplan dans l'espace des données comme frontière de décision de manière à ce que :
 - cet hyperplan maximise le taux de bonne classification des échantillons d'apprentissage.
 - La distance entre le plan et le pixel le plus proche est maximisée.

Le principe de cette méthode est décrit par la figure (II.4) :

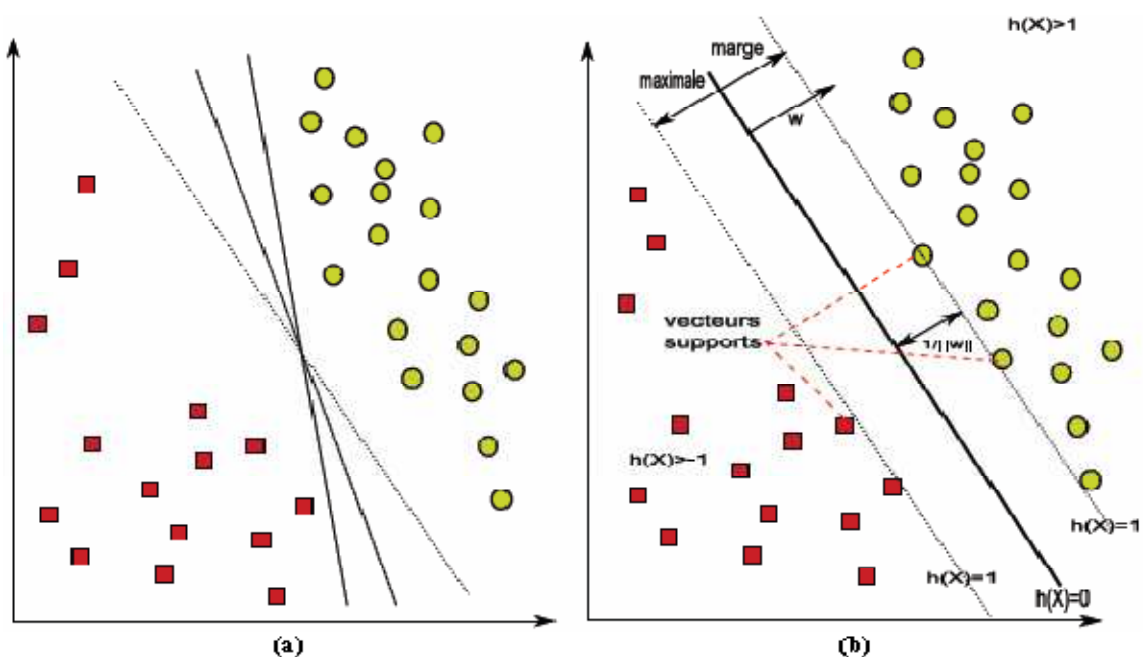


Figure II.4 : Principe des SVM.

Où (a) et (b) représentent :

- Recherche de l'hyperplan optimal ;
- (b) Hyperplan optimal, marge et vecteurs supports

Ce plan est alors défini par une combinaison des échantillons les plus proches de ce plan, qui sont appelés les vecteurs supports [15].

Avidan [16] s'étend à l'approche de flux optique avec un classificateur de support vecteur machine pour le suivi d'objet.

- Le **Boosting** est une classe de méthodes d'apprentissage basée sur la sélection de classificateurs. Ce sont des techniques utilisées dans de nombreuses activités d'apprentissage, aussi bien pour de la catégorisation [17] et le filtrage de texte, que pour des problèmes linguistiques [18], de bioinformatique [19] ou de détection [20] et classement d'objet dans des images. Le principe du Boosting est illustré dans la figure (II.5).

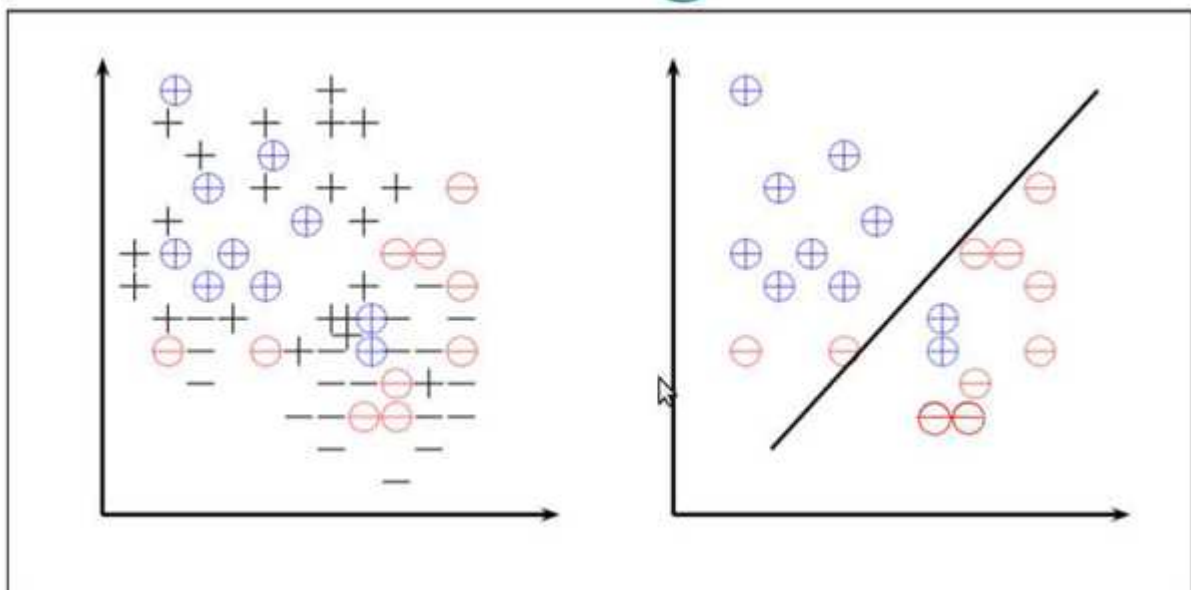


Figure II.5 : principe du Boosting. A gauche : l'ensemble d'apprentissage S (+et -) et le sous-ensemble S_1 (points rouges et bleu entourés). A droite : l'ensemble S_1 et la droite C_1 apprise sur cet ensemble.

Garbner[21] , à proposé une approche de Boosting en ligne. A partir d'un ensemble d'échantillons, on introduit un sélecteur qui choisit parmi plusieurs

classificateurs (ou features), celui qui présente une erreur minimale d'estimation.

II.2.3.2 Modèles génératifs en classification supervisée

La modélisation générative cherche en premier lieu à trouver une structure pour la distribution jointe des entrées X et sorties Y . Elle correspond à une vaste catégorie de classificateurs, dont les classificateurs *de Bayes Naïfs* (NB). On parle de modélisation générative car une loi de probabilité jointe est définie pour toutes les variables possibles, c'est-à-dire pour les données à prédire Y , les données d'entrée X

- **Modèles graphiques** : Les modèles graphiques encore appelés *réseaux bayésiens*, utilisent des modèles modulaires pouvant être représentés sous la forme de graphe. Entre les variables, un graphe de dépendance est alors établi. Ceci s'avère très intéressant lorsque le nombre de variables soumises à l'apprentissage est important [12].
- **statistiques Bayésienne** : consiste à considérer que les paramètres d'un modèle sont aléatoires. Il est donc nécessaire de définir une distribution a priori de ces paramètres ne dépendant pas des observations.

C'est un outil d'apprentissage statistique qui permet d'incorporer, d'une manière probabiliste, des informations, extérieures aux données, cruciales à la reconnaissance. Lorsque la quantité de données disponibles ne suffit pas pour l'estimation fiable du modèle, l'ajout de ces informations permet d'améliorer considérablement les performances d'une méthode d'apprentissage statistique. On parle généralement dans ce cas de *régularisation bayésienne*.

- **Approche Bayésienne naïve** : La classification bayésienne naïve est une approche supervisée, basée sur le théorème fondamental naïf de Bayes décrit par l'équation (II.4) Les méthodes de cette approche sont

qualifiées de "Naïve" ou "Simple" car elles supposent l'indépendance des variables [22]. En effet, l'hypothèse naïve a priori du classificateur est que toutes les propriétés ou événements sont conditionnellement indépendants.

$$P(C_i/V_j) = \frac{P(C_i)P(V_i/C_i)}{P(V_i)} \quad (\text{II.4})$$

Si nous prenons l'exemple de classification de texte nous considérons alors deux ensembles de texte et de classe de texte D et C respectivement tel que :

$D = \{d_i, i=1..|D|\}$, $C = \{C_k, k=1..|C|\}$. Soit un texte d_i dont le vecteur de caractéristiques correspondant est : $V_j = \langle V_{1j}, V_{2j}, \dots, V_{nj} \rangle$. Le principe de la classification bayésienne consiste à affecter au texte d_i sa classe correspondante $C_i \in C$ telle que la probabilité conditionnelle par rapport à V_j est la plus élevée [23].

