

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

ECOLE NATIONALE POLYTECHNIQUE
Département de Génie Electrique



Projet de fin d'études

**En vue de l'obtention du Diplôme
Master en Electronique**

THEME :

*Suivi d'objets en mouvement dans
des séquences d'images*

Présenté par :

Mr. B. YADOUGH

Proposé par :

Mme. S. AIT DAOUD
Mr. C. LARBES

Promotion : Juin 2012

Ecole Nationale Polytechnique
10, Avenue Hacén Badi, El-Harrach, Alger.

Remerciements

J'exprime toute ma reconnaissance à Monsieur C. LARBES d'avoir joué pleinement son rôle de promoteur en étant à mes côtés tout au long de l'étude de nom projet. Ses conseils et orientations m'ont guidé jusqu'à l'aboutissement de ce travail.

Je remercie ma co-promotrice madame S. AIT DAOUD pour ses remarques pertinentes qui ont apporté une amélioration certaine à ce travail. Je la remercie également d'avoir examiné pleinement mon travail et proposé des modifications constructives.

Je tiens à remercier Monsieur L. ABDELOUAL, MA-A à l'Ecole Nationale Polytechnique d'Alger, d'avoir accepté de présider le jury.

Mes remerciements vont aussi à Monsieur M. O. TAGHI, MA-A à l'Ecole Nationale Polytechnique, d'avoir bien voulu examiner mon travail.

Que tous ceux et toutes celles qui ont contribué de près ou de loin à l'aboutissement de mon travail trouvent ici mes sincères remerciements et l'expression de ma reconnaissance.

Dédicaces

Je dédie ce mémoire, aux personnes qui me sont les plus chères :

A mes parents qui m'ont énormément soutenu dans les moments

Les plus difficiles, partagé mes joies et mes peines, qui se sont

Toujours sacrifiés à mes dépends.

A mes frères et à mes sœurs

A toute ma famille.

A mes amis et collègues.

Résumé

Le suivi d'objets en mouvement est l'une des applications essentielles de la détection d'objets mobiles, qui est elle-même un vaste domaine du traitement et d'analyse d'une séquence d'images. Le but de ce travail est d'étudier deux méthodes différentes de suivi d'objets mobiles, le suivi par contour actif qui consiste à suivre l'objet mobile en l'entourant par un contour, et le suivi par maillage qui consiste à suivre l'objet en mouvement par la déformation des polygones constituant le maillage selon le déplacement de l'objet suivi.

Mots clés : suivi, détection, estimation, contour, maillage et séquence

Abstract

Track moving objects is one of the core applications of motion detection, which is itself a broad field of processing. The purpose of this work is to study two different methods of tracking moving objects, the first based on the active contour which follows the moving object by surrounding it by a contour, and the second is tracking by mesh which follows the moving object by the deformation of its polygons according to the displacement of the tracked object.

Keywords : monitoring, detection, estimation, contour, mesh and sequence

ملخص

تتبع الحركة من أهم التطبيقات الناتجة عن اكتشاف الأجسام المتحركة في مجموعة صور متتالية و هذه الأخيرة مجال واسع لاختبار المشهد. الهدف من هذا العمل يقوم على دراسة طريقتين مختلفتين لتتبع الحركة في المشهد: التتبع بواسطة ك و التتبع بواسطة "المياح" التي بدورها تقوم على تصوير الجسم المتحرك بواسطة أشكال تتغير بتحريك الجسم.

: تتبع، والكشف، تقدير، كفاف، أشكال وتسلسل

Table des matières

Introduction générale	1
------------------------------------	---

Chapitre I : Etat de l'art de la détection et l'estimation du mouvement

I.1 Introduction.....	2
I.2 Détection de mouvement.....	2
I.2.1 Définition.....	2
I.2.2 Différents méthodes de détection de mouvement.....	2
I.2.2.1 Détection sans modélisation.....	4
I.2.2.2 Détection par modélisation locale.....	5
I.2.2.3 Détection par modélisation semi-locale.....	11
I.2.2.4 Détection par modélisation globale.....	13
I.3 Estimation et calcul de mouvement.....	14
I.3.1 Définition.....	14
I.3.2 mouvement réel, mouvement apparent, mouvement estimé.....	14
I.3.3 Estimation de mouvement.....	16
I.4 Conclusion.....	20

Chapitre II : Suivi d'objets en mouvement

II.1 Introduction.....	21
II.2 Méthode de suivi par contour actif.....	23
II.2.1 Contour actif par approche contour.....	23
II.2.2 Contour actif par approche région.....	29
II.2.3 Application du modèle de contour actif pour le suivi de joueurs de football.....	30
II.3 Méthode de suivi par maillage.....	34

II.3.1 Définitions et propriétés.....	34
II.3.2 Suivi d'objets en mouvement par maillage.....	39
II.3 Conclusion.....	48
Conclusion générale.....	49

Liste des figures

Chapitre I

FIG I.1: classification des méthodes de détection de mouvement	3
FIG I.2 : modèle de trois gaussiennes	8
FIG I.3 : Résultats pour différentes valeur de	9
FIG I.4 : illustration des mouvements réels et apparents dans un système optique de vue ...	15
FIG I.5 : Estimation directe et inverse du vecteur déplacement	16
FIG I.6 : projection de l'objet sur le plan image	17
FIG I.7 : les variations spatio-temporelles.....	19

Chapitre II

FIG II.1 : La courbe C.....	23
FIG II.2 : organigramme du suivi d'objets mobiles par contour actif.....	26
FIG II.3: représentation de l'évolution du contour actif.....	28
FIG II.4 : forces régissant l'évolution du contour actif.....	29
FIG II.5: processus de suivi par contour actif.....	31
FIG II.6: contours initiaux et finaux obtenus pour les deux premières images d'une séquence.....	32
FIG II.7 : Contours initiaux et finaux obtenus pour les 20 ^e , 40 ^e , 60 ^e , et 80 ^e images de la séquence.....	33
FIG II.8: Contours initiaux et finaux montrant: (a) la sensibilité au fond de la scène (particulièrement les lignes blanches), et (b) l'impossibilité du contour actif à se séparer afin de suivre plusieurs objets.....	34
FIG II.9 : les nœuds, les arêtes, et les faces définissant un maillage.....	35
FIG II.10 : Exemples de maillages triangulaires (a), (b) et (c), et quadrangulaire (d)....	35

FIG II.11 : exemples de maillages conformes et non conformes.....	36
FIG II.12 : exemple de maillage irrégulier.....	37
FIG II.13 : exemple d'utilisation du maillage statique (a) image originale ; (b) maillage non conforme adapté à l'image.....	37
FIG II.14 : Exemple de grille active.....	38
FIG II.15 : Mise en correspondance par maillage triangulaire.....	38
FIG II.16 : Maillage déformable	39
FIG II.17 : principe du calcul de la DFD locale et la DFD globale pour une détection arrière.....	42
FIG II.18 : division d'une maille possédant une ou plusieurs arêtes trop longues.....	44
FIG II.19 : suivi simple sur la séquence foreman.....	44
FIG II.20 : DFD globale sur le suivi de la séquence foreman.....	45
FIG II.21 : suivi d'une partie d'un personnage sur laquelle passe une ombre portée.....	45
FIG II.22 : suivi par maillage d'un personnage occulté, avec peu de rééité (a), et forte rééité (b).....	46
FIG II.23 : image de l'erreur DFD en valeur absolue pour l'image 6395 avec faible (a) et forte (b) rééité.....	46
FIG II.24 : DFD globale pour la séquence, avec faible (a) et forte (b) rigidité.....	47

Introduction générale

En traitement d'images et en vision par ordinateur, le suivi d'objets en mouvement dans un environnement connu ou inconnu est couramment étudié. Il s'appuie sur des propriétés d'invariance des objets d'intérêt. L'invariance peut concerner la géométrie de la scène ou des objets, l'apparence des objets (c'est-à-dire la photométrie ou la couleur) ou encore la cinématique (par exemple les contraintes spatio-temporelles).

En vision artificielle, le suivi d'objets est au cœur de nombreuses applications. Donnons deux exemples. Le suivi peut être un outil pour donner des capacités d'autonomie visuelle à des automates ou des robots. Dans ce cas, la perception visuelle est un préalable à l'action et nécessite un apprentissage pour établir des liens entre les causes et les actions à produire en réponse. Le suivi d'objets peut aussi permettre d'automatiser des tâches répétitives. Par exemple, dans le cas de la vidéo-surveillance, la perception visuelle a pour objectif d'aider l'être humain et de l'informer d'un problème en détectant des événements anormaux. Cela nécessite donc de nouveau un apprentissage.

L'objet de ce projet de fin d'études consiste qu'à partir d'une séquence d'images représentant un objet déformable en mouvement, on détermine, sa position dans chaque image et donc son déplacement ou encore sa déformation d'une image à l'autre.

Le premier chapitre comprend un état de l'art sur la détection et l'estimation de mouvement. En citant les différentes méthodes de détection de mouvement, et les classifieur en fonction de la modélisation de l'arrière-plan. Et en expliquant l'estimation et le calcul de mouvement.

Le deuxième chapitre traite le suivi d'objets en mouvement, en détaillant deux méthodes différentes. L'une sur le suivi d'objets par contour actif qui peut être réparti en deux catégories, la première correspond aux approches contours et la seconde aux approches régions. Et l'autre sur le suivi d'objets en mouvement par maillage.

Chapitre I

*Etat de l'art sur la détection et l'estimation du
mouvement*

I.1 Introduction

L'analyse du mouvement est un problème fondamental du traitement des séquences d'images. Il s'agit en particulier d'un problème central pour des domaines tels que la post-production, la compression vidéo, la vidéo surveillance, la reconnaissance d'actions, l'analyse sémantique de scènes, ...etc.

L'analyse du mouvement dans une séquence d'images est un vaste sujet qui regroupe plusieurs problématiques. On peut notamment citer l'estimation du champ de mouvement, la détection, et la segmentation de chaque région d'une image ayant un mouvement distinct des autres.[1]

L'estimation du mouvement est un élément essentiel pour l'analyse d'une scène. Elle génère un grand nombre d'applications en robotique, en imagerie biomédicale, en compression de vidéos, et beaucoup d'autres encore.[2]

I.2: Détection de mouvement

I.2.1 : Définition

Le terme Détection de mouvement (Motion détection), indique uniquement que l'on parle d'une méthode qui a pour objet de trouver en quels points de l'image un mouvement a eu lieu. Un algorithme ayant cet objectif fournit en sortie une variable quantitative (quantité de mouvement) ou qualitative (booléenne) pour tout pixel de chaque image d'entrée.

La plupart des états de l'art publiés jusqu'à présent sur la détection de mouvement, sont fortement orientés vers un domaine d'application précis, ou alors inclus dans un état de l'art plus vaste traitant des systèmes complets de suivi d'objets mobiles.

I.2.2 : Différentes méthodes de détection de mouvement

Avec classification hiérarchisée des différentes méthodes de détection de mouvement. Nous distinguerons quatre grandes familles de méthodes en fonction de la modélisation de l'arrière-plan.

- Détection sans modélisation de l'arrière-plan :

Ces méthodes consistent à détecter le mouvement par le calcul en tout point de l'image d'une quantité mathématique qui est fonction de l'intensité ou de la couleur de l'ensemble des pixels et qui est censée refléter l'importance du mouvement visible dans la scène. Comme exemple en cite La dérivée temporelle de l'intensité lumineuse, l'entropie spatio-temporelle de l'image et la norme du flot optique.

- Modélisation locale de l'arrière-plan :

Ces méthodes consistent à associer à tout point de l'image une valeur ou une fonction permettant de modéliser l'apparence de l'arrière-plan en ce point. Le modèle d'apparence de l'arrière-plan en un point ne dépend que des observations qui ont eu lieu en ce point.

Les autres pixels de l'image n'interviennent pas. La grande majorité des méthodes présentées dans la littérature bâtissent un modèle statistique (ensemble de paramètres d'une loi, ou ensemble d'échantillons), mais il peut s'agir d'un processus stochastique, d'un filtre prédictif ou simplement d'une valeur d'intensité.

- Modélisation semi-locale de l'arrière-plan :

Ces méthodes sont très semblables à celles de la catégorie précédente, à la différence près que la modélisation de l'arrière-plan en un point dépend des observations qui ont eu lieu dans un certain voisinage de ce point, ou dans la région de l'image à laquelle il appartient.

- Modélisation globale de l'arrière-plan :

Ces méthodes utilisent à chaque instant l'ensemble des observations pour construire un modèle de l'ensemble de l'arrière-plan.

La figure suivante résume cette classification :

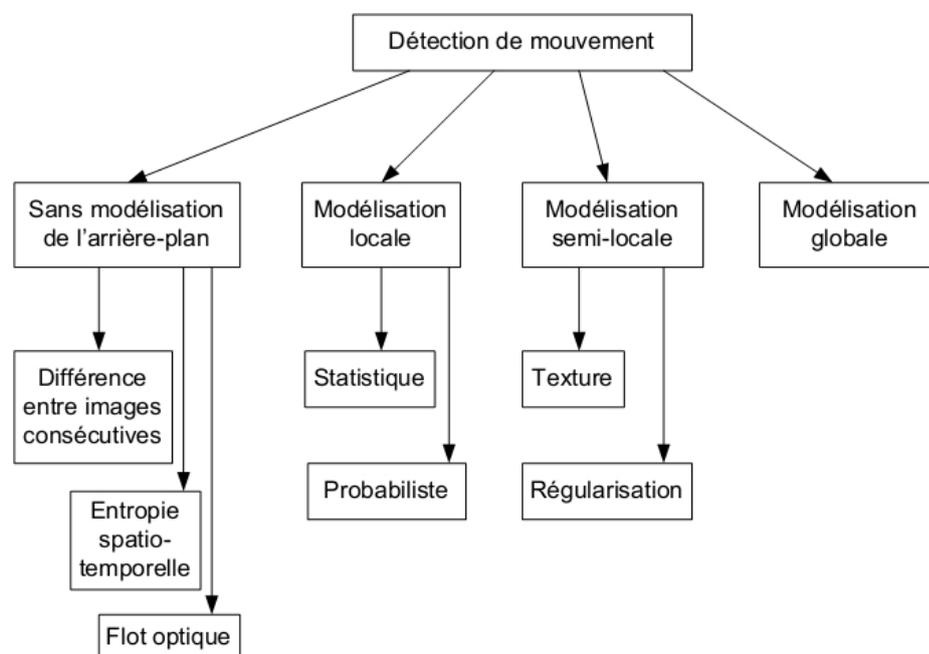


FIG I.1: Classification des méthodes de détection de mouvement

1.2.2.1 : Détection sans modélisation

On désigne par $I(t)$ l'image présente dans le flux d'entrée au temps t et par $I(x,y,t)$ le pixel de coordonnées $x = (x,y)$ dans cette même image. Nous appellerons E le domaine de définition des coordonnées des pixels. Les dérivées spatiales horizontale et verticale de l'image seront notées respectivement $I_x(x,y,t)$ et $I_y(x,y,t)$. La dérivée temporelle sera notée $I_t(x,y,t)$. Le gradient de l'image au point (x,y) et au temps t sera quant à lui abusivement noté :

$$I(x,y,t) = [I_x(x,y,t) \ I_y(x,y,t)]^T$$

Dérivé temporelle :

La manière la plus intuitive de détecter les zones en mouvement dans un champ de vision est de mesurer le changement d'apparence des pixels entre deux trames consécutives, soit la dérivée temporelle en tout point.

Etant un signal discret a une dimension, l'estimation de la dérivée temporelle instantanée du signal au temps t est donnée par :

$$\forall (x,y) \in E \quad \forall t > 0 \quad I_t(x,y,t) \approx |I(x,y,t) - I(x,y,t-1)|$$

Cette méthode se montre peu robuste face à des phénomènes tels que les mouvements lents, les arrêts brefs d'un objet en mouvement, ou encore la présence de trames redondantes dans certaines séquences vidéo. Il convient donc d'effectuer un lissage temporel de la séquence, c'est-à-dire d'appliquer un opérateur de moyenne mobile à la mesure obtenue. Ceci peut être fait à l'aide d'une matrice de même taille que les trames d'entrée, appelée accumulateur, et que nous noterons A .

$$\forall (x,y) \in E \quad \begin{cases} A(x,y,0) = 0 \\ \forall t > 0 \quad A(x,y,t) = w_A A(x,y,t-1) + (1-w_A) I_t(x,y,t) \end{cases}$$

Avec $0 \leq w_A \leq 1$. Le terme w_A pondère la contribution des mesures passées par rapport à la dernière mesure en date. Avec un w_A faible, les effets du lissage temporel sont peu visibles, et les problèmes qui avaient motivé l'utilisation de cette méthode risquent d'apparaître tout de même. Avec un w_A élevé, le lissage est important et un effet de persistance des entités détectées peut se produire (effet fantôme).