Les algorithmes génératifs pour le tracking font, généralement, l'apprentissage d'un modèle pour représenter l'objet cible, puis, l'utilisent pour retrouver la région dans l'image avec une erreur minimal de reconstruction. La méthode IVT [24], utilise un modèle avec un apprentissage incrémental pour adapter les changements d'apparences. Récemment, la classification, en utilisant une matrice creuse est très répondeue [25]. D'après cette étude, les algorithmes de cette catégorie, qui font un apprentissage d'un modèle d'apparence en ligne pour le tracking, se limitent généralement pour des applications en temps réel.

II.3 Conclusion

Dans ce chapitre, nous avons présenté de nouveaux modèles d'apparence adaptés au tracking. D'un coté, les histogrammes multi-échelles qui introduisent un descripteur dans la représentation multi-échelles d'un objet pour compenser les limites des histogrammes mono-échelle en utilisant de l'information obtenue

pour plusieurs niveaux de lissage de l'image d'origine. D'un autre côté, des modèles basés sur l'apprentissage supervisé. En effet, en vision par ordinateur les caméras numériques se généralisent et leurs performances ne cessent de croître. Dès lors, la compréhension automatique des images environnantes est une étape indispensable pour le traitement d'un grand nombre d'images, pour faire face à la grande complexité du problème à traiter, la plupart des méthodes nécessitent des images de très bonne qualité.

III.1 Introduction

La vision par ordinateur et ces différentes applications, telles que la télédétection, l'imagerie médicale et les systèmes à réseaux de caméras pour surveillances, nécessitent des techniques de traitement efficaces, rapides et précises pour aboutir à des applications en temps réel. Dans cette optique, plusieurs travaux, plus ou moins robustes, ont été réalisés et publiés [26,27] notamment dans le domaine de suivi d'objets (Tracking).

Cependant, malgré que de nombreux algorithmes aient été proposés dans la littérature, le suivi d'objet en mouvement reste tout de même un problème de taille à cause de plusieurs facteurs liés soit, à la complexité de la scène en termes de changement de luminance, occlusion et nature du mouvement, soit à la représentation de l'objet à suivre. Pour cela, la mise en évidence d'un modèle efficace pour le tracking à toujours été et reste une préoccupation majeure pour les chercheurs.

Dans ce chapitre, nous allons tester et conclure sur l'efficacité de deux types de modèle pour le tracking : *modèle de contour actif* et *modèle d'apparence*.

III.2 Modèle de contour actif pour le tracking

Dans la littérature, une liste non exhaustive d'algorithmes basés sur des modèles de contours actifs pourrait être faite. Cependant, les limites de ces algorithmes s'avèrent clairement dans leurs applications en temps réel. En effet, face à la complexité des processus relevant de ce domaine, les modèles à appliquer pour le tracking, pour qu'ils soient rapides, exigent une simplicité dans leur écriture mathématique. Ces modèles devraient présenter le moins possible de termes de dérivées partielles (EDP) qui sont gourmands en temps de calcul.

Le modèle que nous avons repris [28], [29] puis testé et évalué dans cette section est construit à partir de deux modèles, géodésique (CAG) qui est basé contour et celui de Chan-Vese (C-V) qui est basé région (voir l'Annexe). Le modèle est décrit avec la formule (III.1).

$$\frac{\partial \phi}{\partial t} = SPF(I(v)) \cdot \alpha |\nabla I|, \quad v \in \Omega. \quad (\text{III.1})$$

Où :

- ✓ α est une constante rajoutée pour augmenter la vitesse de l'évolution du contour.
- ✓ Ω est le domaine de l'image.
- ✓ $SPF(I)$ est une force signée qui régit le sens de l'évolution du contour qui est sous la forme suivante :

$$SPF(I(x)) = \frac{I(x) - \frac{C_1 + C_2}{2}}{\max \left(\left| I(x) - \frac{C_1 + C_2}{2} \right| \right)}, \quad x \in \Omega \quad (\text{III.2})$$

C'est une force générée par la différence de la moyenne des intensités à l'intérieur et à l'extérieur de l'objet (C_1 et C_2 respectivement) et leur moyennes. Cette force applique une pression incitant le contour à se rétrécir ou s'accroître selon son signe.

Pour évaluer l'efficacité de ce modèle en tracking, nous allons effectuer deux tests. Une application sur une vidéo enregistrée et une autre sur le tracking en temps réel.

III.2.1 Application du modèle sur une vidéo

Pour réaliser l'application du modèle décrit par l'équation (III.1) au suivi, un algorithme de segmentation de l'objet d'intérêt dans une séquence d'image est appliqué au préalable (voir l'annexe).

Le processus du tracking est illustré dans la figure (III.1).

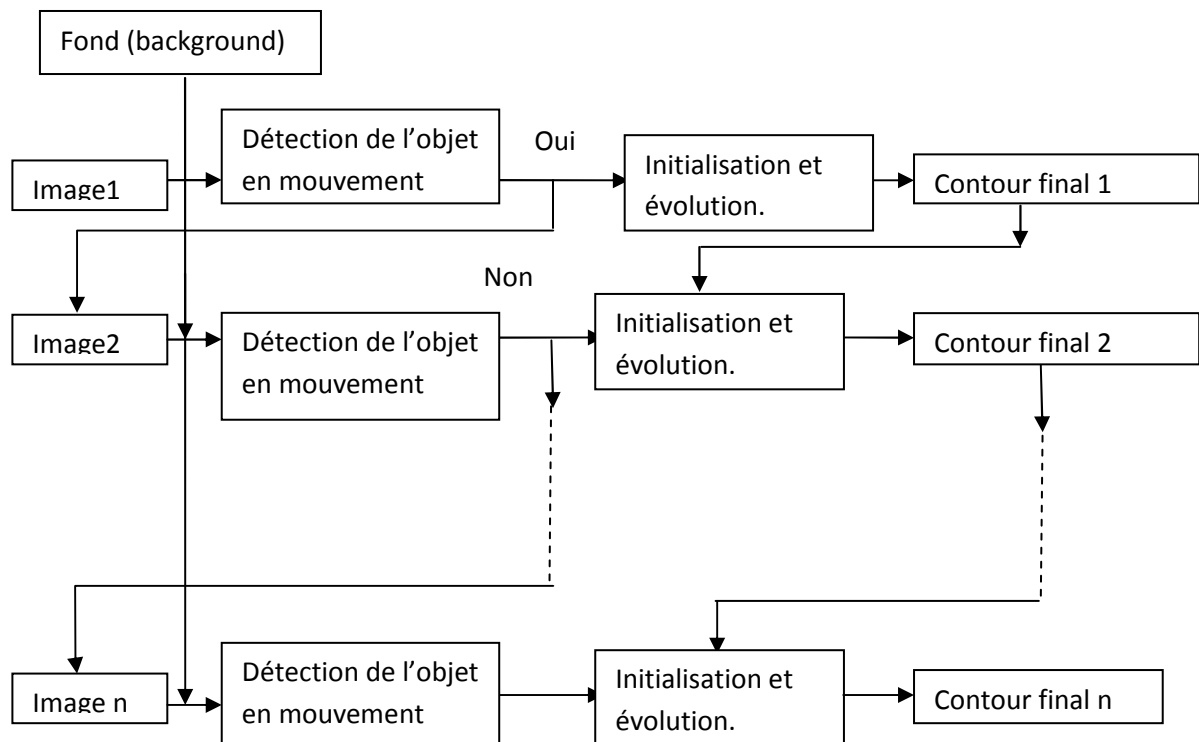


Figure III.1 : croquis schématisant le suivi.

Dans ce processus, les principales étapes du suivi sont :

- **Estimation du fond (background)** : dans notre cas, le fond est considéré comme l'image prise à l'aide d'une caméra fixe, en l'absence d'un objet quelconque.
- **Acquisition de la première image de la séquence** : l'image est prise à l'apparition d'un objet. Un test de détection d'objet est alors effectué par une soustraction directe entre le background et la première image.
- **Suivi de l'objet en mouvement** : la détection et le suivi du mouvement se fait par la soustraction directe entre les images successives jusqu'à la fin de la séquence.

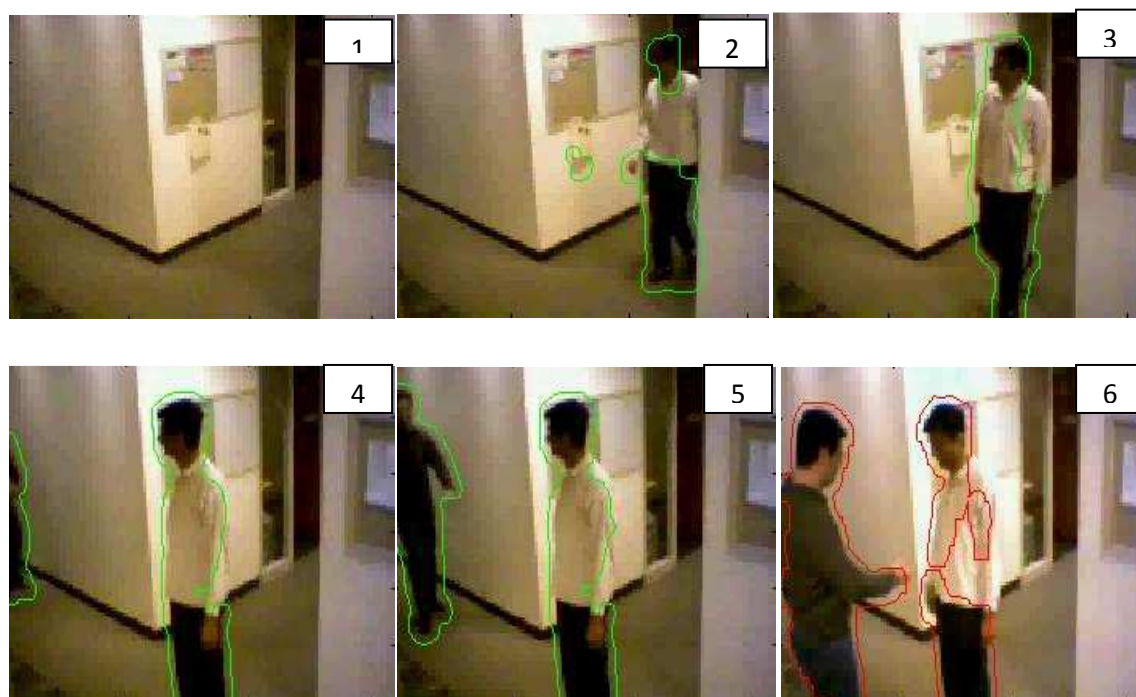
Pour la détection et le suivi de l'objet en mouvement (différence, background-image), le contour initialisé évolue jusqu'à convergence (contour final). Le contour final de la convergence sera le contour initial pour la prochaine image, l'opération se répète jusqu'à la fin de la séquence.

a/ Evaluation du modèle sur une vidéo enregistrée

Il est noter que dans les tests qui suivront, la valeur du paramètre « α » de l'équation (III.1) est fixée à $\alpha = 15$. En effet, cette valeur est décidée empiriquement, du fait qu'il en résulte un meilleur temps de convergence de l'algorithme de détection (voir l'annexe).

Nous rappelons que pour l'expérimentation du modèle adopté ici pour le tracking, nous faisons l'acquisition des images sous un background fixe.

La vidéo utilisée pour ce test est nommée « vipmen.avi » dans MatLab. Les résultats de la segmentation et du suivi, avec application d'une dilatation sur les images de la séquence, sont montrés dans la figure (III.2).



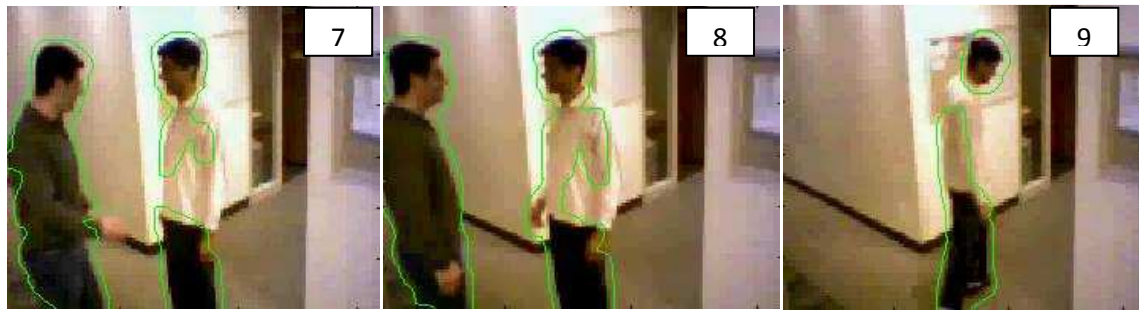


Figure III.2 : résultats de l'application sur une vidéo

Discussion

L'image (1) de la séquence est utilisée comme arrière-plan. Dans les images (2 et 3), l'apparition et le mouvement de la première personne à été détecté et suivi par le contour de couleur verte, qui signifie la convergence. L'arrivée de la deuxième personne dans les images (4 et 5) à bien été segmentée, le contour actif suit son mouvement aussi dans les images (7 et 8). Cependant, le contour, dans ces dernières, présente des discontinuités relatives, probablement, aux limites du modèle adopté, dues, comme par exemple :

- ✓ Au changement d'intensité dans le fond
- ✓ Au mouvement brusque illustré même par l'image (6) par le contour rouge.
- ✓ Au fond complexe, en effet, la chemise blanche n'a pas été bien détectée du fait qu'elle possède les mêmes intensités que le fond. Aussi les pantalons noirs avec la faïence noire.

b/ Evaluation du modèle sur le tracking en temps réel

Pour l'application en temps réel, nous avons utilisé une Web-cam avec une fiche USB, pour l'acquisition.

Nous présentons ici, les résultats des tests effectués sur le mouvement de main ou du corps humain. Les figures (III.3 et III.4) illustrent approximativement les performances obtenues.

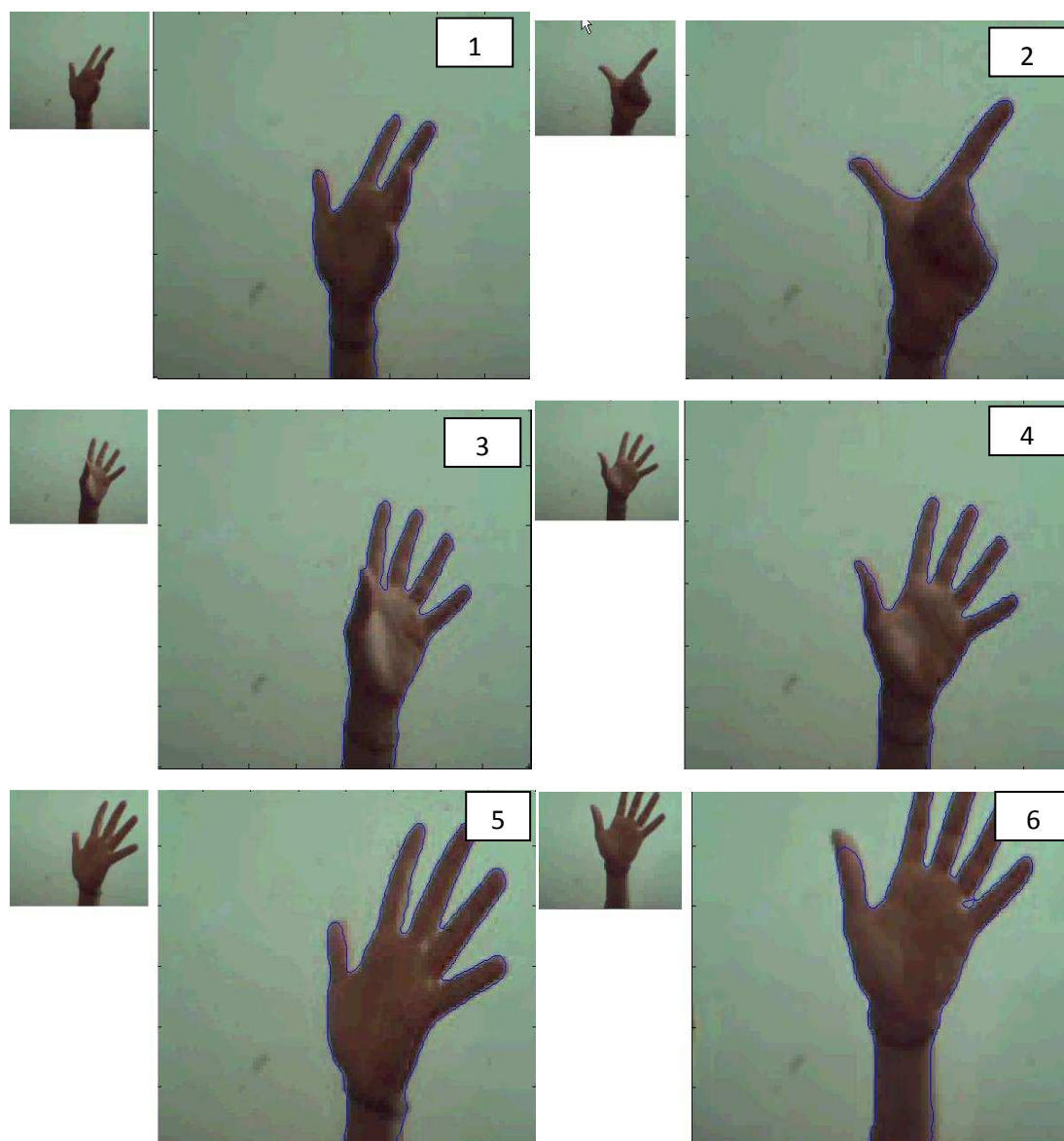


Figure III.3 : *tracking en temps réel du mouvement de doigts et de main*

Discussion

Dans cette séquence, le résultat de la segmentation de la main durant son mouvement en temps réel est assez bon du fait que dans la quasi-totalité des images (1, 2, 3, 4 et 5), le contour actif entoure parfaitement la main suivie. Excepter les quelques défauts de segmentation dus au mouvement brusque et au changement d'intensité du fond, images (2 et 6).