Entropie spatio-temporelle :

L'entropie est une mesure issue de la thermodynamique, associée au degré de désordre d'un système. Dans notre cas, il s'agira de mesurer en chaque point, la (variabilité) de la grandeur mesurée. Concrètement, plus l'intensité lumineuse (ou la couleur, ou le gradient, etc.) aura pris de valeurs dissemblables en un point pendant un certain intervalle de temps, plus l'entropie sera élevée en ce point.

Des auteurs proposent, pour calculer l'entropie, une méthode à base d'histogrammes spatio-temporels. Il s'agit, en chaque pixel désigné par (x,y,t) , de créer un histogramme à partir des points (u,v,τ) d'un voisinage spatio-temporel de diamètre W dans le domaine spatial,

et de longueur L dans le domaine temporel. Ainsi, pour une image à Q niveaux de gris, on note $H_{x,y,t}(q)$ la fréquence de la q -ème classe de l'histogramme ($0 \leq q \leq Q - 1$) associée au point (x,y,t) . Si on normalise cet histogramme, on obtient une fonction de densité de probabilité.

$$\forall (x,y) \in E \quad \forall t \geq L \quad P_{x,y,t}(q) = \frac{H_{x,y,t}(q)}{W^2L} \text{ avec } \sum_{q=0}^{Q-1} P_{x,y,t}(q) = 1$$

Cette expression nous permet de déterminer en tout point l'entropie associée à la répartition des niveaux de gris au voisinage du dit point.

$$E(x,y,t) = - \sum_{q=0}^{Q-1} P_{x,y,t}(q) \log(P_{x,y,t}(q))$$

L'entropie ainsi calculée peut être quantifiée en 256 niveaux, et être représentée sous la forme d'une image appelée STEI (Spatial Temporal Entropy Image). Dans cette image, sont mis en évidence les points où les variations d'intensité lumineuse sont importantes au cours du temps, et dans une moindre mesure, au sein d'un voisinage spatial. Les zones de mouvement sont donc bien mises en évidence, mais les contours (zones de fort gradient) le sont également. On peut atténuer cet effet en donnant plus d'importance aux variations observées dans le domaine temporel que dans le domaine spatial, mais ce phénomène persistera toujours.

Flot optique

Alors que la dérivée temporelle quantifie la variation de l'aspect de chaque pixel considéré individuellement, le flot optique est un champ de vecteurs à deux dimensions représentant la projection sur le plan image du mouvement réel observé (tridimensionnel).

De nombreuses méthodes ont été proposées depuis celles-ci sont détaillées dans plusieurs états de l'art. Quelle que soit la méthode choisie, le calcul du flot optique reste une opération très coûteuse en temps de calcul. Le temps-réel peut néanmoins être atteint en sous-échantillonnant les trames et en choisissant un algorithme rapide.

I.2.2.2 : Détection par modélisation locale

Modélisation par une image :

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, et il n'est pas toujours possible d'obtenir une image de la scène totalement vide. Dans ces conditions, il est nécessaire de mettre à jour l'image de l'arrière-plan.

Des auteurs proposent d'utiliser la différence entre images consécutives pour y parvenir. Ils considèrent la première image de la séquence comme une première approximation du modèle de l'arrière-plan. Ensuite, à chaque nouvelle trame, la différence par rapport à l'image précédente est calculée, et les pixels où aucun mouvement n'est détecté sont utilisés pour mettre à jour le modèle du fond. Pour plus de robustesse, les auteurs préconisent de ne considérer que les points auxquels la dérivée temporelle a été négligeable pendant un certain intervalle de temps.

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 appris, censé représenter 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 avant-plan.

Des auteurs 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. Le même procédé peut être utilisé sur des images en couleur. Ce modèle permet d'obtenir de bons résultats pour des scènes d'intérieur où l'arrière-plan est parfaitement statique, mais en environnement extérieur, des phénomènes périodiques tels que l'ondulation d'une surface d'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.[3]

Stauffer et Grimson ont proposés de modéliser chaque pixel de l'arrière-plan par un mélange de K gaussiennes. La probabilité d'observer en x la valeur z au temps t est donc :

$$P_x(Z_t) = \sum_{i=1}^k w_{i,t} \eta(z_t, \mu_{i,t}, \Sigma_{i,t})$$

Où :

$w_{i,t}$: Poids accordé à la i -ème gaussienne au temps t

η : la fonction gaussienne de densité de probabilité

$\mu_{i,t}$: la moyenne de la i -ème gaussienne au temps t

$\Sigma_{i,t}$: la matrice de variance-covariance de la i -ème gaussienne au temps t

K est compris entre 3 et 5, et les composantes couleurs sont considérées comme indépendantes et de même variance, soit :

$$\Sigma_{i,t} = \sigma_{i,t}^2 I$$

Avec :

I: la matrice identité de dimension (n X n).

$\sigma_{i,t}^2$: vecteur ligne des variances de dimension (n)

On peut également écrire la matrice de covariance sous la forme:[3]

$$\Sigma_{i,t} = \begin{pmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_N^2 \end{pmatrix}$$

Le processus de détection de mouvement se décompose en trois étapes : l'initialisation des paramètres du mélange, classification des pixels et enfin la mise à jour des paramètres du mélange selon la classification attribuée :

- *L'initialisation des paramètres :*

Le poids w , la moyenne μ , et l'écart type σ peuvent être initialisés en utilisant la méthode standard EM (Espérance-Maximisation) en maximisant la vraisemblance de manière itérative. Cette méthode est utilisée en cas d'insuffisance des informations des observations ou manque de données. [4]

- Classification des pixels

Après initialisation des paramètres, une première détection des objets mobiles est possible, en classant les pixels comme appartenant à l'arrière-plan ou le premier plan (objets en mouvement).

Pour cela les K gaussiennes du mélange sont ordonnées selon la valeur du rapport $r_j = w_j / \sigma_j^2$, puis, l'arrière-plan sera modélisé par les B premières distributions gaussiennes, alors que les autres distributions représentent les objets en mouvement. Le nombre de gaussiennes B est déterminé selon la relation suivante :

$$B = \operatorname{argmin}_b \left(\sum_{i=1}^b w_{i,t} > T \right)$$

Où T est un seuil de décision entre distribution du fond et objets en mouvement, sa valeur dépend de la complexité de la scène. Il est recommandé de choisir une petite valeur de T pour les scènes simples, et une grande valeur de T pour les scènes complexes ayant plusieurs modes de mouvement.

un pixel du fond correspond à un grand poids avec une faible variance du fait que le fond est plus présent que les objets en mouvements et que sa valeur est pratiquement constante. Ainsi, il correspond à une ou plusieurs gaussiennes parmi les B distributions gaussiennes de l'arrière-plan.

Les B distributions gaussiennes des fonds seront ordonnées dans une liste dont les distributions les plus probables restent au sommet tandis que les distributions des fonds les moins probables gravitent, et peuvent éventuellement être remplacées par de nouvelles distributions.

À l'arrivée d'une nouvelle trame, à l'instant $t+1$, chaque pixel subit un test d'identification pour le classer comme appartenant à l'arrière-plan ou à un objet en mouvement, en vérifiant l'inégalité suivante :

$$\text{SQRT}((X_{t+1} - \mu_{i,t})^T \Sigma^{-1} (X_{t+1} - \mu_{i,t})) > k$$

Où X_{t+1} est la valeur du pixel dans la trame à l'instant $t+1$, et k un seuil constant déterminé empiriquement égale à 2,5.

Si l'inégalité précédente est vérifiée pour toutes les K distributions gaussiennes, le pixel appartient au premier plan (objet en mouvement). Le graphe suivant illustre un exemple de 3 gaussiennes où le pixel introduit ne correspond à aucune d'elles :

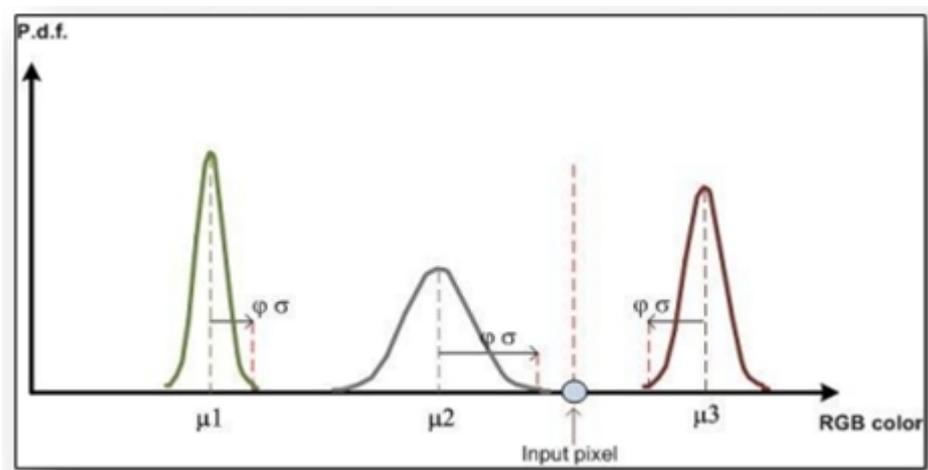


FIG I.2 : Modèle de trois gaussiennes

En revanche, si l'inégalité précédente n'est pas vérifiée, alors le pixel correspond à l'une des K distributions gaussiennes, d'où l'un des deux cas suivants doit avoir lieu :

- Cas 1 : le pixel correspond à l'une des B distributions gaussiennes, donc il appartient à l'arrière-plan ;
- Cas 2 : le pixel correspond à l'une des autres distributions, donc il appartient au premier-plan (objet mobile) ;

- Mise à jour des paramètres

La mise à jour des paramètres est effectuée selon s'il y a correspondance des pixels avec les distributions gaussiennes ou pas :

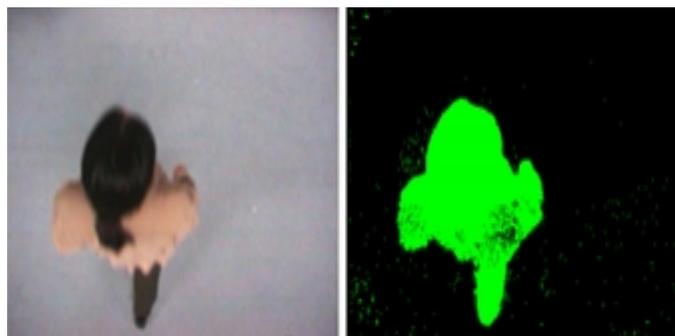
- S'il y a correspondance avec l'une des K gaussiennes, la mise à jour est effectuée sur chacun des paramètres de la gaussienne correspondante comme suite:

Le poids $w_{i,t}$ devient :

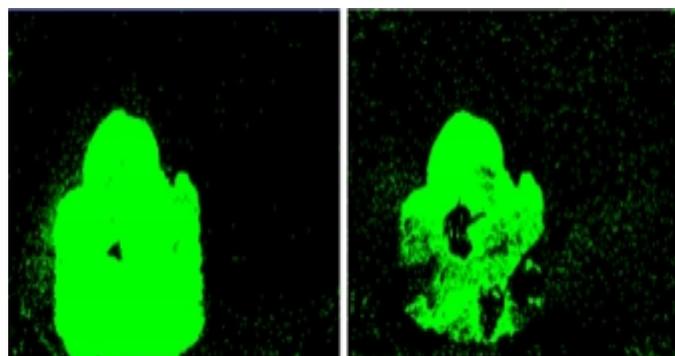
$$w_{i,t+1} = (1 - \epsilon) w_{i,t} +$$

Avec $\epsilon \in [0,1]$ un taux constant qui détermine la vitesse de changement des paramètres. Sa valeur est prédéfinie empiriquement à l'aide des séquences de test, et en fonction de la rapidité de changement du fond. Ainsi, un fond qui ne change que lentement nécessite une valeur faible de ϵ , alors qu'un fond qui change rapidement nécessite une valeur large de ϵ .

La figure suivante illustre un exemple qui montre l'effet des différentes valeurs du taux ϵ sur les résultats de détection :



(a) Image d'origine (b) Résultat pour $\epsilon = 0,1$



(c) Résultat pour $\epsilon = 0,01$ (d) Résultat pour $\epsilon = 0,5$

FIG 1.3 : Résultats pour différentes valeurs de ϵ

On remarque que la valeur que l'image caractérisée par $\alpha=0,1$ donne de meilleurs résultats de détection. Cette valeur se conforme avec la simplicité de la scène.

La moyenne $\mu_{i,t}$ et la variance $\sigma_{i,t}^2$ deviennent :

$$\mu_{i,t+1} = (1-\alpha)\mu_{i,t} + \alpha X_{t+1}$$
$$\sigma_{i,t+1}^2 = (1-\alpha)\sigma_{i,t}^2 + \alpha(X_{t+1}-\mu_{i,t+1})(X_{t+1}-\mu_{i,t+1})^T$$

Avec un taux de mise à jour de la moyenne et la variance égale à :

$$\alpha = \frac{\sigma_b}{\sigma_b + \sigma_{i,t}}$$

Pour les autres distributions dont il n'y avait pas correspondance la moyenne μ et la variance σ^2 restent inchangés, seul le poids $w_{i,t}$ devient :

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

- S'il n'y a aucune correspondance avec les K gaussiennes, alors la distribution la moins probable du rang K est remplacée par une autre distribution ayant comme paramètres :

$$w_{i,t+1} = \text{Le plus faible des poids précédents}$$

$$\mu_{i,t+1} = X_{t+1}$$

$$\sigma_{i,t+1}^2 = \text{La plus grande des variances initiales}$$

Après avoir établi l'étape de mise à jour, la prochaine détection des objets mobiles peut avoir lieu, et le cycle de détection continue. [4][5]

Modélisation prédictive :

Une autre approche, assez semblable à l'approche statistique, consiste à utiliser un filtre de Wiener ou de Kalman pour prédire la prochaine valeur que l'on devrait observer en chaque point. C'est l'écart entre la prédiction et l'observation qui sera utilisé pour estimer l'amplitude du mouvement. Ces méthodes permettent de gérer les problèmes d'arrière-plans non statiques, et les perturbations intervenant à intervalles réguliers (par exemple, la bande horizontale qui apparaît lorsque l'on filme un écran à tube cathodique).

Des auteurs proposent une méthode à trois niveaux sémantiques (local, semi-local, global). La segmentation au niveau local est effectuée à l'aide d'un filtrage prédictif de Wiener. Le filtre de Wiener permet de construire une valeur estimée \hat{z}_t de la valeur que l'on devrait observer à l'instant t , à partir d'un échantillon de N mesures bruitées.

$$\hat{z}_t = \sum_{i=1}^N w_i z_{t-i}$$

Où les w_i sont des coefficients pondérateurs calculés de manière à minimiser l'erreur quadratique moyenne d'estimation (t).

$$\xi(t) = E[|z_t - \hat{z}_t|^2]$$

Plus fréquemment, le filtrage prédictif est réalisé à l'aide d'un filtre de Kalman. La méthode suppose que la meilleure information que l'on puisse avoir sur l'état d'un système est obtenue par le calcul d'une estimation qui fait explicitement mention du bruit enregistré lors de l'observation.

1.2.2.3 : Détection par modélisation semi-locale

Détection par région

Etant donnée la nature progressive des mouvements généralement observés dans les séquences à analyser, et compte tenu des imprécisions dues aux caméras vidéo utilisées, certains auteurs 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.

Des auteurs considèrent l'ensemble des zones carrées 3X3 non recouvrantes de l'image et bâtissent un modèle de Markov caché (MMC) pour décider si cette zone appartient à l'arrière-plan, a un objet mobile, ou à une ombre. Le MMC possède donc 3 états : arrière-plan, objet mobile, ombre. Le nombre de symboles observables est 256, puisqu'on observe une intensité lumineuse. Afin de prendre en compte l'intensité moyenne de la zone ainsi que son homogénéité, deux observations sont faites à chaque instant. Ces observations sont obtenues par application d'un filtre de convolution moyenneur 3X3 au centre de la zone et par le calcul de la norme du gradient de Sobel. Les auteurs déclarent que ces deux mesures sont statistiquement indépendantes. Les probabilités initiales et de transition sont initialisées en annotant manuellement une séquence de test et en mesurant en tout point la durée moyenne du passage par chacun des états et la proportion de temps passé dans chaque état. Les probabilités d'émission sont différentes pour chaque état. Quand le système est dans l'état (objet mobile), la distribution des intensités est considérée comme uniforme car on ne peut pas préjuger de l'apparence des objets en mouvement. Les paramètres du modèle sont ensuite optimisés de manière incrémentale par l'algorithme de Baum-Welch qui est une version généralisée de l'algorithme EM (Expectation Maximization). La classification des régions entre arrière-plan, objet mobile ou ombre se fait en sélectionnant la probabilité d'émission la plus élevée.

Caractérisation par la texture :

La prise en compte du voisinage des points peut également être réalisée en calculant en tout point un vecteur caractérisant la texture à cet endroit, et en utilisant l'ensemble des vecteurs calculés comme espace de représentation des données.

Des auteurs utilisent le codage LBP (Local Binary Patterns). Le code LBP du point (x,y) est le mot binaire obtenu en concaténant l'intensité souillée de tous les pixels situés dans un voisinage de (x,y) . Le voisinage est constitué de P points $\{(u_i, v_i)\}_{i=1}^P$ uniformément répartis sur un cercle de rayon R centré en (x,y) . La valeur du seuil utilisé est l'intensité du point (x,y) .

$$LBP_{P,R}(x,y) = \sum_{i=1}^P s(I(u_i, v_i) - I(x,y))2^{i-1}, \text{ avec } s(g) = \begin{cases} 1 & \text{si } g \geq 0, \\ 0 & \text{sinon} \end{cases}$$

D'une manière assez similaire à celle de Stauffer et Grimson, les auteurs conçoivent le modèle d'arrière-plan en un lieu donné comme un ensemble de modes dotés d'un poids. Ici, les modes ne sont pas des distributions gaussiennes, mais des histogrammes de codes LBP calculés sur des blocs carrés partiellement recouvrant. Le nombre K d'histogrammes utilisés pour modéliser une région de l'arrière-plan est choisi empiriquement. Pour un bloc donné, à chaque nouvelle image, un nouvel histogramme h est calculé et comparé à chacun des K histogrammes $\{(m^i)\}_{i=1}^K$ qui constituent le modèle. La mesure de similarité utilisée est l'intersection d'histogrammes.

$$\cap(h, m^i) = \sum_{j=1}^N \min(h_j, m_j^i)$$

Où N est le nombre de classes des histogrammes. Si aucun des histogrammes du modèle n'est suffisamment proche de h , l'histogramme du modèle qui possède le poids le plus faible est remplacé par h .

Si, en revanche, le modèle contient un histogramme suffisamment similaire à h , m est mis à jour ainsi que le poids de l'ensemble des histogrammes du modèle :

$$m = \alpha h + (1 - \alpha)m$$

$$\forall i \in \llbracket 1, K \rrbracket \quad w_i = \beta M_i + (1 - \beta)w_i$$

Où w_i est le poids de l'histogramme m^i , M_i est un indicateur valant 1 si m^i est l'histogramme du modèle le plus similaire à h et 0 sinon, et α et β sont des coefficients choisis dans l'intervalle $[0,1]$ de manière à régler la vitesse de mise à jour.

Comme la méthode de Stauffer et Grimson, la segmentation entre avant-plan et arrière-plan est réalisée en classant les histogrammes du modèle par poids décroissants et en attribuant à l'arrière-plan les b premiers histogrammes dont la somme des poids dépasse un certain seuil.

Régularisation a posteriori :

Plusieurs auteurs proposent une application dite (multicouches), c'est-à-dire, à plusieurs niveaux sémantiques. Le niveau le plus bas bâtit un modèle local de l'arrière-plan qui permet de réaliser une première estimation de la segmentation des objets mobiles. Ensuite, un second procédé analyse les résultats obtenus et les régularise de manière à augmenter la consistance des régions détectées dans un voisinage spatial et/ou temporel.

Après avoir segmenté l'arrière-plan à l'aide d'une modélisation locale, Des auteurs considèrent que parmi les points détectés, ceux ou la dérivé temporelle est importante appartiennent nécessairement a des objets mobiles. Dans ce cas, un histogramme normalisé des régions trouvées est construit, et les régions sont utilisées comme germes par un algorithme de segmentation par croissance de région utilisant l'histogramme comme critère d'arrêt.

1.2.2.4 : Détection par modélisation globale

Basculement entre plusieurs modèles :

Pour prendre en compte la totalité de la scène dans le processus de segmentation entre avant-plan et arrière-plan, Des auteurs décident de garder en mémoire k modèles de l'arrière-plan. 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.

Initialement, les k modèles sont acquis en exécutant un algorithme k-means sur les images d'une séquence d'apprentissage. En cours de traitement, les modèles peuvent être mis à jour si une large majorité des pixels de l'image sont détectés comme étant en mouvement.

Un modèle de Markov caché modifie automatiquement sa topologie quand il est confronté à un environnement dynamique. Les auteurs utilisent cet outil pour gérer les difficultés liées aux changements soudains d'éclairage. Typiquement, le MMC aura deux états (jour/nuit, par exemple) et à chaque état sera associé un modèle statistique local de l'arrière-plan.

Espaces vectoriels :

Une autre manière 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 méthode des *eigenbackgrounds* (arrière-plans propres) constitue la première application de l'analyse de données à la détection de mouvement. [3]

1.3: Estimation et calcul du mouvement

1.3.1 : Définition

L'estimation de mouvement est une quantification du mouvement simple, par vecteurs de translation, mettant en œuvre des méthodes de calcul de trajectoire. Les applications de l'estimation de mouvement sont surtout la réduction de la redondance temporelle pour la compression et l'analyse de scène. [6]

1.3.2 : mouvement réel, mouvement apparent, mouvement estimé

Les images représentent souvent la projection de scènes réelles 3D. C'est pourquoi le mouvement observé (ou mouvement apparent) dans une séquence temporelle d'images représente le plus souvent la projection du mouvement 3D dans le plan image. Une image peut également représenter une coupe d'un objet 3D. Dans ce cas, le champ de mouvement observé à l'intérieur de cette coupe pourra représenter à la fois un déplacement dans le plan image et également la projection sur ce plan d'un mouvement 2D ou 3D à partir d'une séquence temporelle d'images 2D ou 3D dont le contenu varie en fonction du temps. On doit différencier le mouvement réel, le mouvement apparent sur l'image, et le mouvement estimé.

Champ de mouvement réel et champ de mouvement apparent :

Le mouvement (ou déplacement) réel anime la scène réelle, dans un espace tridimensionnel. Ce mouvement réel est observé par une prise de vue (caméra). Quant au mouvement observé ou apparent sous la forme d'une séquence d'images 2D ou 3D si l'on s'appuie seulement sur l'intensité des pixels. Dans ce cas on observe, en fait des changements de la distribution spatiale de l'intensité lumineuse. Le mouvement ainsi perçu est appelé champ de mouvement apparent ou flot optique qui est en général différent du champ réel de mouvement.

Le champ de déplacement apparent représente en général la projection (orthographie) du mouvement réel, il est appelé aussi mouvement projeté. Ce dernier représente une approximation du mouvement réel.

On suppose que le point P_t d'un objet réel 3D, à l'instant t , devient le point P_{t+1} à l'instant $t+1$. La projection perspective des points P_t et P_{t+1} est noté par p_t et p_{t+1} , respectivement avec $I_t(x,y)$ et $I_{t+1}(x,y)$ leurs intensités. Le mouvement apparent $\overrightarrow{p_t p_{t+1}}$, correspond à la projection perspective du mouvement réel.

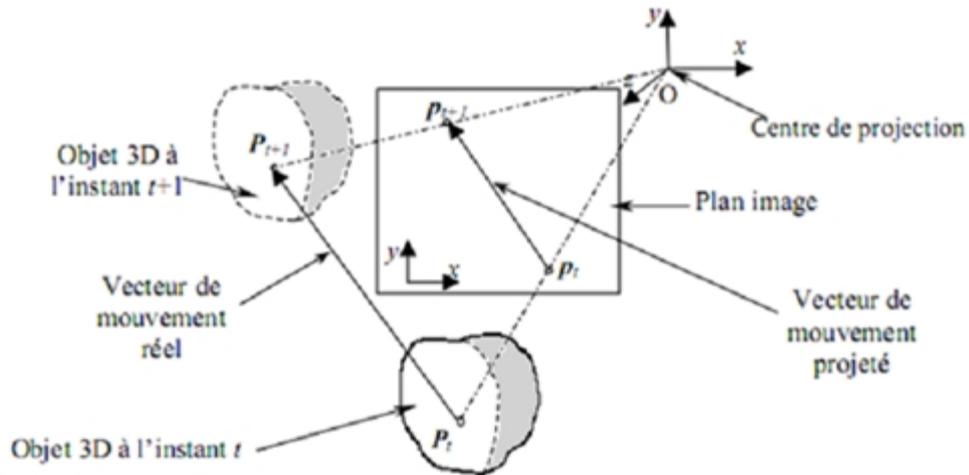


FIG I.4: Illustration des mouvements réels et apparents dans un système optique de vue

Dans ce cas, le mouvement projeté peut représenter la résultante du mouvement de plusieurs objets. On pourrait par exemple obtenir un mouvement nul, du fait de combinaisons inversées de deux ou plusieurs objets en mouvement. Dans le cas continu, le champ de déplacement apparent est défini pour tous les points $P(x, y, t)$ du plan image par un vecteur déplacement $d(x, y, t)$.

Champ de mouvement apparent et champ de mouvement estimé :

Le vecteur déplacement estimé $d(p) = (d_x(p), d_y(p))$ correspond au déplacement du point $p = (x, y)$ du plan image, est déterminé à partir du champ de mouvement apparent c'est-à-dire des variations locales d'intensité lumineuse $I(x, y, t)$ entre les instants t et $t + t$, ou t est la distance en temps inter-images, dans le cas continu. Le vecteur vitesse estimé $v = (v_x(p), v_y(p))$ est défini comme la variation temporelle du déplacement par unité de temps $(v_x, v_y) = (\frac{dx}{dt}, \frac{dy}{dt})$. Ceci explique pourquoi une séquence temporelle d'images ne permet pas d'estimer le mouvement apparent observable dans la séquence, et non le champ de vitesse réel. On nomme champ de déplacement, le champ de vecteurs déplacements estimés. La différence entre le champ estimé et le mouvement apparent existe si le gradient spatial d'intensité est trop faible. Pour que le mouvement réel soit observable, il faut que la variation du niveau de gris (respectivement de couleurs) soit suffisamment grande dans les régions où il y a un mouvement. L'exemple classique qui met en évidence l'importance de ce facteur, est celui d'une sphère ayant une distribution uniforme d'intensité (éclairage homogène), qui effectue un mouvement de rotation autour de son axe propre, dans une scène à éclairage constant. Même si en réalité il y a mouvement, celui-ci ne peut être observé à partir d'une séquence.

De plus si l'illumination de la scène varie. Un mouvement observable dans une séquence d'images d'intensité ne correspond pas toujours à un mouvement réel. La même sphère immobile soumise à une illumination variable au cours de la séquence génère un mouvement apparent (c'est-à-dire une variation d'intensité) artificiel.

1.3.3 : Estimation de mouvement

Pour estimer le mouvement à partir du mouvement apparent, il faut admettre que l'intensité reste inchangée ou elle varie d'une façon prédictible. Cette hypothèse dite de conservation s'exprime par l'équation DFD (Displaced Frame Difference) des différences entre les images déplacées autrement dit les images prises entre les instants t et $t+dt$.

$$DFD = I(x+dx, y+dy, t+dt) - I(x, y, t) = 0$$

L'estimation du mouvement réel à partir du mouvement apparent peut être abordée de deux manières différentes :

- L'estimation des vecteurs déplacements dans le plan image $d(x, y, t) = (dx, dy)$ estimés entre les images t et $t + dt$
- L'estimation du vecteur vitesse $v(x, y, t) = (v_x, v_y)$. Les vecteurs déplacements estimés peuvent varier en espace et en temps.

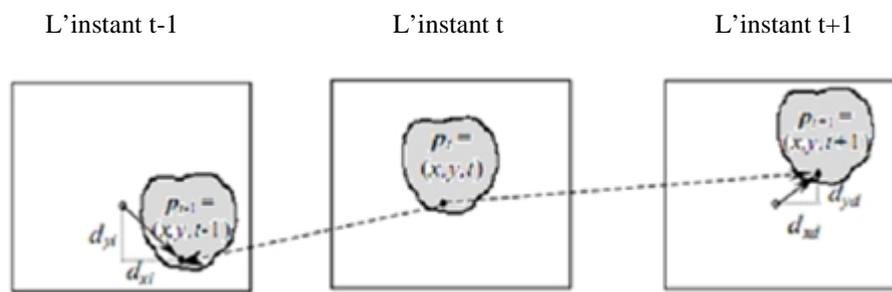


FIG I.5 : Estimation directe et inverse du vecteur déplacement

Pour la figure précédente :

- Dans le cas de l'estimation directe on a noté: $d_d(x, y, t) = (d_{xd}(x, y, t), d_{yd}(x, y, t))$.
- Dans le cas de l'estimation inverse on a noté: $d_i(x, y, t) = (d_{xi}(x, y, t), d_{yi}(x, y, t))$.

Les échantillons $I(x, y, t)$ permettent de déterminer les vecteurs vitesses. On observe que si la vitesse reste constante dans l'intervalle dt entre deux images et si dt est petit, alors la vitesse estimée peut être assimilée au déplacement : $v = \frac{d(x,y,t)}{dt}$

Pour un mouvement accéléré, on doit élargir nos choix en prenant plus de deux images afin de mieux estimer le champ de vitesse.

En conclusion, le champ de mouvement estimé peut être caractérisé par le champ de vecteurs vitesse ou par le champ de vecteurs déplacement. L'estimation du mouvement représente donc deux approches équivalentes. L'estimation du mouvement est très sensible au bruit présent dans les images qui peut être interprété comme étant le résultat d'un mouvement dans la scène réelle.

Le calcul du mouvement :

C'est l'extraction des caractéristiques physiques de l'environnement à partir d'une information fournie par l'image qui ne permet de les recouvrir que partiellement. Par exemple, on déduit la vitesse réelle d'un objet qui se déplace dans un environnement tridimensionnel or l'information recueillie de l'image ne nous donne que deux paramètres la vitesse horizontale v_x et la vitesse verticale v_y (c'est-à-dire les deux composantes du flot optique), ils sont appelés aussi les vitesses apparente.

Pour vaincre cette difficulté, il nous faut des informations de plus, la distance de l'objet et la distance focale de la caméra pour trouver les vraies coordonnées de l'objet avec seulement les deux coordonnées apparentes sur notre écran. On envisage plusieurs méthodes de calcul de mouvement, chacune de ces méthodes s'appuie sur un algorithme qui lui est particulier, en général le temps réel est exigé même pour la compression dans le cas d'une transmission indirecte. Le temps de calcul est donc une préoccupation primordiale dans le traitement des séquences d'images. La figure ci-dessous représente une méthode d'estimation de mouvement, en se basant sur l'évolution de l'objet en question sur un intervalle de temps entre t et $t + \Delta t$.

La matrice résultante nous nous révèle la région ayant subi un changement et ceci s'effectue en comparant les éléments de la matrice de différence avec l'image de référence.

$$\begin{cases} x = f \frac{X}{Z} \\ y = f \frac{Y}{Z} \end{cases} \text{ X, Y, et Z sont les coordonnées d'un point de la scène.}$$

x et y sont les coordonnées du même point projeté dans le plan image et f la distance focale de la caméra.