Le temps de convergence de l'algorithme pour le suivi, est dans la plage de [0.45 à 0.65] secondes, pour une valeur de 15 de α .



Figure III.4 : *tracking en temps réel du mouvement de corps humain.*

Discussion

Les résultats de cette application sont presque similaires à ceux présentés dans la figure (III. 3) en termes de segmentation et du tracking. Quelques défauts dus aux changements d'apparence sont toujours visibles, néanmoins, le suivi est plutôt performant.

c/Conclusion sur les performances du modèle

Les différents résultats obtenus par l'application du modèle décrit par l'équation (III.1), sur une vidéo enregistrée ou en temps réel, nous amènent à conclure sur son efficacité et sur ses limites.

En effet, le modèle adopté prend en compte l'information du background pour la détection et le suivi d'objets. Par conséquent, le changement d'intensité lumineuse affecte sensiblement les résultats. Néanmoins, le modèle de contour actif utilisé parvient à suivre le mouvement des objets d'intérêt avec succès surtout quand l'apparence de la scène ne varie que peu.

Malgré que le modèle est basé région et qu'il exploite les informations de tous les pixels pour faire évoluer le contour, un temps de calcul record à été enregistré comparativement à d'autre temps réalisés par d'autres modèles.

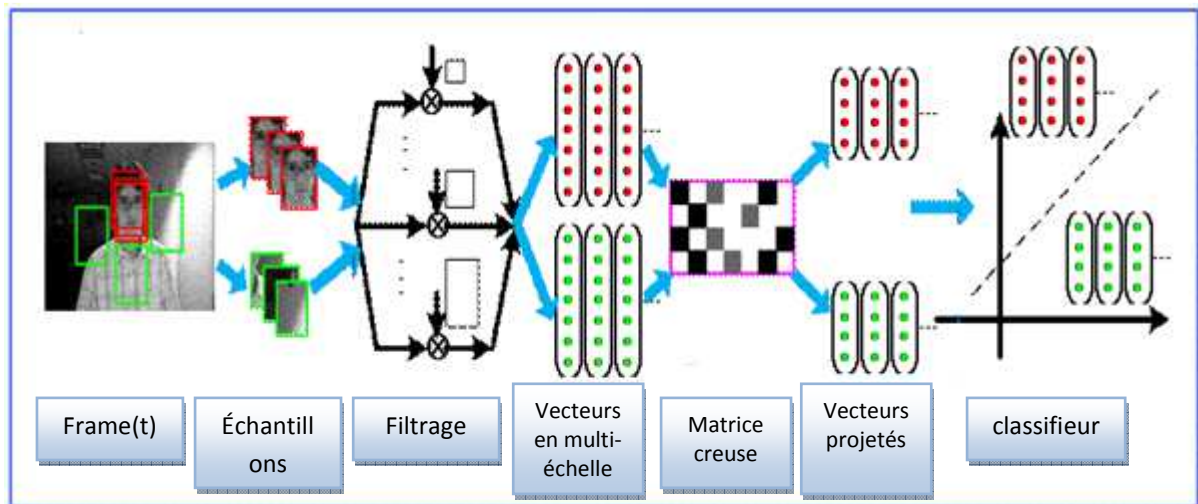
En fin, nous pouvons conclure que ce modèle convient plus pour la segmentation d'images statiques que pour le tracking. Dès lors, il serait peut être plus judicieux de tester un autre type de modèle basé sur l'apparence pour le tracking.

III.3 un modèle d'apparence pour le tracking

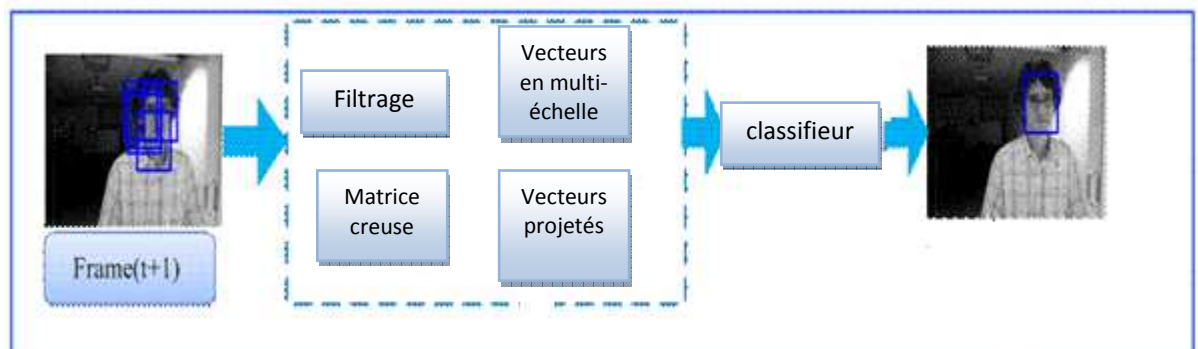
Comme nous l'avons déjà évoqué dans le chapitre précédent, développer un modèle d'apparence efficace pour le tracking s'avère une tâche ardue, compte tenu de plusieurs facteurs relatifs au changement d'intensité lumineuse, l'occlusion, de pose etc.

Le modèle que nous avons adopté ici pour le suivi est un modèle d'apparence qui est basé sur l'apprentissage supervisé on-ligne et qui est à la fois génératif et discriminatif. En effet, le modèle est génératif du fait qu'un objet peut être bien représenté par des caractéristiques (features) extraites dans le domaine de compression. Il est également discriminatif car il exploite ces features pour séparer la cible de son background environnant en faisant une classification binaire par l'intermédiaire d'un classifieur de naive-bayes [30].

Les principales étapes de l'algorithme de classification sont illustrées dans la figure (III.5) ((a) et (b)).



(a) : mise-à-jour du classifieur à la frame (t).



(b) : tracking à la frame (t+1).

Figure III.5 : principales étapes de l'algorithme de compression pour le tracking.

Dans ce modèle, la sélection des features se fait par la préservation de l'information et de la réduction de dimensionnalité à partir de la représentation multi-échelle de l'image, en se basant sur les théories de compression [12], [13]. Pour cela on utilise une matrice de mesure très creuse (a very sparse matrix) qui satisfait la propriété d'isométrie restreinte (RIP) [15] pour faire une projection (compression) de l'espace de l'image vers un espace compressé et de dimension réduite.

◆ Principe de la projection (compression)

Soit une matrice $R \in \mathbb{R}^{n \times m}$ dont les lignes ont des éléments de longueur unité pour passer de l'espace de grande dimension de l'image $X \in \mathbb{R}^m$ vers un espace de dimension inférieur $V \in \mathbb{R}^n$ avec :

$$V = RX \quad (\text{III. 3})$$

Avec : $n \ll m$.

La matrice R doit satisfaire le lemme Johnson- Lindenstrauss [16] pour pouvoir reconstruire X avec une grande probabilité, à partir de V , avec une erreur minimale. Cela peut garantir alors que V préserve presque toutes les informations contenues dans X .

◆ Mode opératoire de l'algorithme de ce modèle

Dans cette approche, le problème est décrit comme étant une tâche de détection (voire figure III.5). Il suppose que la fenêtre de suivi de la première image a été déterminée. Pour mettre à jour le classifieur, des **échantillons positifs** et sont prélevés **près de l'emplacement actuel de la cible** et **d'autres négatifs loin du centre de la cible**. Pour prévoir l'emplacement de l'objet dans l'image suivante, l'algorithme tire quelques échantillons autour de la cible en cours et désigne l'échantillon ayant une classification maximale.

La figure (III.6) illustre graphiquement la compression d'un vecteur X de grande dimension à un vecteur V , de dimension réduite.