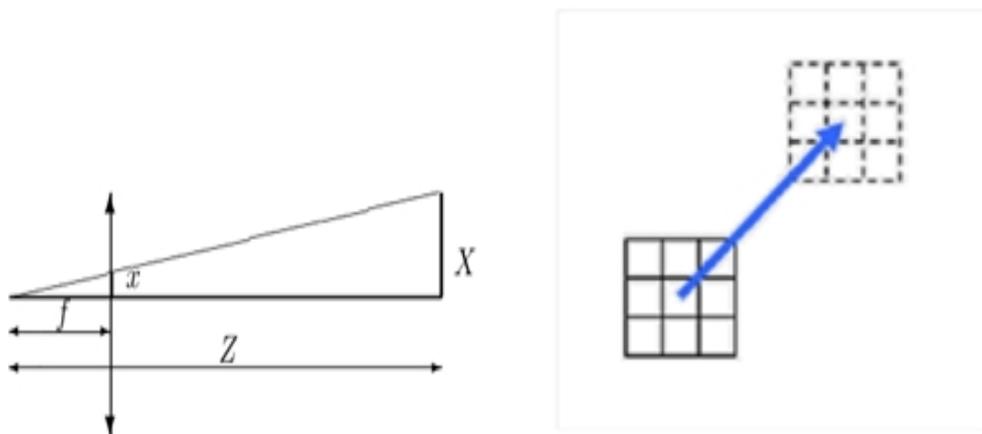


FIG I.6 : projection de l'objet sur le plan image

Le champ des vitesses réelles :

Admettant qu'on échantillonne 25 images/sec ($f_c = 50\text{Hz}$ pour les téléviseurs de norme PAL) donc la différence temporelle entre deux images qui se succèdent est de $t = \frac{1}{25} \text{ s} = 40\text{ms}$.

Ensuite on fait la différence X et Y et en divisant par t on aura la vitesse instantanée suivant l'axe des x et l'axe des y dans l'intervalle de temps $[t, t+40\text{ms}]$. Une fois qu'on a établi l'expression du flot optique on peut déduire les vitesses de nos objets si on s'intéresse à la surveillance du territoire ou bien la poursuite des cibles. $V_x = \frac{f}{Z} v_x$ et $V_y = \frac{f}{Z} v_y$, avec V_x et V_y sont les vitesses réelle de l'objet suivant l'axe des x et des y , mais on est toujours confronté à un problème qu'on ne peut pas déterminer la vitesse normale au plan de l'image c'est-à-dire V_z .

Le calcul du flot optique :

Le problème qui se pose est la mesure du mouvement dans une séquence d'images. En général, l'évolution de l'image au cours du temps est due principalement à deux facteurs :

- des sauts entre deux séquences successives, qui sont rares et ponctuels
- le déplacement relatif des objets filmés et de la camera.

Le mouvement relatif des objets et de la camera est un champ de vecteurs, a trois composantes, des vitesses des objets filmés dans le référentiel de la camera. Ce champ de vecteur correspond au mouvement réel. La scène étant projetée sur le plan du film de la camera, on définit sur le plan film (ou plan focal de l'image) de la camera un deuxième champ de vitesse qui est le champ de vitesses projeté. On note p l'opérateur de projection (qui peut être linéaire ou non). Pour chaque point a de l'image, qui est le projeté $p(a)$ d'un point réel A de vitesse V , le flot optique en x est alors le vecteur $v = \frac{dp(x)}{dt}$. L'objet de la mesure du flot optique est de l'estimer sur la base d'une séquence d'images filmées $I(t, a(x_a, y_a))$.

La mesure du flot optique a un certain nombre d'applications possibles. Elle peut servir en tant que telle pour faire de la compression de séquences d'images vidéo par compensation de mouvement (prédiction d'images sur la base d'un champ de déplacement). La mesure du flot optique sert également à l'analyse des scènes: le mouvement apparent des objets d'une scène peut permettre de reconstruire une scène tridimensionnelle si on dispose d'informations supplémentaires sur la nature du mouvement réel. Ces techniques servent donc pour la construction de modèles tridimensionnels d'objets réels (acquisition tridimensionnelle) pour la réalité virtuelle, ou encore en robotique, pour construire une représentation de l'environnement d'un robot en déplacement.

Le calcul du flot optique consiste à extraire un champ de vitesses à partir d'une séquence d'images en supposant que l'intensité (ou la couleur) est conservée au cours du déplacement. Sous cette hypothèse, on peut établir une relation entre la vitesse apparente v (déplacement dans l'image d'indices visuels tels que des régions délimitées par des contours

supposé représenter la projection du mouvement 3D des objets de la scène et/ou du mouvement de la caméra) et les variations spatio-temporelles de l'intensité.

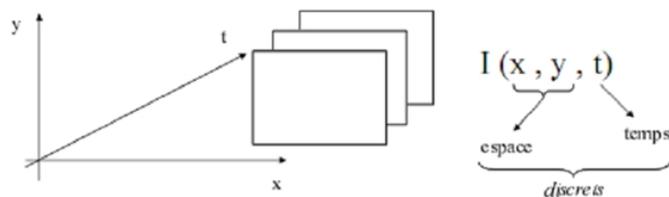


FIG I.7 : Les variations spatio-temporelles

Soit $I(x, y, t)$ la fonction de l'intensité lumineuse au pixel $p(x, y)$ dans l'image prise au temps t . Si la luminance d'un point sur l'image ne varie pas de manière significative entre les temps t et $t+dt$ et si pendant cet intervalle de temps le point $p(x, y)$ se déplace de $d(dx, dy)$, alors en développant en série de Taylor :

$$I(x + dx, y + dy, t + dt) = I(x, y, t) + \frac{\partial I}{\partial x} dx + \frac{\partial I}{\partial y} dy + \frac{\partial I}{\partial t} dt + \dots$$

Et puis qu'on a mis l'hypothèse de conservation d'intensité donc :

$$I(x + dx, y + dy, t + dt) = I(x, y, t)$$

Ce qui fait qu'après qu'on néglige les termes d'ordre supérieur :

$$\frac{\partial I}{\partial x} dx + \frac{\partial I}{\partial y} dy + \frac{\partial I}{\partial t} dt = 0$$

Dérivant maintenant par rapport au temps :

$$\frac{\partial I}{\partial x} \frac{dx}{dt} + \frac{\partial I}{\partial y} \frac{dy}{dt} + \frac{\partial I}{\partial t} = 0 \quad \rightarrow \quad \frac{\partial I}{\partial x} v_x + \frac{\partial I}{\partial y} v_y + \frac{\partial I}{\partial t} = 0$$

Le gradient spatial est donné par :

$$\nabla I = \left(\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y} \right)$$

Le flot optique est donné par: $v = (v_x, v_y)$ puisque il traduit les vitesses du mouvement dans l'image. Donc l'équation précédente se résume:

$$\nabla(I) \cdot v + \frac{\partial I}{\partial t} = 0$$

• : Est le produit scalaire, Cette équation s'appelle l'équation de contraintes du flot optique. [6]

1.4 : Conclusion

L'analyse du mouvement dans une séquence d'image repose sur la détection et l'estimation du mouvement.

La détection de mouvement a pour but de décider quelles parties des images (pixels ou régions) correspondent à des objets mobiles. Il s'agit d'une détection binaire, et L'estimation inclut la précédente en y ajoutant la contrainte que le résultat fourni doit être quantitatif.

Chapitre II

Suivi d'objets en mouvement

II.1 : Introduction

Le suivi visuel d'objets dans des séquences d'images est devenu une des tâches les plus utilisées, parmi celles développées pour la vision par ordinateur. Les différents champs d'application comprennent entre autres: la vidéoconférence (suivi des interlocuteurs), la vidéosurveillance (détection, suivi, reconnaissance du comportement de personnes, d'intrus), la segmentation et l'édition de séquences vidéo, sans oublier les applications plus traditionnelles de la robotique : suivi d'obstacles pendant une phase d'évitement, suivi d'indices visuels dans une tâche asservie sur la vision, suivi d'un opérateur (corps, visage, main...) pour définir des modes d'interaction évolués entre l'Homme et la Machine...[7]

Le but central du suivi visuel consiste à déterminer la configuration d'une cible ou les caractéristiques d'un objet dans une séquence d'images, tant que cette cible est en mouvement apparent dans les images; ce mouvement dans les images peut provenir d'un réel mouvement 3D de la cible dans le champ de vue de la caméra (scénario de type vidéosurveillance), du mouvement de la caméra si celle-ci est montée sur un actionneur: robot, platine site / azimut... (scénario du robot qui suit des amers fixes durant ses déplacements) ou de la combinaison de ces deux mouvements (scénario du véhicule qui suit une cible mobile, ou qui évite un obstacle dynamique). [7]

En se basant sur la nature du modèle géométrique utilisé, il est possible de distinguer quatre classes parmi les méthodes de suivi :

- **Suivi basé sur des modèles spécifiques :**

Cette approche est dépendante du champ d'application. En particulier, dans le domaine de l'analyse de séquences télévisuelles, citons par exemple le suivi de visages ou de personnes, la reconnaissance d'objets caractéristiques. Dans d'autres domaines, on trouve des approches par modèles spécifiques en CAO, ou en imagerie médicale (suivi d'organes).

De tels modèles spécifiques sont certainement plus robustes et/ou efficaces dans les cas particuliers prévus explicitement. Par contre ils ne fonctionnent que dans ces cas-là, ce qui oblige à manipuler plusieurs modèles distincts dès que l'on désire effectuer le suivi d'objets de natures différentes, et à préciser pour chaque suivi quel modèle spécifique utiliser.

A l'opposé, les trois approches suivantes font moins d'hypothèses sur la nature des objets à suivre, tentant ainsi d'être utilisable de façon moins spécifique.

- **Suivi de primitives ponctuelles**

Des primitives ponctuelles sont détectées pour chaque image, et un procédé de mise en correspondance les relie d'une image à l'autre. Généralement, les primitives présentent une caractéristique discriminante, et invariante au court du temps (par exemple des coins ou des points à fort contraste), ce qui assure qu'elles correspondent à un point réel des objets suivis, et que si elles sont détectées sur une image, elle le sont aussi sur les autres. Il est donc possible pour chaque primitive ainsi mise en évidence de calculer sa trajectoire au fil de la séquence.

Il apparaît que ce modèle ne définit pas l'étendue d'un objet, mais seulement la position de certains points le composant. Pour représenter cette étendue, il existe deux approches :

Les contours, ou les régions, qui se différencient principalement par les critères utilisés pour estimer le suivi.

- **Suivi de contours :**

La principale hypothèse du suivi de contours (aussi appelé contour actif) est que le bord de l'objet suivi présente des caractéristiques particulières (contraste de texture et/ou discontinuité de mouvement), ce qui sert à recalculer la position du contour d'une image à la suivante en faisant converger celui-ci vers les points ayant ces caractéristiques.

Ce suivi considère donc uniquement l'information présente sur les contours des objets. Ceci présente le risque que le calcul du déplacement de l'objet soit perturbé par l'apparence du fond. Ce point de robustesse est important si l'on désire traiter des vidéos d'origine télévisuelle pour lesquelles les différents objets et le fond peuvent avoir une texture perturbée.

- **Suivi de régions :**

Une ou plusieurs régions sont données sur une image. Le suivi doit pouvoir déterminer leurs positions dans les images suivantes. Une caractéristique générale du suivi de régions consiste en des hypothèses d'homogénéité du mouvement à l'intérieur d'une région, et d'invariance de l'apparence de celle-ci dans le temps. En cela, il se distingue fondamentalement du suivi de contours, qui se base à l'inverse sur les discontinuités (de texture et de mouvement) présentes dans l'image.

Pour représenter des mouvements non rigides, il est nécessaire d'adopter un modèle plus riche, capable de déformations locales. Le modèle de mouvement par maillages répond à ce critère. Son utilisation évite un morcellement en petites régions (approche par partition) ou une approximation grossière (modèle affine global) lorsque des objets déformables sont suivis. Contrairement aux partitions, il ne permet pas de prendre en compte explicitement les discontinuités de mouvement, ce qui peut être en partie évité si le maillage ne recouvre qu'un seul objet.[15]

II.2 : Méthode de suivi par contours actifs

Les contours actifs ou « Snakes » ont été intensivement utilisés pour la segmentation et le suivi d'objets pour leur capacité à intégrer les processus de détection et de chaînage des contours en un seul processus de minimisation d'énergie. Cette méthode est basée sur la minimisation d'énergie le long d'une courbe quand elle est déformée par des forces internes et externes. Ces forces sont définies afin d'imposer respectivement, des propriétés pour la forme du contour et pour sa position dans l'image [7][10].

Le modèle de contours actifs se présente sous la forme d'une courbe fermée ou ouverte dont l'initialisation est située à proximité du contour qu'on veut obtenir et dont l'évolution s'effectue selon un processus itératif de déformation contrôlé par un test de convergence.

Les travaux existants en segmentation par contour actif peuvent être répartis en deux catégories. La première correspond aux approches contours et la seconde aux approches régions.

II.2.1 : Contour actif par approche contour

Historiquement, ils ont été introduits par Kass, Witkin et Terzopoulos en 1988 et font figure de référence en la matière. Le mot « snake » (serpent en français) se justifie par la déformation du contour lors de sa convergence. Le critère selon lequel le contour se déforme en faisant évoluer le contour initial est associé à la recherche d'une minimisation d'une fonctionnelle interprétable en terme de contraintes physiques: élasticité, raideur. La convergence du contour vers la frontière recherchée est considérée comme une condition de stabilité. La méthode de détection des contours actifs nécessite l'initialisation d'une ligne de contour en entourant la zone des objets à détecter et en faisant évoluer cette courbe de façon à minimiser son énergie E.[10]

Puisque le « snake » peut former un graphe composé, nous pouvons le représenter sous forme paramétrique par : $V : [0,1] \rightarrow \mathbb{R}^2$

Le contour actif peut être décrit par une courbe C, fonction du temps t et de l'abscisse curviligne s par la formule suivante :

$$C = \{v(s, t) = (x(s, t), y(s, t)) / s \in [a,b] \text{ et } t \in [0,T] \}$$

Où a et b représentent les extrémités du snake. La figure suivante représente la courbe C :



FIG II.1: La courbe C

Donc, le contour actif est formé d'une série de points mobiles et répartis sur une courbe en deux dimensions. La courbe est placée dans la zone d'intérêt de l'image ou autour d'un objet. [11]

La segmentation est réalisée à travers un processus de minimisation d'une énergie notée E exprimée par:

$$E(C) = E_{\text{int}}(C) + E_{\text{ext}}(C) + E_{\text{img}}(C)$$

E_{int} : Une énergie propre due uniquement à la forme du contour, dite énergie interne.

E_{ext} : C'est une énergie qui exprime des contraintes supplémentaires qui peuvent être imposées par l'utilisateur pour obtenir le contour qu'il veut.

E_{img} : Une énergie potentielle imposée par l'image. C'est elle qui va attirer la ligne du contour vers les contours réels présents dans l'image.

- L'énergie interne:

L'énergie interne sert à maintenir une certaine topologie cohérente du contour, en empêchant des nœuds individuels sur le contour de se balader trop loin de leurs nœuds voisins. Le but est de limiter l'influence des effets externes sur la déformation du contour.

Cette énergie est dite interne puisqu'elle est caractéristique des propriétés physiques du contour.

$$v(s) = \alpha(s)|v_s(s)|^2 + \beta(s)|v_{ss}(s)|^2$$

$$|v_s(s)|^2 = \left(\frac{dx(s)}{ds}\right)^2 + \left(\frac{dy(s)}{ds}\right)^2$$

$V_s(s)$: la dérivée première de $v(s)$ par rapport à s .

$V_{ss}(s)$: la dérivée seconde de $v(s)$ par rapport à s .

$\alpha(s)$: l'élasticité du contour

$\beta(s)$: la rigidité du contour

Le terme de premier ordre est contrôlé par $\alpha(s)$ qui représente l'élasticité de contour. Le terme du second ordre est contrôlé par $\beta(s)$ qui représente la rigidité du contour. Le choix des paramètres α et β impose les caractéristiques de déformation du contour. Par exemple dans le cas d'un contour où il n'y a pas de forces externes avec $\alpha > 0$ le contour tente de former un cercle et il tend vers un point (cercle de rayon nul) avec le temps; en plus, une valeur positive de $\beta > 0$ empêche le contour d'avoir des discontinuités locales, il ne peut pas former des coins aigus.

- L'énergie externe

Cette énergie sert à introduire des informations de haut niveau. Ceci se traduit par la présence de forces de répulsions en certains points de la courbe ou par la présence de forces d'élasticité simulant l'allongement d'un ressort entre deux points de la courbe. Les informations de haut niveau peuvent être associées à la présence de frontières préalablement détectées (contraintes) ou de points de contrôle (points de passage obligés).

- L'énergie image

L'énergie de l'image E_{img} (c) fait intervenir les caractéristiques d'image que l'on cherche à mettre en valeur. L'énergie de l'image caractérise donc les éléments vers lesquels on veut attirer le snake sur l'image, il s'agit de détecter des contours, on s'intéresse donc aux lignes de fort gradient:

$$E_{img} = -gradient(i) = - \int_a^b |\nabla^* I(v(s))|^2 ds$$

Ou $\nabla^* I(v(s))$ représente le gradient de l'image au voisinage de la courbe $v(s)$. Un minimum d'énergie sera atteint si la courbe passe par les points de gradient maximal de l'image lisse par un filtre gaussien. Le signe négatif indique que les gradients les plus grands minimiseront cette énergie et attireront le « snake ». [10]

La figure suivante schématise un exemple de processus d'un suivi d'objet en mouvement par contour actif [11]:

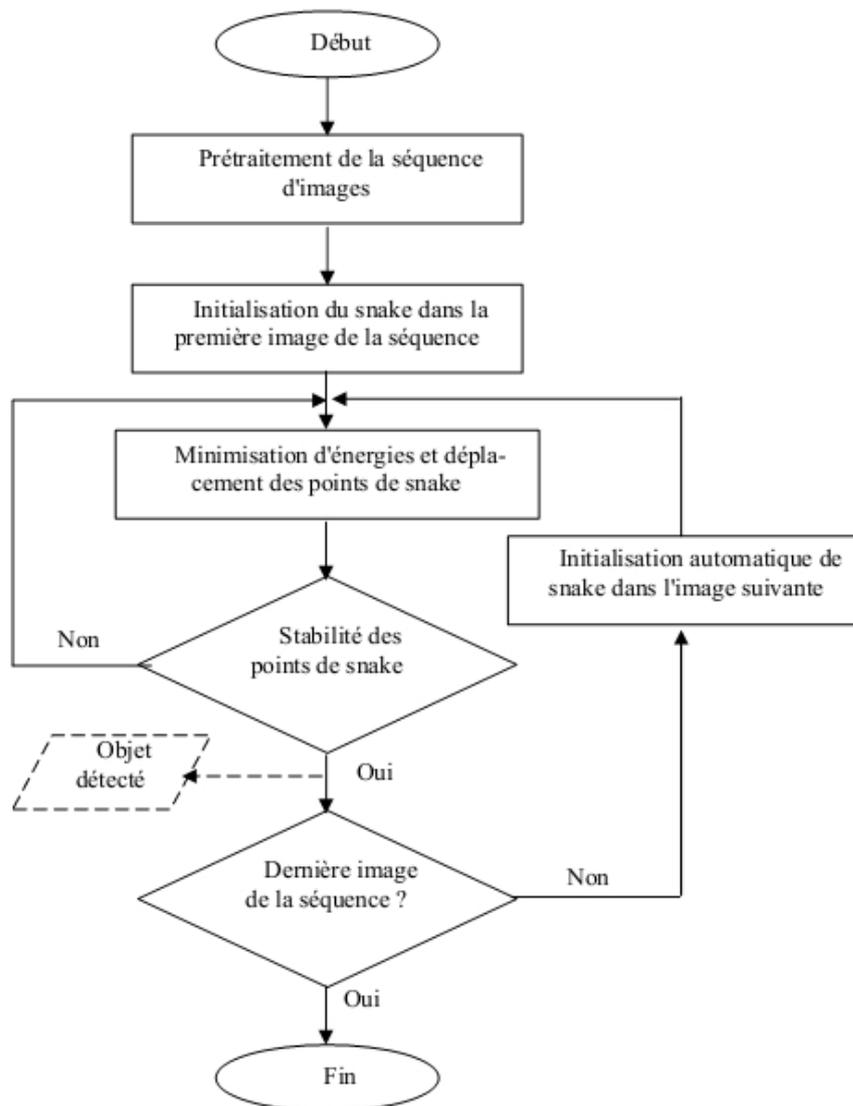


FIG II.2 : Organigramme d'un suivi d'objets mobiles par contour actif

Déformation du contour actif :

Pour déformer le contour actif, on pose $v(x, y)$ comme étant la coordonnée généralisée pour la réduction des équations du mouvement donné par l'équation de Lagrange. D'où la minimisation de la fonction J suivante :

$$J = \int_{\mathcal{T}} L(v(s, t)) ds$$

Où $L(v(s, t))$ est le Lagrangien du contour, défini par :

$$L(v) = K(v) - U(v)$$

$K(v)$: Energie cinétique du contour due à son mouvement.

$U(v)$: Energie potentielle du contour due à sa position

L'équation de l'énergie cinétique :

$$K(v) = \frac{1}{2} \int_{\Omega} \mu(s) |v_t|^2 ds$$

$\mu(s)$: Densité linéique

$v_t = \frac{\partial v}{\partial t}$ Dérivée partielle de $v(s,t)$ par rapport à t .

L'énergie potentielle dépend de l'énergie interne du modèle physique (élasticité et rigidité) et de l'énergie potentielle externe associée à l'image:

$$U(v) = E_{snake} = \int_{\Omega} E(v(s)) ds = \int_{\Omega} [E_{interne}(v(s)) + E_{externe}(v(s))] ds$$

La solution du problème de minimisation de l'énergie donnée par l'expression de l'équation de la fonction J se résume à trouver un chemin qui possède un état stationnaire donc le minimum doit toujours satisfaire les équations d'Euler Lagrange :

$$\frac{d}{dt} \left(\frac{\partial L}{\partial v_t} \right) - \frac{\partial L}{\partial v} = 0$$

$$\alpha(s) = \frac{1}{\gamma} w_1(s)$$

En remplaçant les expressions de K et U le Lagrangien s'exprimera comme suit:

$$L = \frac{1}{2} \int_{\Omega} [\mu(s) |v_t|^2 - w_1(s) |v_s(s)|^2 - E_{externe}(v(s))] ds$$

Les équations d'Euler-Lagrange relatives au mouvement de contour deviennent:

$$\frac{\partial}{\partial x} (L) - \frac{\partial}{\partial t} \left(\frac{\partial}{\partial x_t} (L) \right) + \frac{\partial}{\partial s} \left(\frac{\partial}{\partial x_s} (L) \right) + \frac{\partial^2}{\partial^2 s} \left(\frac{\partial}{\partial x_{ss}} (L) \right) = - \frac{\partial}{\partial x_t} \left(\frac{1}{2} \int_{\Omega} \gamma(s) |v_t|^2 ds \right)$$

(s) : Facteur de viscosité.

On remplace la valeur de Lagrangien on obtient :

$$\mu(s) x_{tt} + \gamma(s) x_t - \frac{\partial}{\partial s} (w_1(s) x_s) + \frac{\partial^2}{\partial^2 (s)} (w_2(s) x_{ss}) = - \frac{1}{2} \frac{\partial}{\partial x} (E_{externe}(v))$$

On a la même équation pour y tout en remplaçant x par y .

Cohen a mis la densité linéique $\mu(s)$ égale à zéro, (s) est une constante et:

$$\beta(s) = \frac{1}{\gamma} w_2(s)$$

$$E_{externe} = \frac{1}{2\gamma} E_{externe}$$

On obtient donc l'équation suivante :

$$\frac{\partial v(s, t)}{\partial t} - \frac{\partial}{\partial s} \left(\alpha(s) \frac{\partial v(s, t)}{\partial s} \right) + \frac{\partial^2}{\partial s^2} \left(\beta(s) \frac{\partial^2 v(s, t)}{\partial s^2} \right) + \nabla E_{externe}(v(s, t)) = 0$$

On ajoute les conditions aux limites et les conditions initiales.

L'évolution du contour est basée sur le déplacement du contour de façon itérative de telle sorte que les forces internes et les forces externes soient équilibrées.

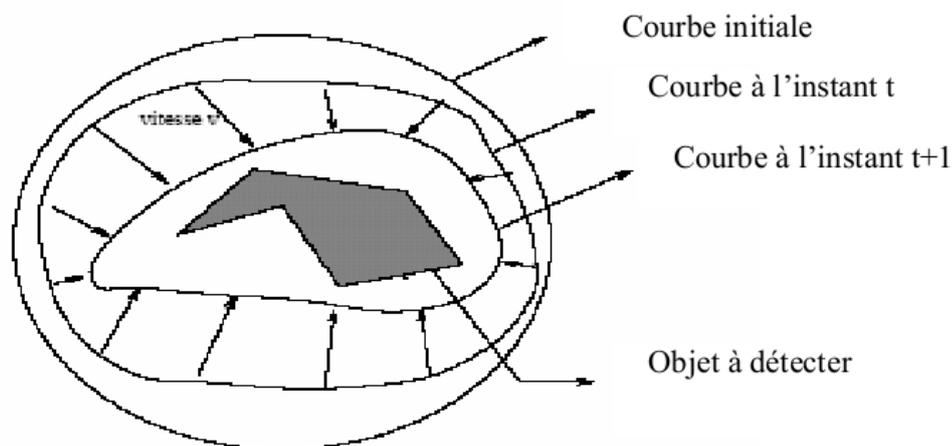


FIG II.3 : Représentation de l'évolution du contour actif

L'état final est obtenu lorsque $\frac{\partial v(s,t)}{\partial t}$ s'approche de zéro. Plusieurs méthodes ont été utilisées pour la résolution de cette équation. M. Berger a mis en évidence la difficulté du choix des divers paramètres, ainsi que l'influence prépondérante de l'initialisation sur la convergence du contour actif. F. Leymarie et M.D. Levine ont réalisés une étude sur le choix des paramètres, la résolution numérique et la stabilité des solutions dans une application précise qui est la poursuite d'objet évoluant dans le plan (cellule vivante évoluant sur une place).

II.2.2 : contour actif par approche région

Le principe des contours actifs basés régions est de prendre en compte des propriétés de l'image sur des régions et non plus localement. La fonctionnelle à minimiser dans le cas de termes régions s'écrit comme une intégrale sur la région et non plus sur le contour.

Elle est de la forme :

$$J(\Omega) = \int_{\Omega} k(x, \Omega) dx$$

Où k est appelé descripteur de région. C'est une fonction qui dépend des propriétés de la région. Il est important de noter que souvent les approches région combinent à la fois des termes basés régions et des termes basés contours, principalement pour prendre en compte les caractéristiques des régions et ajouter des contraintes sur le contour.

Les approches basées régions peuvent reposer sur une modélisation statistique des régions à segmenter. Un descripteur de moyenne ou bien de variance sur la région est, par exemple, parfois utilisé. [12]

Les premiers à avoir utilisés des descripteurs de région sont Cohen et al, et Ronfard. Cohen et al ont présentés une méthode de reconstruction de surface en utilisant des contours actifs. Ils incluent un critère contenant des termes basés régions. Ronfard a proposé, pour la première fois, que la force d'évolution appliquée au contour actif devait être proportionnelle à la différence des caractéristiques de la région des objets et de celles de la région du fond. [12][13]

Un exemple des travaux réalisés par Ronfard consiste à appliquer à chaque point du contour dont le voisinage appartient au fond une force dirigée vers l'intérieur du contour. Inversement, chaque point du contour possédant un voisinage appartenant à l'objet sera soumis à une force en direction de l'extérieur. Toutes ces forces sont définies avec une direction normale au contour comme indiqué dans la figure suivante :

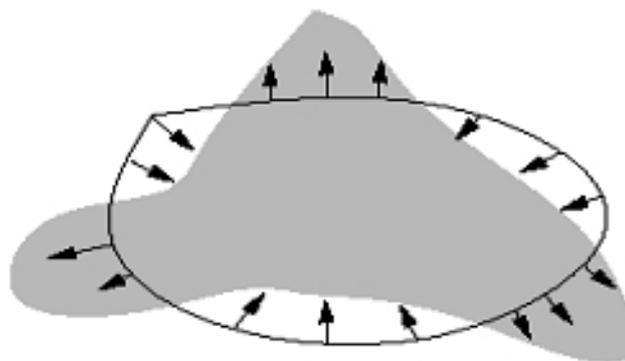


FIG II.4: Forces régissant l'évolution du contour actif

Ronfard a associé à un contour une énergie $W^{contour}$. L'évolution du contour actif dépend de cette énergie.

La région peut être définie comme l'ensemble des pixels voisins décrit par une même texture. Dans une image on peut avoir plusieurs régions.

L'intensité des pixels d'une région k est approchée par une combinaison linéaire des fonctions $K_i(x,y)$. L'énergie de la région k est définie comme suit:

$$W_k^{region} = \iint \|I(x,y) - m_k\|^2 dx dy$$

$I(x,y)$: l'intensité des pixels.

m_k : Une intensité moyenne.

Cette définition de l'énergie permet de définir celle associée au contour $W_k^{contour}$. Elle est choisie comme étant la Ward-distance entre les deux régions R_{in} et R_{out} , elle permet de mesurer le degré de similitude entre deux régions.

$$W^{contour}(C) = D[R_{in}, R_{out}] = W^{region}(R_{in} \cup R_{out}) - W^{region}(R_{in}) - W^{region}(R_{out})$$

Si les deux régions sont identiques alors cette mesure est minimale.

Le déplacement d'un point définit un nouveau contour $C + \delta C$. Ce contour a une énergie: $W^{contour}(C + \delta C)$. La variation d'énergie entre les deux contours permet de définir la force qui sera appliquée aux nœuds.

II.2.3 : application du modèle de contour actif pour le suivi de joueurs de football

Nous choisirons comme application le suivi de joueurs de football au cours d'un match. Dans une telle scène, il ne s'agira pas, précisons-le, d'analyser les mouvements techniques du joueur, mais seulement de suivre sa position.

Une des méthodes développées utilise un modèle de contour actif mais ne procède à aucun filtrage ou segmentation précédant la déformation du contour, ce qui permet de diminuer le temps de calcul. L'hétérogénéité de la texture des zones de l'image correspondant aux joueurs, associée à la relative homogénéité de la pelouse, conduit à faire le choix d'un modèle utilisant une énergie ballon de réduction. Tenant compte à la fois des propriétés simultanées des objets suivis (taille, texture complète) et du fond considéré comme de couleur uniforme, la méthode présentée se montre alors plus efficace.

Afin de réduire le temps de calcul, aucun prétraitement n'est nécessaire à la déformation du contour actif et aucune estimation de mouvement n'est effectuée pour le suivi. La méthode se décompose en deux parties, exécutées successivement pour chaque image : l'initialisation sous forme de rectangle puis la déformation du contour actif.

- Initialisation sous forme de rectangle :

Pour une image donnée, l'initialisation par un rectangle est effectuée en deux étapes. Tout d'abord un rectangle parallèle aux contours de l'image est créé. Sa position et sa taille sont déterminées de telle sorte qu'il englobe le contour actif final obtenu à l'image précédente. La taille de ce rectangle est ensuite augmentée afin d'englober le contour actif final supposé pour l'image courante. Le contour actif initial est obtenu en plaçant les points qui le composent de manière uniforme sur le contour du rectangle.

Un traitement spécifique est nécessaire pour la première image d'une séquence, ou a priori aucune information sur le contour n'est disponible. Actuellement, l'initialisation sous forme de rectangle est remplacée par une initialisation manuelle du contour effectuée par l'utilisateur ou une initialisation automatique obtenue par une détection des objets en mouvement.

- Déformation du contour actif :

Le contour actif, dont le nombre de points dépend de la précision voulue, est ensuite déformé et se fixe autour de l'objet. Il est réduit jusqu'à convergence avec les forces décrites précédemment. Nous avons choisi de réduire le « snake » plutôt que de l'accroître car il est plus efficace de réduire le contour en traversant des zones homogènes (comme la pelouse) que de l'accroître en traversant des zones hétérogènes (comme les joueurs) riches en information de contour et de couleur. [8]

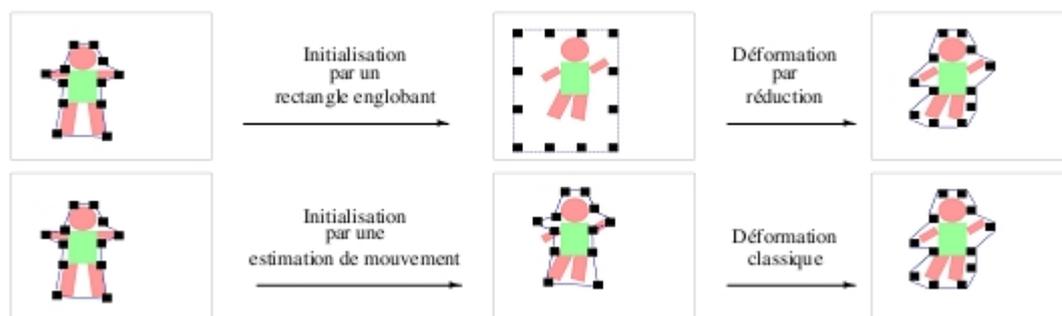


FIG II.5: Processus de suivi par contour actif

La figure suivante illustre que le processus de suivi déboute toujours par le résultat précédent (image de gauche) pour obtenir le résultat courant (image droite). Cependant, cette contribution ne nécessite pas d'estimation de mouvement (ligne de haut) contrairement aux autres approches par contour actif (ligne du bas) [14]

Expérimentations :

La méthode a été testée sur des séquences vidéo couleur issues de retransmissions télévisées de matchs de football. Ces séquences ont été acquises à une fréquence de 15 images par seconde. Les images couleur sont codées en 24 bits (RGB) et leur taille est de 384×288 pixels.