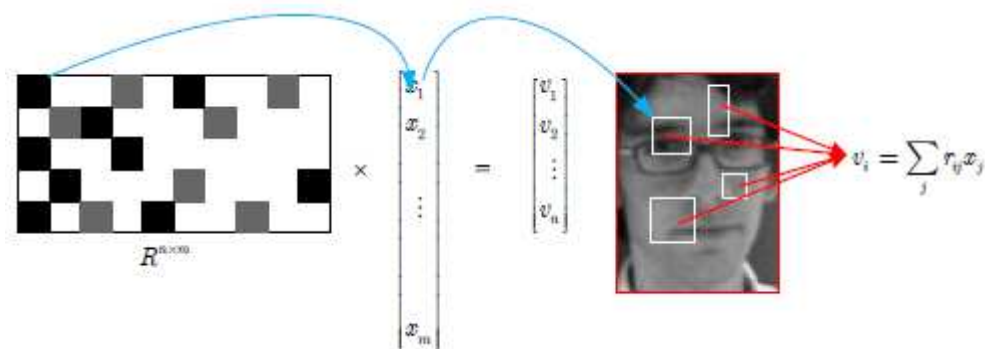


Figure III.6 : compression par une matrice creuse, R

Dans la matrice R, les rectangles noirs, gris et blanc représentent, respectivement, les entrées négative, positive et nulle. La flèche en bleu illustre que l'une des entrées non-nulles d'une ligne de R, multipliant un élément dans X, est équivalent à un filtre rectangulaire en convolution avec l'intensité à une position donnée dans l'image d'entrée.

Pour chaque échantillon $z \in \mathbb{R}^m$, sa représentation dans l'espace à dimension réduite, est donnée par ; $v = (v_1, \dots, v_n)^T \in \mathbb{R}^m$ avec $n \ll m$. Les éléments de v sont supposée indépendamment distribués et classifiés avec le classifieur de Nive Bayes [22].

Le modèle du classifieur utilisé ainsi que sa mise-à-jour sont donnés par:

$$H(V) = \log \left(\frac{\prod_{i=1}^n P(v_i|y=1)P(y=1)}{\prod_{i=1}^n P(v_i|y=0)P(y=0)} \right) = \sum_{i=1}^n \log \left(\frac{P(v_i|y=1)}{P(v_i|y=0)} \right) \quad (\text{III.4})$$

Où, $P(y=1)=P(y=0)$; avec y une variable binaire ($y \in [0, 1]$) qui représente le label d'un échantillon.

Les distributions conditionnelles $P(v_i|y = 1)$ et $P(v_i|y = 0)$ dans $H(V)$ sont supposées être des distributions Gaussiennes avec quatre (4) paramètres $(\mu_i^1, \sigma_i^1, \mu_i^0, \sigma_i^0)$ où :

$$P(v_i|y = 1) \sim N(\mu_i^1, \sigma_i^1) ; \quad P(v_i|y = 0) \sim N(\mu_i^0, \sigma_i^0) \quad (\text{III.5})$$

Les paramètres scalaires dans (III.4) sont mis à jour d'une manière incrémentale comme suit :

$$\mu_i^1 \leftarrow \lambda \mu_i^1 + (1 - \lambda) \mu^1$$

$$\sigma_i^1 \leftarrow \sqrt{\lambda \sigma_i^1{}^2 + (1 - \lambda)(\sigma^1)^2 + \lambda(1 - \lambda)(\mu_i^1 - \mu^1)^2} \quad (\text{III.6})$$

Où :

- λ est un paramètre d'apprentissage.
- $\sigma^1 = \sqrt{\frac{1}{n} \sum_{k=0}^{n-1} |y=1} (v_i(k) - \mu^1)^2$
- $\mu^1 = \frac{1}{n} \sum_{k=0}^{n-1} |y=1} v_i(k)$

L'algorithme de compression pour le tracking peut être résumé dans la figure (III.7). Cet algorithme a pour entrée la première image de la séquence sur laquelle un ensemble de patches D^γ est extrait, tel que $D^\gamma = \{Z | \|I(z) - I_{t-1}\| < \gamma\}$ où, I_{t-1} est la position de la cible dans la frame (t-1). Il ya ainsi l'extraction de features avec dimensionnalité réduite.

L'algorithme utilise la formule H de l'équation (III.4) pour chaque vecteur d'échantillons $v(z)$ afin de retrouver la localisation optimale I_t . Deux autres patches d'image sont échantillonnés, $D^\alpha = \{Z | \|I(z) - I_t\| < \alpha\}$ et $D^{\zeta, \beta} = \{Z | \zeta < \|I(z) - I_t\| < \beta\}$ avec $\alpha < \zeta < \beta$, pour γ extraire deux ensembles d'échantillons pour la classification et la mise-à-jour des paramètres du classifieur selon l'équation (III.6).

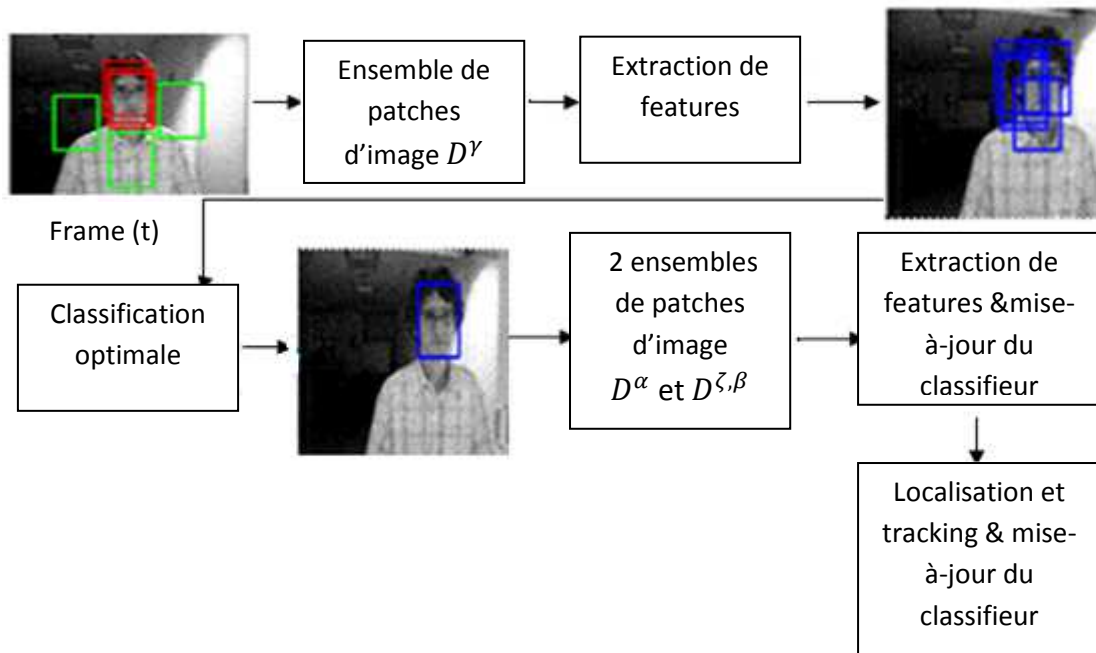


Figure III.7 : principales étapes de l'algorithme du tracking.

III.3 Expérimentation et évaluation du classifieur

Etant donnée la position de cible dans la frame courante, le paramètre est initialisé à $\alpha = 4$, ce qui génère 45 échantillons positifs. Pour générer les échantillons négatifs, les auteurs ont pris : $\zeta = 8$ et $\beta = 30$. 50 échantillons négatifs sont sélectionnés aléatoirement. Les autres paramètres ont pour valeurs les quantités suivantes : $n = 50$, $\gamma = 20$ et le paramètre d'apprentissage $\lambda = 0.85$.

III.3.1 Robustesse à la détection pour le tracking

Les méthodes de suivi par détection se heurtent souvent à des problèmes d'ambiguïté inhérents tels que, le changement d'apparence, comme illustré dans la figure (III.8).

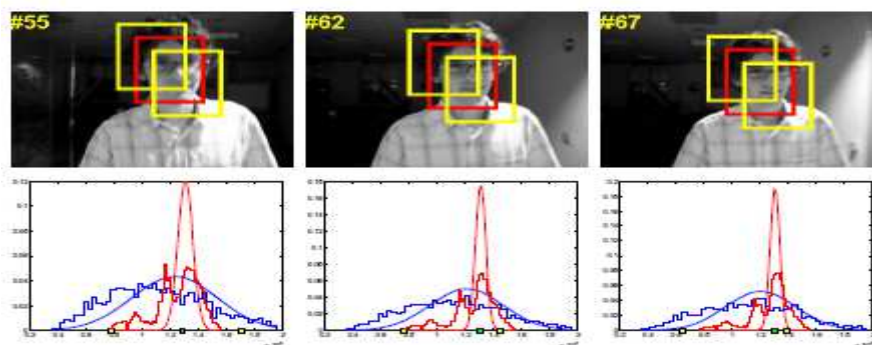


Figure III.8 : *robustesse du classifieur face au changement d'apparence.*

Alors que l'apparence de la cible change au fil du temps, la plus part des échantillons positifs les plus corrects, en rouge, sont similaires dans la plus part des frames. Tandis que, les échantillons les moins corrects, en jaune, sont différents du fait qu'ils impliquent des informations du fond. Ainsi, l'algorithme est stable puisqu'il est capable de sélectionner l'échantillon positif le plus correct car sa probabilité est plus grande que celle de l'échantillon positif le moins correct.