Le contour actif est initialement composé de 16 points. Le nombre de points peut varier au cours de la séquence selon la précision souhaitée. L'étape de déformation est limitée à un maximum de 30 itérations. Cependant, le contour actif converge la plupart du temps avant ce nombre d'itérations. La convergence n'a pas été obtenue pour 30 % des images. A chaque itération, l'énergie est calculée dans un voisinage de 5×5 pixels autour du point de contour obtenu lors de l'itération précédente. Le gradient est calculé à l'aide de l'opérateur de Sobel et la valeur utilisée pour le seuillage du gradient a été fixée à 500.

La figure suivante montre le résultat de la déformation du contour actif sur les deux premières images d'une séquence :



FIG II.6: Contours initiaux et finaux obtenus pour les deux premières images d'une séquence

Le contour initial est initialisé manuellement sur la première image (un utilisateur place les points du contour sur l'image) tandis qu'il est initialisé automatiquement sur la seconde image en se basant sur la position finale du contour actif obtenue dans la première image.

Les résultats présents sur la figure suivante montrent le suivi d'un joueur de football en mouvement sur plusieurs images extraites d'une séquence de 100 images correspondant à une durée de 7 secondes :

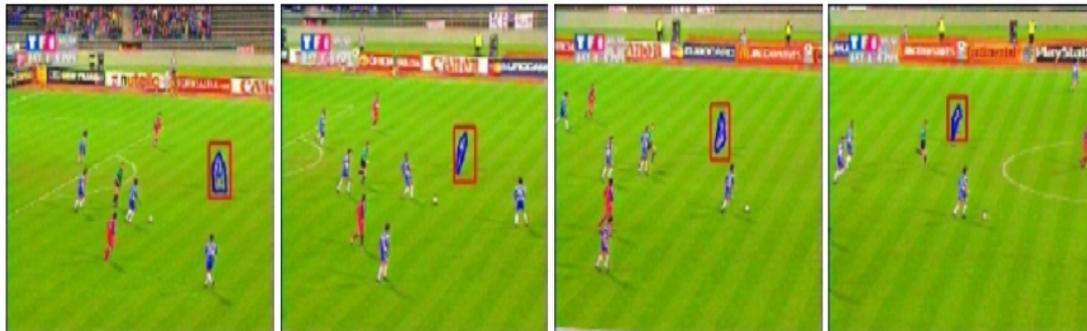


FIG II.7: Contours initiaux et finaux obtenus pour les 20°, 40°, 60°, et 80° images de la séquence.

Les contours initiaux sont déterminés automatiquement à l'aide des contours finaux obtenus pour les images précédentes. L'algorithme de suivi n'est perturbé ni par le mouvement du joueur, ni par le mouvement de la camera.

Discussion :

Avantage :

La méthode présentée est capable de suivre des objets non rigides tels que les joueurs de football. Elle utilise une implémentation rapide du modèle de contour actif, sans filtrage global ni estimation de mouvement. Elle analyse des séquences d'images couleur et est robuste au mouvement de la camera.

Insuffisance :

La limite principale de la méthode proposée est liée à l'environnement de l'objet suivi, et plus particulièrement aux autres objets en mouvement présents dans la séquence ainsi qu'au fond de la scène. En fait, la méthode est capable de suivre un joueur tant qu'il n'est pas trop proche des autres joueurs présents dans la scène. Si l'objet suivi est proche d'un autre objet, le contour actif englobe les deux objets et la méthode n'est pas capable de séparer les deux objets lorsque ceux-ci s'éloignent l'un de l'autre. Le fond de la scène est lui aussi source de problèmes. Dans le cas d'images de match de football, le fond contient fréquemment des lignes blanches pour représenter les limites du terrain. Les pixels représentant ces lignes sont caractérisés par de fortes valeurs de gradient. Le contour actif se fixe alors sur ces lignes blanches. Ces deux problèmes sont illustrés par la figure suivante :

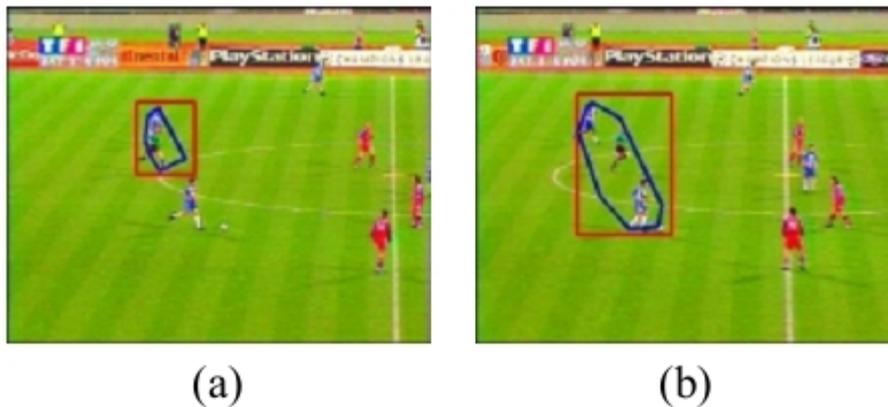


FIG II.8: Contours initiaux et finaux montrant: (a) la sensibilité au fond de la scène (particulièrement les lignes blanches), et (b) l'impossibilité du contour actif à se séparer afin de suivre plusieurs objets. [8]

II.3 : méthode de suivi par maillage

Le concept de représentation par maillage est relativement récent en codage d'image fixe ou vidéo, alors que ce type de représentation est exploité depuis bien longtemps pour la modélisation et l'analyse de structure par l'intermédiaire en particulier de la méthode des éléments finis. La méthode des éléments finis, qui consiste à diviser un système physique au comportement complexe en composantes simples au nombre de paramètres limité pour simuler, par leur juxtaposition le comportement d'un système complet. De là est née l'idée de représenter une image en procédant à une approximation par morceaux par l'assemblage de facettes, chacune remplie d'une texture appropriée. Ainsi, même si ces petites briques prises indépendamment ne permettent pas d'expliquer ou de rendre un comportement général, leur assemblage dans une structure globale permet de reproduire une structure au comportement beaucoup plus complexe.[16]

II.3.1 : Définitions et propriétés

Le maillage est une opération de génération d'une grille 2D de l'image par définition d'une partition, en éléments polygonaux, basée sur des caractéristiques homogènes pour la représentation réduite des données de cette image. Le but de cette opération est que l'image approchée par cette représentation soit la plus fidèle possible à l'image originale. Les sommets des polygones sont appelés nœuds et ces nœuds sont reliés entre eux par des arcs.[16]

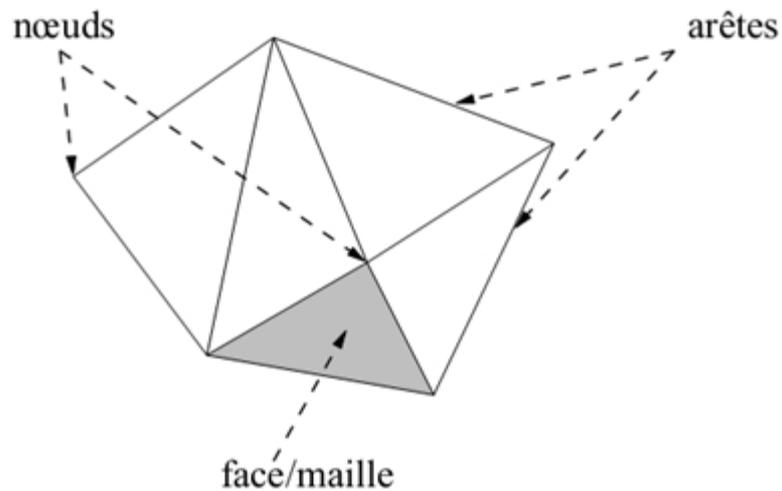


FIG II.9 : Les nœuds, les arêtes, et les faces définissant un maillage.

Même si, en théorie, les polygones de la division sont quelconques, en pratique, dans un maillage, tous les polygones sont du même ordre. Les maillages triangulaires, quadrangulaire et hexagonal ne sont formés respectivement que de polygones à trois, quatre et six côtés (c'est-à-dire d'ordres trois, quatre et six, donc des triangles, des quadrilatères et des hexagones). Les maillages les plus utilisés sont les maillages triangulaires et quadrangulaires. la figure suivante illustre un exemple de maillages triangulaires et quadrangulaires :

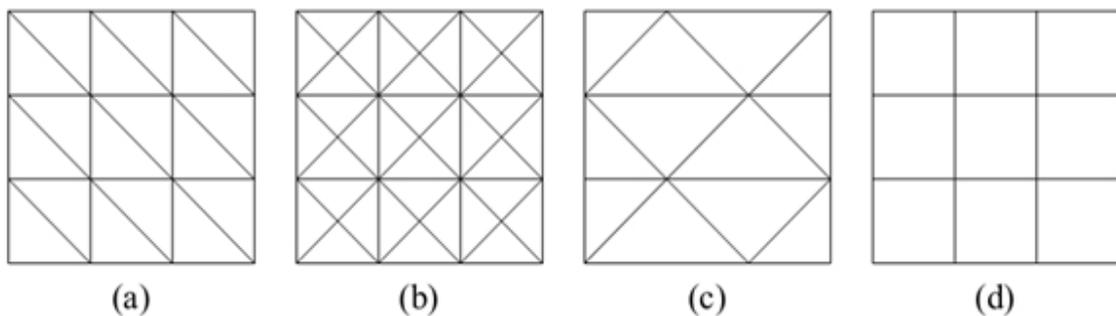


FIG II.10 : Exemples de maillages triangulaires (a), (b) et (c), et quadrangulaire (d)

Maillage conforme :

Un maillage est dit conforme si l'intersection de deux arêtes distinctes est un nœud ou l'ensemble vide, sinon il est dit non conforme. Il est à noter qu'à la construction le choix est toujours de générer un maillage conforme. Il pourra devenir non conforme lors du déplacement d'un de ses nœuds. Cette propriété de conformité est une condition suffisante assurant la continuité des polynômes d'interpolation d'une face à l'autre. La figure suivante donne exemples des maillages conformes et non conformes :

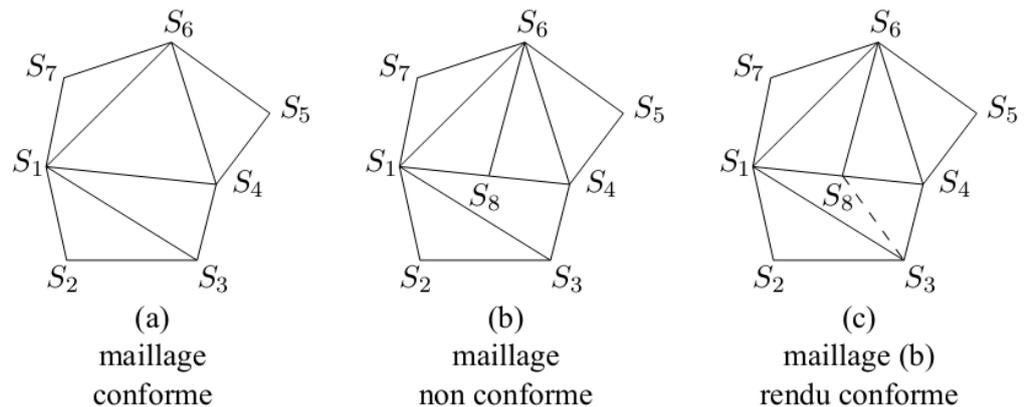


FIG II.11: Exemples de maillages conformes et non conformes

Tous les maillages de la figure précédente sont conformes dans le cadre d'un maillage composé de polygones d'ordre quelconque. Par contre, seul le maillage (b) est non conforme dans le cadre d'un maillage triangulaire. Ce maillage est rendu conforme par l'ajout de l'arête en tiret du maillage (c).

Maillage régulier/irrégulier :

Un maillage composé de polygones identiques est dit régulier sinon il est dit irrégulier.[17]

L'avantage d'un maillage régulier réside dans la simplicité de sa définition, et dans le fait que les positions des points d'une cellule (triangle ou quadrilatère) sont identiques d'une cellule à une autre, à un facteur de translation ou de symétrie près. Par conséquent, ces positions peuvent être mémorisées et permettre un balayage rapide des points de l'image.

L'inconvénient de ce maillage est sa non-flexibilité: par exemple, les arêtes du découpage ont fort peu de chances de coïncider avec les contours (gradients spatiaux ou temporels respectivement pour une image fixe ou une séquence vidéo), ce qui rend ces maillages uniformes souvent inappropriés pour refléter des zones voisines de ces contours, compliquant par conséquent l'exactitude du rendu des scènes. D'un point de vue suivi d'objets vidéo un seul patch peut contenir des mouvements multiples mais cette unique "facette" ne pourra pas rendre fidèlement ces différents mouvements. Ce problème peut se résoudre par un découpage plus fin de cette partie ou par un repositionnement ce qui résulte en un maillage irrégulier.

Une partition irrégulière permet un choix non contraint du nombre de sommets et leur concentration dans des zones d'intérêt mais aussi un placement libre de façon à être plus précis ou plus adapté au contenu [16]. La figure suivante illustre un exemple de maillage irrégulier obtenu à partir d'un nuage de points plus dense sur les contours d'objets. [17]

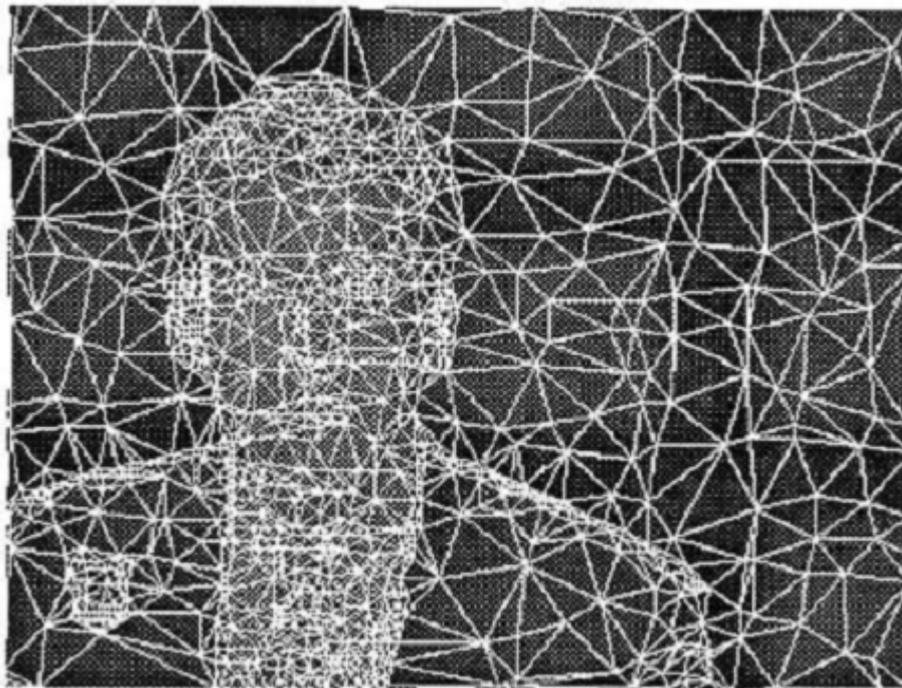


FIG II.12: Exemple de maillage irrégulier

Un maillage est dit statique si la position de ses nœuds est invariante dans le temps. La figure suivante illustre un exemple de maillage statique dans la représentation compressée d'une image :

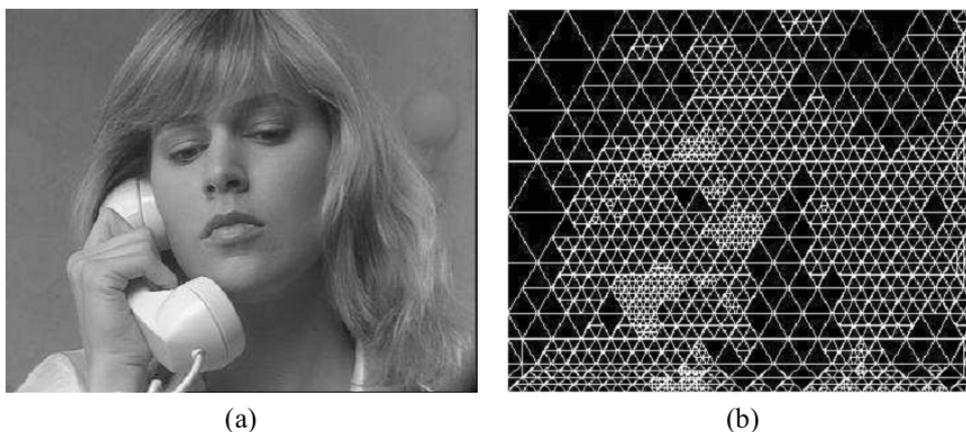


FIG II.13: Exemple d'utilisation du maillage statique (a) image originale ; (b) maillage non conforme adapté à l'image [17]

Par extension, le maillage déformable est caractérisé par la position des nœuds comme variable dépendante du temps. Un exemple d'utilisation des maillages dynamiques est la représentation synthétique d'un champ de mouvement entre deux images : [15][17]

- *Grilles quadrangulaires :*

Le mouvement de l'image est estimé à partir du déplacement des points de contrôle (ou nœuds) formant une grille uniforme. Cette grille partitionne l'image en quadrangles. La grille est construite sur l'image au temps t , le déplacement des nœuds est effectué sur l'image au temps $t+1$ afin de diminuer l'erreur de compensation de mouvement comme illustré dans la figure suivante :

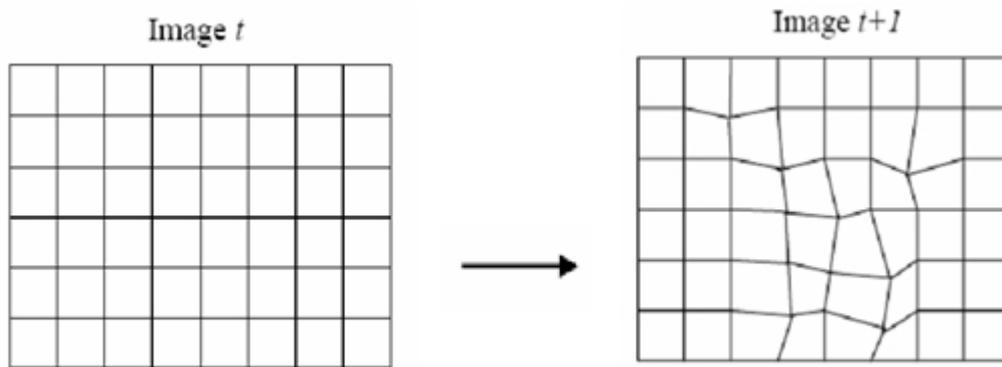


FIG II.14 : Exemple de grille active (quadrangles)

Le mouvement de chaque pixel est déterminé par interpolation du mouvement des nœuds des points de contrôle autour du pixel considéré. Chaque région possède alors son modèle de mouvement, déterminé par le mouvement des points de contrôle.

- *Maillages triangulaires :*

La méthode de mise en correspondance par maillages triangulaire est similaire à l'approche de mise en correspondance par grille, à la différence près que les régions formées sont triangulaires au lieu d'être rectangulaires comme illustré dans la figure suivante : [16]

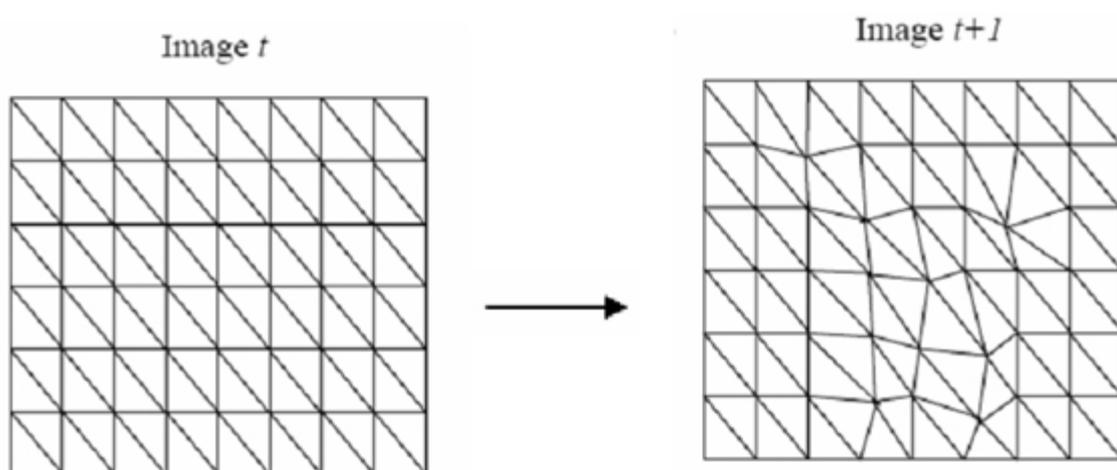


FIG II.15 : Mise en correspondance par maillage triangulaire

A partir d'un maillage déformable, il est possible de définir une fonction de déformation associée à chaque point du domaine maillé, et non pas seulement à ses nœuds. Pour cela, il faut rajouter un modèle d'interpolation du mouvement, qui détermine le mouvement d'un point quelconque à partir du mouvement des nœuds. Les modèles classiquement utilisés font dépendre le déplacement d'un point uniquement du déplacement des nœuds de la maille auquel il appartient. Par exemple : pour les maillages triangulaires, le déplacement d'un point est une fonction affine du déplacement des nœuds de la maille, pour les maillages quadrangulaires, il s'agit d'une interpolation bilinéaire. Le point p de la figure suivante se déplace ainsi en p' , et son déplacement est calculé par interpolation des trois déplacements des nœuds mis en évidence :

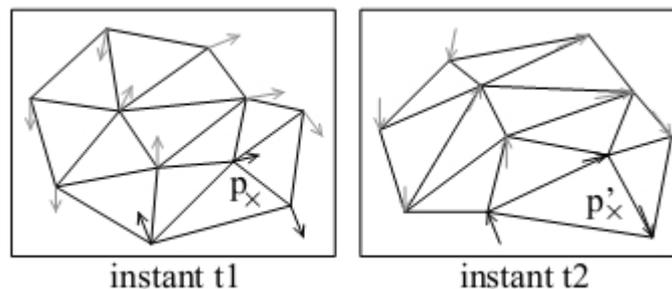


FIG II.16 : Maillage déformable

Il est intéressant de noter qu'étant donné un maillage à deux instants donnés t_1 et t_2 , cette transformation est non ambiguë, totale et symétrique : il est possible de définir pour chaque point du domaine au temps t_1 son unique correspondant au temps t_2 , et réciproquement.

II.3.2: suivi d'objets en mouvement par maillage

La structure de maillage déformable définie précédemment est capable de représenter une certaine classe de mouvements. Pour que la notion de mouvement soit valide dans une vidéo, les images successives doivent représenter la même scène, filmée depuis deux positions voisines, et à des instants rapprochés, c'est pourquoi on reste dans le cadre d'un plan audiovisuel unique. Il reste à déterminer les paramètres de ce mouvement à partir de l'observation. Par rapport au maillage ceci s'exprime de la façon suivante: étant donné un maillage sur une des images, comment déplacer les nœuds de façon à ce que le déplacement induit par le modèle correspond au déplacement apparent des objets situés sous les nœuds de cette image à la suivante.

Deux approches sont envisagées pour estimer le mouvement des nœuds : à caractère seulement locale, ou prenant en compte le maillage et son modèle d'interpolation dans leur globalité.

- Approche locale :

Le déplacement est estimé indépendamment pour chaque nœud à l'aide d'un champ dense de mouvement (estimation deux optique) ou par block matching. Présentant une certaine instabilité, cette approche est surtout utilisée pour l'initialisation. En effet, étudier seulement un voisinage d'un nœud pour estimer son mouvement ne permet pas toujours de connaître le déplacement de l'objet précisément (problème de l'aperture).

Ce problème apparaît en estimation de mouvement dans des séquences vidéo pour tous les points ne présentant pas des caractéristiques de contraste convenable. En particulier, il est impossible de connaître le mouvement dans une zone de couleur homogène. Plus généralement, en se limitant à des observations locales autour d'un point, son mouvement ne peut être calculé que dans une direction présentant un contraste.

Pour remédier à cela, il est nécessaire de faire des hypothèses supplémentaires de régularité du mouvement, ce qui permet d'intégrer l'information disponible sur une plus grande zone.

- Approche par minimisation d'énergie globale :

Etant donné un ensemble de déplacements pour les nœuds, une énergie quantifie la qualité de celui-ci par rapport à la paire d'images : une faible énergie correspond à des déplacements cohérents avec l'observation, alors qu'une énergie plus élevée marque leur incompatibilité.

L'estimation des paramètres de déplacement s'exprime alors sous la forme de la minimisation de cette énergie. L'information de mouvement peut ainsi, par une définition appropriée de l'énergie, être intégrée sur toute la surface du maillage, apportant une meilleure stabilité.

Différence d'image déplacée (DFD)

L'énergie utilisée pour l'estimation du mouvement par maillage est dans la littérature toujours basée sur la différence d'images déplacées. Celle-ci est définie comme la différence entre l'une des images et le résultat de la déformation de l'autre image conformément aux déplacements des nœuds. Ainsi, étant donné un ordre temporel sur les deux images de la paire considérée, deux formulations symétriques (avant et arrière) peuvent être adoptées, dépendant du choix de l'image déformée.

- Formulation avant :

Les algorithmes de codage fonctionnent de la façon suivante: pour reconstruire la seconde image du côté du récepteur en connaissant la première image, il suit de transmettre les déplacements des nœuds, et une image d'erreur résiduelle. Cette image d'erreur est en fait la différence entre l'image prédite (obtenue par déformation de la première image conformément aux paramètres de déplacement du maillage) et la vraie seconde image, d'où le

nom de différence d'images déplacées (DFD). Plus elle contient des valeurs faibles, plus efficace sera sa compression.

On cherchera donc dans ce cadre à minimiser une énergie globale exprimée généralement comme la somme des carrés des erreurs locales :

$$DFD = \sum_{p' \in R_2} (I_2(p') - I_1(p' + \delta_{2 \rightarrow 1}(p')))^2 = \sum_{p' \in R_2} (DFD_{2 \rightarrow 1}(p'))^2$$

Ou :

R_2 représente le domaine recouvert par le maillage dans l'image 2, $I_2(p)$ (resp. $I_1(p)$) représente la valeur du pixel p dans la seconde (resp. première) image.

$\delta_{2 \rightarrow 1}(p')$: est le déplacement associé au point p' de la seconde image vers la première par le modèle de mouvement,

DFD (p') est l'erreur locale de prédiction au point (p').

- Formulation arrière :

De manière symétrique, en déformant la seconde image, et en la comparant avec la première, on définit la DFD avec déformation arrière :

$$DFD = \sum_{p' \in R_1} (I_1(p') - I_2(p' + \delta_{1 \rightarrow 2}(p')))^2 = \sum_{p' \in R_1} (DFD_{1 \rightarrow 2}(p'))^2$$

Avec des notations analogues à celles de l'équation précédente. La différence fondamentale entre ces deux formulations tient au domaine sur lequel la sommation est calculée, et au fait que le maillage est fixé sur la première image et pas sur la seconde.

La figure suivante explique l'algorithme DFD par formulation arrière :

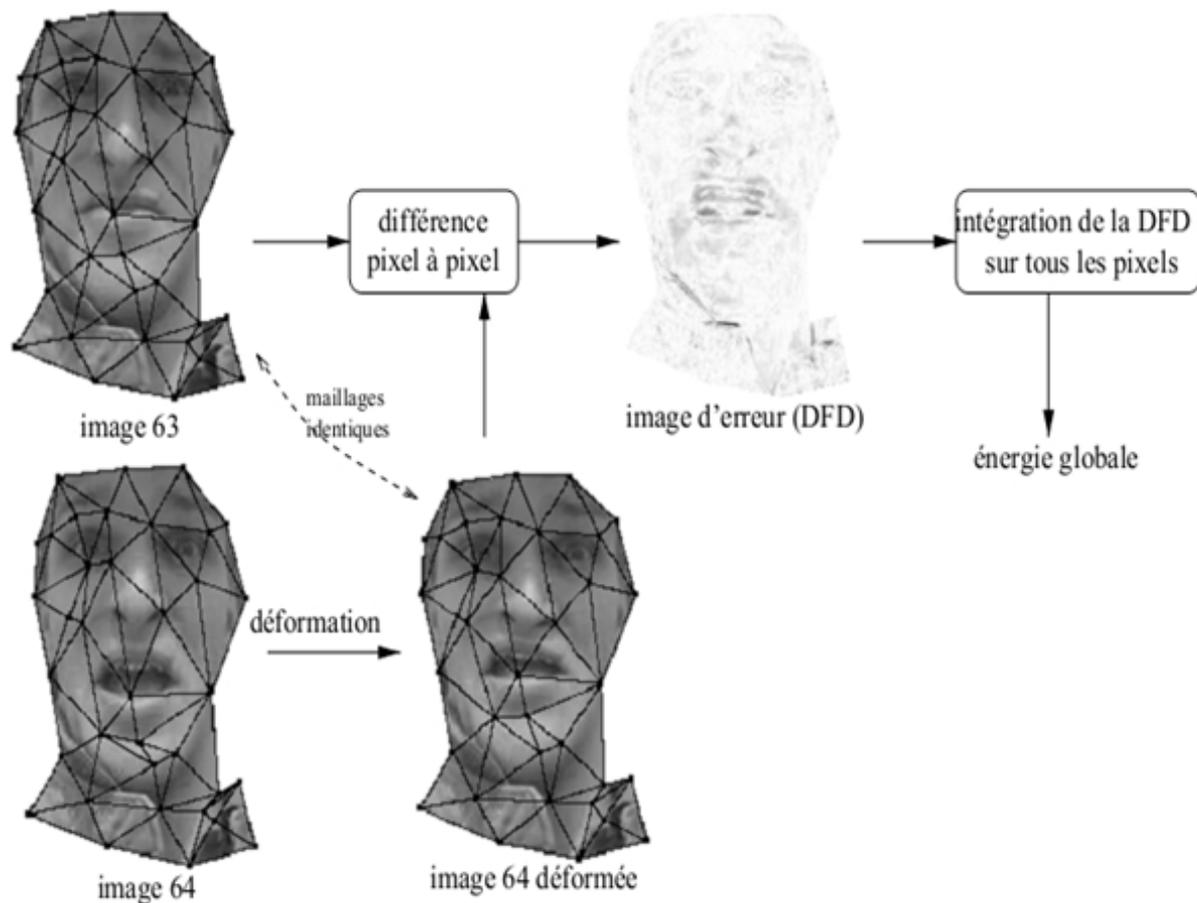


FIG II.17 : Principe du calcul de la DFD locale et la DFD globale pour une détection arrière

Modèle du suivi :

Etant donné que l'on suit un objet dans sa globalité, un maillage lui sera associé, de façon à recouvrir sa partie visible. Ainsi, contrairement à certaines approches visant à estimer les mouvements sur la totalité de l'image, il est concentré uniquement sur la partie de l'image contenant l'objet. Ceci présente un avantage par rapport à la régularité du mouvement: le maillage ne recouvrant qu'un seul objet, les discontinuités de mouvement seront beaucoup moins nombreuses que dans le cas d'un maillage global.

Le maillage seul représente la zone suivie. Bien que seuls des domaines polygonaux puissent être représentés ainsi, cela permet d'approximer toute région suffisamment régulière, avec une précision réglable : pour avoir des contours plus précis, on peut diviser les mailles, rajoutant ainsi de nouveaux nœuds sur le contour.