III.3.2 évaluation du classifieur sur différentes séquence vidéo

Pour évaluer et conclure sur les performances du classifieur en tracking, nous appliquons l'algorithme sur plusieurs séquences vidéo présentant des problèmes relatifs au changement de luminosité, de pose, de mouvement brusque, d'échelle etc.

La première séquence de test contient 462 frames. Quelques résultats de l'application sont présentés dans la figure (III.9).

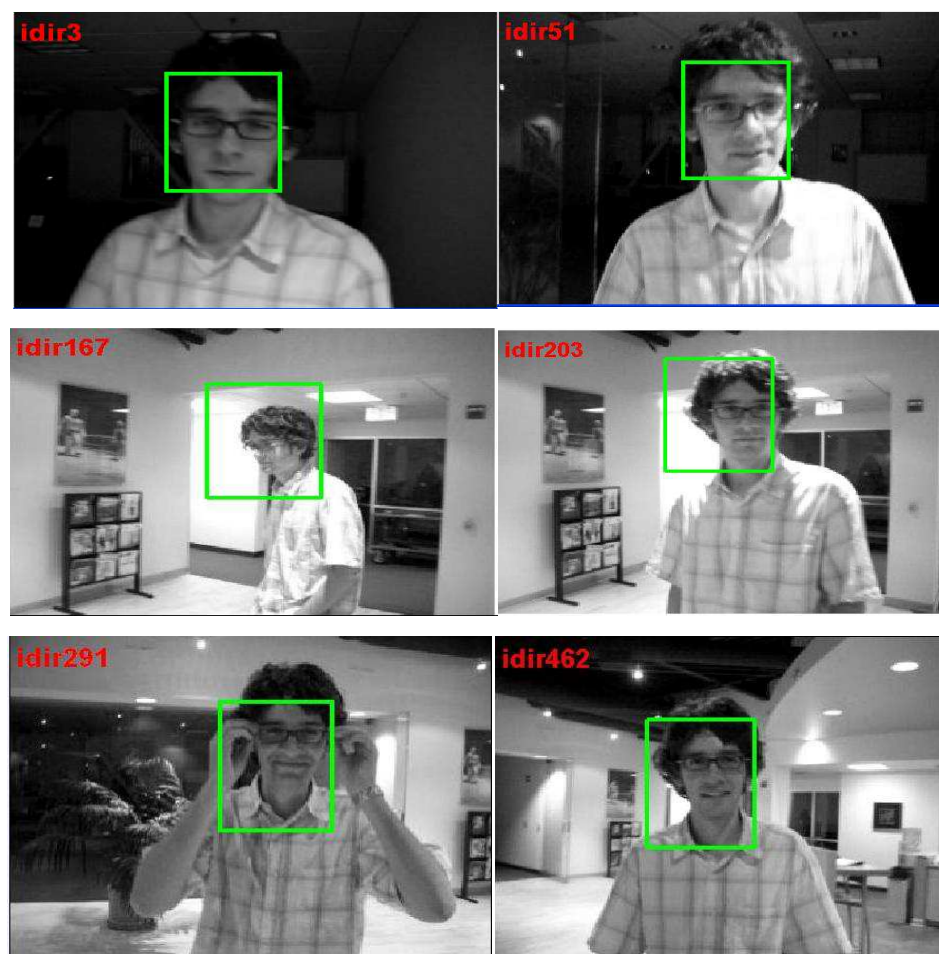


Figure III.9 : résultats montrant l'efficacité du classifieur malgré le changement d'illumination de la scène et de la pose de la cible.

Discussion

Pour la séquence de la figure (III.9), nous pouvons clairement remarquer que la pose de l'objet change progressivement. Quand à l'éclairage de la scène lui, passe d'un changement presque radical (idir3et 51 à idir167) à un changement progressif (idir203 à idir462). Nonobstant de ces contraintes, le tracker parvient à suivre, d'une façon fiable et robuste, la cible contenue dans le rectangle vert.

Nous pouvons dire ainsi que le tracker est robuste au aux changements d'illumination et de pose. En effet, ceci est du au fait que :

- l'apparence de l'objet cible a été bien modélisée par l'application de la projection décrite par l'équation (II.3).
- la mise à jour on-ligne du classifieur est utilisée pour séparer l'objet du fond.

La deuxième séquence testée compte 84 frames, les résultats obtenus, en lui appliquant le classifieur, sont illustrés dans la figure (III.10).

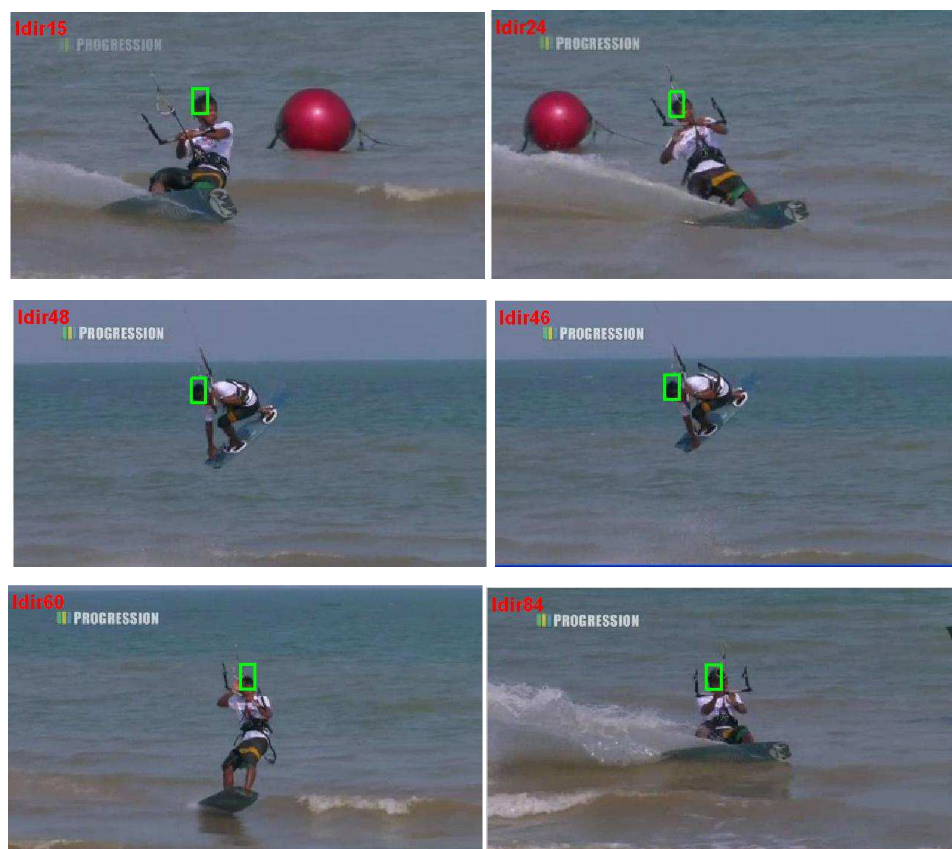


Figure III.10 : performance du classifieur face aux mouvements brusques.

Discussion

En plus du changement d'illumination de la scène, nous pouvons noter que la position du surfeur, dans les frames présentées, change tellement et par fois même, d'une façon brusque en faisant des rotations. Face à ces contraintes, le

tracker réussi, grâce au classifieur et son aspect discriminant, à localiser et à suivre l'objet d'intérêt (avec le rectangle vert).

III.3.3 Les performances du classifieur

Les résultats obtenus dans les figures (III.9 et III.10) nous permettent de conclure sur l'efficacité du classifieur de Bayes-Naïf adopté et appliqué au tracking. En effet, le tracker parvient à localiser sa cible et de la suivre d'une façon efficace dans de circonstances complexes relatives soit :

- au changement d'illumination de la scène, progressivement ou brusquement ;
- Au changement de pose et d'échelle ;
- A la vitesse du mouvement et à sa brutalité (mouvement brusque).

Nous pouvons ainsi conclure que le classifieur adopté, gère le changement d'intensité et d'illumination puisqu'il ne tient pas compte de l'information du background. Il réalise aussi de bonnes performances en termes de localisation de la cible même quand elle est soumise à des mouvements brusques et à des scènes avec de fortes occlusions.

III.4 Conclusion

D'après les résultats de leurs applications au tracking, nous constatons que le modèle de contours actifs adopté est moins performant par rapport au modèle d'apparence utilisé. En effet, là ou où le modèle du contour actif adopté échoue à suivre sa cible, face des contraintes de complexité de la scène, le modèle d'apparence performe et arrive à localiser et suivre sa cible grâce à la simplicité et à la robustesse de son algorithme. En effet, l'algorithme formule la tâche du suivi comme un problème de classification binaire avec mise-à-jour on-ligne dans

le domaine compressé, via une matrice de projection dite creuse, pour la compression des caractéristiques des cibles de premier plan et d'arrière-plan .

L'algorithme combine les mérites des modèles d'apparence génératifs et discriminants pour tenir compte des changements de scène et fournit des résultats concurrents en termes de précision, de robustesse et d'applicabilité en temps réel.

Néanmoins, le modèle de contours actif utilisé réalise avec succès, la segmentation locale ou globale pour le suivi d'un seul ou plusieurs objets selon le choix. Par contre, le classifieur adopté ne peut suivre qu'un seul objet dans la scène.

Conclusion générale

Nous rappelons que le but principal de notre travail était de trouver un ou plusieurs types de modèles qui peuvent répondre aux exigences du suivi d'objet, en présence de complexités et de contrainte dans la scène, notamment à leurs applications au tracking en temps réel. Pour ce faire, deux types de modèles ont été appliqués.

Premièrement, un modèle de contour actif, implémenté par les level-set, qui combine les avantages d'un modèle de contour basé région (C-V) et un autre basé contour (GAC). Les résultats de l'application du modèle adopté nous ont permis de conclure sur son efficacité en tracking en termes de segmentation, de suivi et de rapidité. Sa performance en temps réel est d'autant plus intéressante lorsque l'éclairage de la scène ne varie que très peu ou alors, pas du tout puisqu'il en est sensible.

Deuxièmement, un modèle d'apparence basé sur le classifieur de Bayes-Naïf est appliqué à des séquences vidéo pour le tracking d'un objet d'intérêt. Le classifieur considère le problème du tracking comme une classification binaire avec mise-à-jour on-ligne. En effet, l'algorithme du modèle combine les avantages des modèles d'apparence, discriminatif et génératif pour tenir compte et gérer les changements dans la scène. Les résultats de l'application du classifieur ont révélé son efficacité face aux limites du modèle de contour actif liées au changement d'illumination, à la vitesse du mouvement, au changement brusque d'intensité et de pose. Néanmoins, avec ce classifieur, nous n'avons de l'information que sur la partie de l'image qui sert à suivre le mouvement de la cible, contrairement à l'approche du contour utilisée, qui tient compte de toute l'information de l'image au niveau du pixel.

Conclusion générale

Bien que les deux types de modèles présentent quelques limites, il serait judicieux de lancer comme perspectives la combinaison de leurs avantages pour les exploiter à bon escient notamment :

- Dans des dispositifs d'asservissement visuel,
- Dans le diagnostic de certaines anomalies dans le domaine médical.

Pour terminer, nous espérons que le travail présenté puisse être jugé utile et intéressant et qu'il sera exploité et amélioré par les promotions à venir.

A.I : Définition des contours actifs

Les contours actifs sont définis comme une courbe continue fermée ou non, à extrémités fixes ou non. Ils se déforment à partir d'une position initiale située près de l'objet d'intérêt pour quelques modèles, mais pas nécessairement pour d'autres [13]. Le modèle est soumis à des forces qui le déforment.

Soit Γ une courbe déformable spatialement et temporellement. Son domaine de définition est normalisé à $[0, 1]$. Soient $v(s, t)$ la position d'un point de la courbe à un instant t et x, y ses coordonnées cartésiennes dans l'image (Figure A.1) :

$$v(s, t) = (x(s, t), y(s, t)) \forall (s, t) \in [0, 1] \times [0, +\infty] \quad (\text{A.1})$$

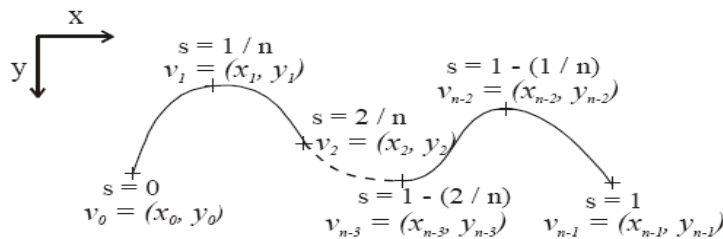


Figure A.1 Exemple de coordonnées cartésiennes et abscisses curvilignes.

Avec s le paramètre de position spatiale le long de la courbe.

Cette représentation nous renseigne que les contours actifs peuvent être un outil de résolution d'un problème de minimisation d'une fonctionnelle d'énergie.

A.II : Modélisation d'un contour actif selon sa représentation

Principalement, deux approches s'opposent pour la modélisation des contours actifs. La première, qualifiée de formulation Lagrangienne représente les contours actifs de manière explicite (ces contours sont souvent paramétriques). La seconde, qualifiée de formulation Eulérienne, représente les contours actifs de manière implicite [13].

A.II.1 : Approche explicite

Parmi les modèles représentés explicitement, on peut considérer que les snakes sont les premiers exemples de contours actifs ayant des propriétés intrinsèques. Ce modèle de courbe élastique, introduit par Kass et *al.* a pour propriété de se positionner autour de l'objet qu'on désire segmenter, en minimisant une fonctionnelle. Cependant, pour ce genre de modèles qui sont dits basés contour, l'initialisation du contour actif demeure un problème, puisque si le contour est initialisé dans une zone où il n'y a plus d'énergie potentielle (force externe), le contour n'évoluera plus, donc, la convergence n'est jamais atteinte. Ce qui explique l'aspect local des contours actifs basés contour.

Par ailleurs, il existe d'autres méthodes qui sont représentées explicitement et qui peuvent s'affranchir de la contrainte d'initialisation, c'est celles basées région.

A.II.2 Approche implicite

Plusieurs méthodes de segmentation par contours actifs se basent sur la modélisation implicite. Parmi ces approches, on trouve celle utilisant les ensembles de niveaux, introduites par Osher et *al.* Son principal avantage est de

permettre la gestion automatique des changements de topologie. En effet, le principe de la méthode consiste à considérer le contour actif comme le niveau zéro d'une fonction $\phi(s, t)$ de dimension supérieure (Figure A.2), avec :

$$\phi(s, t + dt) = \phi(s, t) + \Delta t V(s, t) \quad (\text{A.2})$$

Où : V est la vitesse de propagation. Cette formule, qui est représentée d'une manière générale, gère l'évolution du contour. La fonction ϕ étant une fonction signée de distances par rapport au contour.

Les principaux avantages de cette approche sont :

- * La courbe (ou le contour) peut se diviser ou se fusionner au cours de son évolution : d'où la gestion automatique des changements de topologie.
- * Les résultats obtenus sont très précis, vu que chaque pixel du contour est représenté et repéré par rapport à chaque pixel appartenant à l'image.
- * Par la connaissance des positions de chaque point du contour, on peut reconnaître sa forme.
- * Le contour, considéré comme la courbe de niveau 0, est défini par son voisinage, donc il ne possède pas de structure intrinsèque (paramétrique) (Figure II.2).

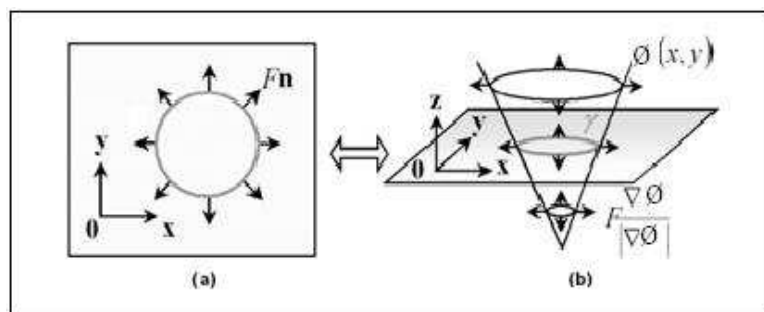


Figure II.2 Représentation implicite d'un contour actif.

A.II.3 construction du modèle de Contour Actif adopté en chapitre III

On rappelle, que la fonctionnelle à minimiser dans les CAG est donnée sous la forme suivante :

$$\frac{\partial \Gamma}{\partial t} = g|\nabla I| \left(\operatorname{div} \left(\left(\frac{\nabla I}{|\nabla I|} \right) + \alpha \right) \right) + \nabla g \cdot \nabla I \quad (\text{A.3})$$

Où : g est la fonction d'arrêt.