Les déplacements sont calculés par minimisation de la DFD. Contrairement aux approches de codage, une déformation arrière est effectuée. En effet, dans la déformation avant, le domaine R_2 sur lequel on somme les DFD dépend des paramètres de déplacement. Ceci n'intervient pas lorsque les nœuds du maillage sont forcés à rester sur les bords de l'image, mais à une influence sur le résultat dans notre cas, où le maillage est une sous partie de l'image. En effet, le nombre de pixel considérés varie, ce qui peut conduire au problème suivant : dans des régions présentant un faible contraste, la minimisation de la DFD consiste en fait à minimiser le nombre de pixels sur lequel s'effectue la sommation, ce qui correspond au resserrement du maillage sur lui-même. Dans le cas d'une déformation arrière, la sommation fait toujours intervenir le même nombre de pixels, ce qui élimine le risque d'une telle dégénérescence.

Comme souligné précédemment, l'approche par minimisation de DFD se base sur une hypothèse classique d'invariance des niveaux de gris. La minimisation de ce critère utilise une méthode de descente de gradient au 1^{er} ordre. Comme toute méthode de descente de gradient, il existe un risque de converger vers un minimum local, spécialement lorsque l'initialisation est éloignée du résultat. Pour remédier à cela, un processus de multi-résolution est utilisé.

Les images de référence sont réduites par filtrage gaussien et sous-échantillonnage d'un pixel sur deux en largeur et en hauteur. On obtient ainsi une pyramide d'images correspondant à des réductions de l'image originale de facteurs 1, 2, 4, . . . etc. dans chaque dimension. Sur l'image la plus réduite, la minimisation conduit à une estimation grossière des déplacements. Ceux-ci sont mis à l'échelle de l'image d'échelle supérieure, pour affiner les résultats, et le processus est répété jusqu'à arriver à l'image originale.

L'utilisation de la DFD seule n'assurant pas la stabilité des résultats, on introduit une contrainte de rigidité pénalisant les variations de la longueur des arêtes du maillage :

$$E_{rigidité} = K \sum_{e \in A} (r_e - r'_e)^2$$

Ou :

A : représente l'ensemble des arêtes du maillage,

r_e (resp. r'_e) la longueur de l'arête e dans la première (resp. deuxième) image.

K : une constante réglant la rigidité du maillage,

Cette énergie supplémentaire s'ajoute à la DFD, et est prise en compte lors de la minimisation de l'énergie globale.

Pour définir le maillage initial sur la première image, il ne peut pas être exigé de l'opérateur le placement explicite de tous les nœuds et des arêtes. Ainsi un module permettant la définition de la région à suivre par l'opérateur est nécessaire. Il s'agit d'un problème d'interface entre l'opérateur humain et le système de suivi, qui se révèle assez indépendant de la méthode de suivi effectivement utilisée.

Dans nos expérimentations, cette étape est traitée de la manière suivante : la région à suivre est décrite par un contour polygonal entré manuellement par l'opérateur. En effet, la notion de contour est certainement plus aisée à manipuler pour celui-ci. A partir de là, la traduction vers les structures internes à la méthode de suivi est automatique : le système calcule le maillage recouvrant la zone, puis le suivi pour le reste du plan. Le résultat de ce suivi peut être affiché surimposé aux images originales de la séquence.

L'algorithme suivant peut être utilisé pour la formation du maillage : Le contour polygonal est d'abord triangulé par une triangulation de Delaunay contrainte par les arêtes du contour, et les triangles extérieurs sont éliminés. Ensuite, les triangles contenant des arêtes de longueur supérieure à un seuil fixé à l'avance sont divisés comme représenté à la figure suivante (les arêtes en gras sont divisées en leur centre, les arêtes rajoutées sont en pointillés).

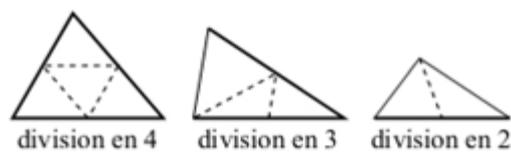


FIG II.18 : Division d'une maille possédant une ou plusieurs arêtes trop longues

La séquence *foreman* est utilisée classiquement pour estimer la qualité des procédés de codage/compression. La caméra est mobile (portée à l'épaule), et filme un personnage dont la tête bouge dans le champ de vision et cligne des yeux. Celui-ci se met à ouvrir la bouche à partir de l'image 63. D'une image à la suivante, le mouvement est généralement de moins de 2 pixels, sauf à l'ouverture de la bouche où il est de l'ordre de 10 pixels autour de celle-ci. L'hypothèse de mouvement lent est donc valide. D'autre part, les conditions d'illumination changent peu, validant l'hypothèse de constance des niveaux de gris.

La figure ci-dessous représente le suivi par maillage de la tête du personnage avec l'utilisation d'une contrainte de rigidité de facteur $K=1000$. Cette valeur est donnée pour comparaison avec les autres expériences, mais n'a bien sûr pas de signification absolue.



FIG II.19 : Suivi simple sur la séquence *foreman*

Le suivi du visage est fidèle sur les 75 images. La DFD globale associée reste toujours inférieure à 3.10^6 , bien qu'il y ait une légère augmentation à l'ouverture de la bouche. L'objectif fixé est ainsi atteint : localisation de l'objet dans chaque image. Le suivi des mouvements internes permet en outre de détecter un étirement de la zone autour de la bouche. Les qualités de ce suivi apparaîtront certainement plus claires au vu d'exemples où la validité des hypothèses est moins bonne.

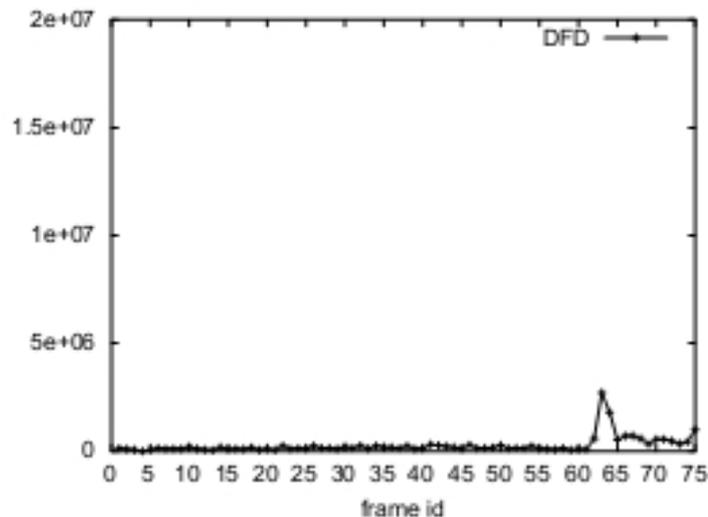


FIG II.20 : DFD globale sur le suivi de la séquence foreman

Influence des variations de luminosité :

Les images 8993 à 9006 de la séquence comme illustré dans la figure ci-dessous montrent sur la gauche du champ un personnage s'avançant, avec une ombre d'abord présente sur la droite, et qui le traverse vers la gauche. Les paramètres du système sont les mêmes que précédemment.



FIG II.21 : Suivi d'une partie d'un personnage sur laquelle passe une ombre portée

Dans cet exemple, le maillage s'accroche à tort sur l'ombre, ce qui montre clairement la limite de l'hypothèse d'invariance des niveaux de gris (gray level constancy) pour l'estimation du mouvement. En effet, celle-ci est rendue invalide par la présence de l'ombre dont le mouvement diffère du mouvement réel du personnage.

Effet des occlusions :

La séquence présentée à la figure ci-dessous représente le suivi du buste d'un personnage, progressivement masqué par un autre personnage passant devant lui à la vitesse approximative de 16 pixels par image. Le suivi a été calculé pour deux critères de rigidité différents: $K=1000$ pour (a), et $K=8000$ pour (b).

Dans le cas peu rigide, le personnage en avant plan amène le maillage à se déformer fortement: les mailles présentes sur des zones recouvertes sont ainsi dans un premier temps déplacées vers la gauche, ce qui conduit à une accumulation des mailles sur la droite de la partie visible du personnage cache. Dans un second temps, l'instabilité numérique du système prend le dessus. Ainsi, les quelques mailles qui se retrouvent partiellement sur le personnage de devant le suivent à partir de ce moment, ce qui entraîne une incohérence: une partie des mailles suit le personnage initial, une autre le personnage en avant-plan.

Dans le cas rigide, le maillage reste stable, gardant à peu près la même forme que le maillage initial. Pendant que le personnage initialement suivi occupe la plus grande partie du maillage, c'est celui-ci qui impose à toutes les mailles son déplacement, grâce aux contraintes de rigidité. Par la suite, le maillage suit le mouvement dominant de la région qu'il recouvre.



FIG II.22 : Suivi par maillage d'un personnage occulté, avec peu de rééité (a), et forte rééité (b)

Il est intéressant de rapprocher ces résultats géométriques avec les images d'erreur en valeur absolue associées comme illustré dans la figure suivante :



FIG II.23 : Image de l'erreur DFD en valeur absolue pour l'image 6395 avec faible (a) et forte (b) rééité

Les niveaux de gris clair représentent une faible erreur, alors que les teintes plus sombres correspondent aux zones à forte erreur. La progression de l'occlusion dans l'image d'erreur est plus visible avec un maillage rigide: on voit en effet nettement un front d'erreur marqué sur la droite, qui localise correctement les zones recouvertes. Dans le cas du maillage peu rigide, la déformation de celui-ci limite l'apparition de ce front d'erreur.

La figure ci-dessous représente l'erreur DFD globale au maillage pour chaque image de la séquence. Pour situer les courbes par rapport à la séquence: l'occlusion commence à l'image 6393, et le personnage initial est entièrement caché à partir de l'image 6397.

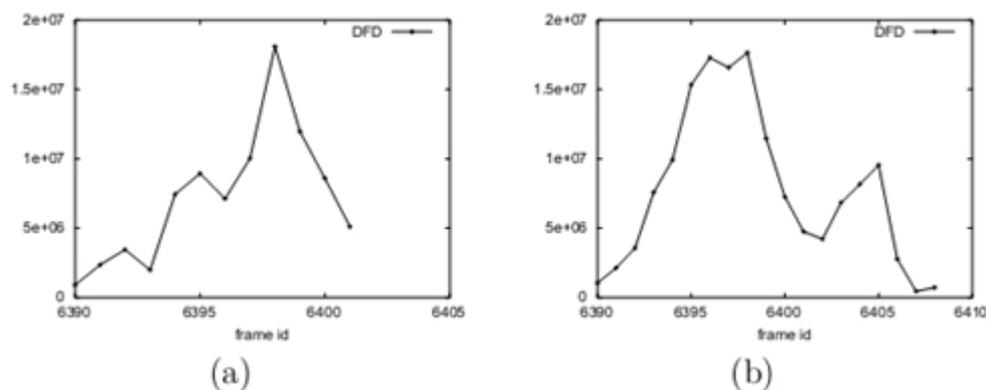


FIG II.24 : DFD globale pour la séquence, avec faible (a) et forte (b) rigidité

Au niveau de la DFD globale, le cas peu rigide voit un décollage plus tardif de l'erreur, ce qui rejoint l'observation faite sur l'image d'erreur: la déformation plus importante du maillage compense l'erreur locale. La DFD associée au maillage rigide représente plus clairement l'occlusion: il y a une montée rapide de l'erreur autour de l'image 6393, puis une redescente après la fin de l'occlusion.

Discussion

Avantage :

Le modèle du suivi d'objets mobiles est un modèle riche, capable de déformations locales. Son utilisation évite un morcellement en petites régions ou une approximation grossière lorsque les objets déformables sont suivis.

Inconvénients :

Les expérimentations ont mis en évidence des comportements non souhaités dans les cas où les hypothèses de base ne sont pas vérifiées : changements de luminosité et occlusion.

La résolution de ces problèmes peut être approchée de deux manières: la prévention de ceux-ci, par des contraintes supplémentaires sur le maillage, ou bien leur détection, suivi éventuellement d'une correction.

II.4 : Conclusion

Le suivi visuel d'objets dans des séquences d'images est devenu une des tâches les plus utilisées, parmi celles développées pour la vision par ordinateur. Plusieurs méthodes du suivi sont envisagées, qui se diffèrent selon la nature du modèle géométrique utilisé.

Parmi les méthodes du suivi on a vu la méthode du suivi par contours actifs qui consiste à initialiser une courbe et la faire évoluer de façon à minimiser l'énergie. Et la méthode du suivi par maillage qui consiste à suivre les objets mobiles par la déformation d'un maillage.

Conclusion générale

Le présent mémoire aborde le suivi d'objets en mouvement à partir d'une séquence vidéo après les avoir détectés et estimer leur déplacement.

Le but central de cette étude consiste à déterminer la configuration d'une cible ou les caractéristiques d'un objet dans une séquence d'images, tant que cette cible est en mouvement apparent dans les images.

Pour répondre à ce besoin, plusieurs méthodes du suivi sont envisagées, qui se différencient selon la nature du modèle géométrique utilisé.

Le travail a été consacré à la méthode du suivi d'objets en mouvement par contours actifs, qui a été appliquée au suivi de joueurs de football au cours d'un match. Elle a permis une implémentation rapide du modèle de contour actif, sans filtrage global ni estimation de mouvement, et elle analyse des séquences d'images couleur et est robuste au mouvement de la caméra. Ainsi que la méthode de suivi d'objets en mouvement par maillage qui consiste à suivre les objets mobiles par la déformation d'un maillage par la variation des positions des nœuds des polygones qui constituent le maillage selon le déplacement de l'objet mobile.

Enfin, en espérant que ce modeste travail contribuera si peu soit-il à la réalisation des systèmes de télésurveillance et aidera les étudiants à mieux comprendre ce vaste domaine.

Liste des références

[1]: Matthieu Fradet. Contributions à la segmentation de séquences d'images au sens du mouvement dans un context semi-automatique, these de doctat, Université de RENNES 1,2010.

[2]: Mathias PINGAULT. Estimations fréquentielle et temporelle du mouvement en transparence aditive dans les séquences d'images, these de doctat, Université JOSEPH FOURIER.

[3]: NICOLAS VERBEKE. Suivi d'objets en mouvement dans une séquence vidéo, these de doctat, Université PARIS DESCARTES,2007

[4]: INI_pdf

[5]: Jaime Gallego Vila. Foreground Segmentation and Tracking based on Foreground and Background Modeling Techniques, Master Thesis, Universitat Politecnica de Catalunya, 2009

[6]: M. LAOUAR NIZAR, M. LARABA MOHAMMED SEDDIK, Detection d'un Mouvement dans une Séquence Vidéo par Filtres Morphologiques, Mémoire d'ingénieur en électronique ENP, 2009

[7]: Antonio MARIN HERNANDEZ, Vision dynamique pour la navigation d'un robot mobile, these de doctorat, spécialité Informatique et Télécommunications, Laboratoire d'Analyse et d'Architecture des Systèmes du CNRS, 2004

[8] : S. Lefèvre, C. Fluck, B. Maillard, N. Vincent : Un modèle de contour actif pour le suivi rapide d'objets en mouvement. Application au suivi de joueurs de football, Université de Tours, 2001

[9]: Muriel GASTAUD, Modélisation de contours actifs pour la segmentation d'images et de vidéos, these de doctorat, Université de Nice-Sophia Antipolis, 2005.

[10] N. Khalifa, A. Malek et K. Hamrouni, Segmentation d'images par contours actifs : Application à la détection du ventricule gauche dans les images de scintigraphie cardiaque, Ecole Nationale d'Ingénieurs de Tunis (ENIT),2005.

[11]: A. Fekir, N. Benamrane et A. Taleb-Ahmed, Détection et suivi d'objets dans une séquence d'images par contours actifs.

[12] : Ariane HERBULOT, Mesures statistiques non-paramétriques pour la segmentation d'images et de vidéos et minimisation par contours actifs, these de doctorat, Université de Nice - Sophia Antipolis, 2007.

[13]: Frédéric PRECIOSO, Contours actifs paramétriques pour la segmentation d'images et vidéos, these de doctorat, Université de Nice - Sophia Antipolis, 2004

[14] : Sebastian LEFÈVRE, Nicole VINCENT, Contours actifs pour le suivi d'objet en temps-réel: multi-topologies et multi-résolutions, Université René Descartes – Paris V.

[15]: Rémi Megret, Suivi par maillages dans des séquences télévisuelles, mémoire de DEA, INSA de Lyon, 2000.

[16]: Amina BOUNAS, Développement d'une technique de maillage : application à la segmentation et au suivi d'objet, mémoire de magistère en électronique, Université Mentouri – Constantine.

[17]: Pascaline PARISOT, Suivi d'objets dans des séquences d'images de scènes déformables De l'importance des points d'intérêt et du maillage 2D, these de doctorat, Université de TOULOUSE, 2009.