Par contre, la fonctionnelle du modèle de Chan-Vese (A.4) est donnée comme suit :

$$\frac{\partial \Gamma}{\partial t} = \delta g \left[\mu \nabla \left(\frac{\nabla I}{|\nabla I|} \right) - \nu - \lambda_1 (I - C_1)^2 + \lambda_2 (I - C_2)^2 \right] \quad (\text{A.4})$$

- **La force signée de pression (SPF) :** C'est une force générée par la différence de la moyenne des intensités à l'intérieur et à l'extérieur de l'objet (C_1 et C_2 respectivement) et leur moyennes. Cette force applique une pression incitant le contour à ce rétrécir ou s'accroître selon son signe. La formule de la SPF est donnée par l'équation suivante :

$$SPF(I(x)) = \frac{I(x) - \frac{C_1 + C_2}{2}}{\max \left(\left| I(x) - \frac{C_1 + C_2}{2} \right| \right)}, \quad x \in \Omega_I \quad (\text{A.5})$$

La représentation de l'apparence est faite par la moyenne des intensités des différentes régions de l'image. Dans la formule (A.5), le numérateur donne le signe de la SPF, ainsi gérer la sens de l'évolution du contour. Le dénominateur lui,

est introduit pour donner une valeur à la SPF allant de $[-1, 1]$, i.e. la normalisation. Une fois que la SPF est calculée, elle sera remplacée dans l'équation (A.3) au lieu de la fonction g . Ce qui nous permet d'aboutir à la formulation suivante :

$$\frac{\partial \Gamma}{\partial t} = \mathbf{fsp}(I) \cdot \left(\operatorname{div} \left(\left(\frac{\nabla I}{|\nabla I|} \right) + \alpha \right) \right) |\nabla I| + \nabla \mathbf{fsp}(I) \cdot \nabla I \quad (\text{A.6})$$

Avec : $\operatorname{div} \left(\left(\frac{\nabla I}{|\nabla I|} \right) \right)$ le terme de courbure, il a pour rôle la régularisation du contour.

En plus à l'introduction de l'information région dans la formule (A.4), le modèle adopté s'en passe des termes suivants :

- $\nabla \mathbf{fsp}(I(\mathbf{x})) \cdot \nabla I$: C'est le terme basé contour utilisé pour attirer le contour vers les bords des objets à détecter. Ce qui est assuré, d'une manière plus robuste [20], avec la nouvelle force basée région SPF.
- $\operatorname{div} \left(\frac{\nabla I}{|\nabla I|} \right)$: C'est un terme de courbure utilisé pour régulariser le contour. Dans le modèle qu'on a adopté, la fonction Level-set est initialisée en tant que constante ayant un signe opposé à l'intérieur et à l'extérieur du contour, ce qui fait que $|\nabla \phi| = 1$. Le terme de régularisation peut être remplacé par $\Delta \phi$, qui est le Laplacien de ϕ . En se basant sur la théorie de (scale-space), on peut dire que l'évolution de la fonction level-set (ϕ) avec le Laplacien, est équivalent au filtrage avec le noyau Gaussien. En s'inspirant de cette idée, le modèle utilise un filtre à noyau Gaussien, indépendamment, du critère d'évolution afin de régulariser le contour, et cela, on effectuant un produit de convolution après chaque itération du processus de convergence.

Nous aboutissons au modèle adopté, décrit par la formule (A.7), qui est implémenté par les ensembles de niveaux :

$$\frac{\partial \phi}{\partial t} = SPF(I(v)) \cdot \alpha |\nabla I|, \quad v \in \Omega. \quad (\text{A.7})$$

Où : α est une constante qui est rajoutée pour augmenter la vitesse de l'évolution.

Si nous voulons segmenter un seul objet désiré (segmentation locale), les conditions suivantes doivent être satisfaites :

→ La fonction ϕ ne doit prendre que deux valeurs possibles :

- ✓ $\phi = +1$ pour toute les valeurs positive de ϕ .
- ✓ $\phi = -1$ ailleurs.

→ Le contour doit être initialisé tout près de l'objet à détecter.

→ Le signe de la SPF doit être choisit d'une manière à ce que l'évolution du contour se fera du milieu moins intense au plus intense, dans le cas où l'objet est d'intensité moindre par rapport à celle de l'image, ou bien, du milieu le plus intense au moins intense et cela si l'objet à segmenter est d'intensité plus importante que celle de l'image.

Bibliographie

- [1] Wren et al : « Real-time tracking of the human body ». *IEEE Trans. On Pattern Analysis and Machine Intelligence*, 19(7): 780-785. (1997).
- [2] Toyama et al/ : « Principles and practice of background maintenance ». In *Proc. 7th IEEE Int. Conf. On Computer Vision (ICCV'99)*, volume I, pages 255-261, Kerkyra (Corfou), Grèce. (1999).
- [3] Rittscher et al/ : « A probabilistic background model for tracking ». In *Proc. 6th European Conf. On Computer Vision (ECCV 2000)*, volume II, pages 336-350, Dublin, Irlande. (2000).
- [4] Heikkilä et Pietikäinen : « A texture-based method for modeling the background and detecting moving objects ». *IEEE Trans. On Pattern Analysis and Machine Intelligence*, 28(4):657-662. (2006).
- [5] Yilmaz, A et al/ : « Object tracking: A survey ». *ACM Comput. Surv.* 38, 4, Article 13, 45 pages. (Dec. 2006),
- [6] Nikolas VERBEKE : « suivi d'objet en mouvement dans une séquence vidéo ». Thèse, PARIS DESCARTES, (2007).
- [7] Julien Richefeu ; « Détection et analyse du mouvement sur système de vision à base de rétine numérique ». Thèse Doctorale, université Paris6, (2006).
- [8] Alain PUJOL : « Contributions à la Classification Sémantique d'Images ». Thèse, université de Lion, (2009).
- [9] Mikram et al : « Multi-Scale Histograms for Kernel-Based Object Tracking ». 'ISCCSP'. (3006).
- [10] Koenderink : « The Structure of Images ». *Biological Cybernetics* 50, 363-370. (1984).

Bibliographie

- [11] Witkin : « Scale-Space Filtering ». International Joint Conference on Artificial Intelligence (IJCAI), 1019-1022. (1983).
- [12] Guillaume Bouchard : « Les modèles génératifs en classification supervisée et applications à la catégorisation d'images et à la fiabilité industrielle ». THESE, Grenoble 1, (2005).
- [13] Weiss and Kulikowski : «Computer Systems That Learn, Classification And Prediction Methods From Statistics, Neural Nets, Machine Learning And Expert Systems ». San Mateo, California Morgan Kaufman Publishers. (1991).
- [14] Vapnik : « *Statistical Learning Theory*. Wiley ». New York. (1998).
- [15] Huang et al : « Image indexing using color correlograms », 'IEEE Conference on Computer Vision and Pattern Recognition (CVPR)', pages 762-768. (2002).
- [16] Avidan, S.: « Support vector tracking ». PAMI 26, 1064-1072. (2004) ».
- [17] Robert.E et al : « A boosting-based system for text categorization ». *Machine Learning*, 39(2/3):135-168, (2000).
- [18] I. Laptev « Periodic Motion Detection and Segmentation via Approximate Sequence Alignment ». *ICCV'05*, Bijing, China, pp. I:816-823. (2005).
- [19] Steven Abney et al, 2003
- [20] I. Laptev : « *Learning realistic human actions from movies* ». In Proc. *CVPR'08*, Anchorage, US. (2008).
- [21] Grabner, H, Grabner, M, Bischof, H: « Real-time tracking via online boosting. In: *BMVC*, pp. 47-56 (2006) ».

Bibliographie

- [22] Hand D et al: principal of data mining ». The MIT Press, 578 pages.(2001).
- [23] Jaillet S et al: « Adéquation des modèles de représentation aux méthodes de catégorisation ». Revue des Nouvelles Technologies de l'Information (RNTI-E-4) Fouille de données complexes, (novembre 2005).
- [24] Ross, D., Lim, J., Lin, R., Yang, M.-H.: « Incremental learning for robust visual tracking». IJCV 77, 125-141 (2008).
- [25] Mei, X., Ling, H.: « Robust visual tracking and vehicle classification via sparse representation ». PAMI 33, 2259-2272 (2011).
- [26] Piccardi, M : « Background subtraction techniques: A review ». In Proc. IEEE Conf. On Systems, Man, and Cybernetics, volume IV, pages 3099-3104, La Hague, Pays-Bas, (2004).
- [27] Nicolas Verbeke et Nicole Vincent : « A PCA-based Technique to Detect Moving Objects ». In Proceedings of the 15th Scandinavian Conference on Image Analysis (SCIA 2007), pages 641{650, Aalborg, Danemark, (juin 2007).
- [28] Idir FEDDAG, Ouardia CHILALI, Samir HACHOUR et Youcef IABADENE : « Le suivi d'objets en mouvement par un modèle de contour actif ». SNIB'10. (2010).
- [29] Idir FEDDAG, Ouardia CHILALI, Samir HACHOUR et Youcef IABADENE : « Study of an active contour model: application in real time tracking ». ICEEA'10 University of Bejaia. (2010).
- [30] Kaihua Zhang et al, « Real-Time Compressive Tracking », in European Conference on Computer Vision (ECCV 2012), Florence, Italy, October, 2012.