

## Suivi de segmentation spatio-temporelle pour la compression de séquences animées.

V. Garcia-Garduno, C. Labit

IRISA/INRIA - Rennes , Campus de Beaulieu 35042 Rennes Cedex, France  
FAX :(33)99 38 38 32, victor.garduno@irisa.fr, labit@irisa.fr

### RÉSUMÉ

Dans le cadre des schémas de codage de séquences d'images animées, les techniques de compression basées sur une segmentation en régions homogènes au sens de mouvement s'avèrent être une alternative efficace et moins arbitraire que les techniques usuelles de compensation de mouvement basées pixel ou blocs de taille et géométrie arbitraires (segmentation en blocs carrés, quadtree ..). Nous proposons dans cette étude une technique de compression de séquences animées basée sur une segmentation spatio-temporelle de la scène observée partitionnée en régions homogènes au sens du mouvement modélisé par des modèles affines et dont un module de suivi temporel assure la cohérence tout au long de la séquence. Nous détaillons ci-après les différents éléments de cette chaîne de compression.

### 1 Introduction

Pour améliorer le caractère arbitraire des segmentations par blocs [9], ou quadtree [8],[5], il est possible d'étendre la méthode de segmentation basée mouvement sur la base de régions à géométries quelconques. Celle-ci opère de la manière suivante : à partir d'une segmentation spatiale fine de la scène [2], et grâce à un critère de fusion spatio-temporel au sens du mouvement une nouvelle segmentation est obtenue<sup>1</sup>. Dans le cadre de cette étude on considère que l'on dispose d'une telle segmentation. Etant donné le coût opératoire pour l'obtention de cette segmentation, cette approche n'a de sens que si la segmentation n'est pas recalculée systématiquement entre deux images successives mais bénéficie d'une procédure de suivi temporel efficace des régions segmentées, ce qui assure à la fois une cohérence temporelle de la segmentation et un gain en coût de codage important. Cet article se décompose en 4 parties :

- la description de la modélisation et estimation du champ apparent 2D.
- la description d'un module de prédiction courte-terme sur la base des trois étapes : prédiction, ajustement et estimation, en précisant les notations utilisés et les hypothèses de départ
- la description d'un module de prédiction long-terme par techniques de filtrage récursif intégrant un horizon temporel plus large [3]
- des résultats expérimentaux sur une séquence synthétique.

### ABSTRACT

In the field of image sequence coding schemes, motion-compensated compression techniques based on region-oriented segmentation appear to be a more efficient and less arbitrary alternative to the usual block-based or pixel-based motion compensation techniques. In this study we don't introduce any arbitrary segmentation grid (square blocks, quadtree ..) but propose a motion-based segmentation module ; affine motion models are identified and used to define a temporal tracking of regions which ensures the coherence throughout the sequence. The different parts of this motion compensation are detailed below and some preliminary results are shown.

### 2 Modélisation et estimation du champ de mouvement apparent 2D

Avant toute étape de segmentation basée mouvement, il convient de spécifier le cadre de modélisation du champ de mouvement 2D (optical flow) choisi. Nous avons étudié [6] dans un cadre de schéma de compression par compensation de mouvement basée région l'utilisation de modèles linéaires simplifiés permettant, sur une région spatiale, l'identification de mouvements translationnels parallèles au plan image, de mouvements de divergence (ou translation parallèle à l'axe optique), de mouvements de rotation (à axe de rotation parallèle à l'axe optique). Nous ne présentons que le principe général de l'algorithme et le formalisme. Dans [10] peuvent être trouver les détails de cette méthode. Nous ne nous intéressons qu'au cas du modèle de mouvement linéaire simplifié  $\Theta_{MLS}$  à 4 paramètres  $(t_x, t_y, k, \Theta)$ . Il est possible de généraliser la méthode dite "du gradient", utilisée par exemple dans le cadre des méthodes pel-récurrentes, en estimant non plus directement le déplacement  $\vec{d}$  pixel par pixel mais le vecteur de déplacement  $\Theta_j$  région par région [7]. Pour chaque région  $R_j$ , l'équation itérative suivante est défini :

$$\Theta_{j,MLS}^{i+1} = \Theta_{j,MLS}^i - \frac{\epsilon}{2N_j} \vec{G}^i$$

avec :

$$\vec{G}^i = \sum_{(x,y) \in R_j} \left\{ \begin{array}{l} \frac{\partial}{\partial t_x} DFD^2(x, y, \vec{d}_{j,MLS}^i) \\ \frac{\partial}{\partial t_y} DFD^2(x, y, \vec{d}_{j,MLS}^i) \\ \frac{\partial}{\partial k} DFD^2(x, y, \vec{d}_{j,MLS}^i) \\ \frac{\partial}{\partial \theta} DFD^2(x, y, \vec{d}_{j,MLS}^i) \end{array} \right\} \quad (1)$$

<sup>1</sup>Ce travail est mené au sein de la convention de recherche CCETT/IRISA 92ME 03 en collaboration avec le laboratoire IRESTE de NANTES

notations:

DFD : "displaced frame difference"



$N_j$ : nombre de pixels de la région  $R_j$ .

$\epsilon$ : matrice de gain.

$\Theta_{j,MLS}^i$ : estimation à l'iteration  $i$ .

Mise en oeuvre de la méthode.

L'algorithme du gradient peut se resumer selon le schéma suivant:

- **Initialisation :**  
Pour chaque région  $j$  de l'image faire :  
initialisation de  $\Theta_{j=1..N}^0$
- **Estimation :**  
calcul de  $\Theta_{estim}^{i+1}$   
si  $DFD(\vec{p}, \vec{d}_{estim}^{i+1}) < DFD(\vec{p}, \vec{d}_j^i)$   
alors  $\Theta_j^{i+1} = \Theta_{estim}^{i+1}$   
sinon  $\epsilon = \epsilon/2$   
 $i = i + 1$   
si  $i < \text{itermax}$  refaire estimation  
sinon fin de l'estimation
- **Relaxation :**  
 $k=0$   
Pour chaque région  $j$  de l'image faire :  
si  $DFD(j, \Theta_{s \in \text{voisinage de } j}) < DFD(j, \Theta_j)$   
alors  $\Theta_j = \Theta_s$  et  $k = k + 1$   
si  $k \neq 0$  faire estimation  
sinon fin

Les deux phases à optimiser plus particulièrement sont : l'initialisation du vecteur  $\Theta_j^0$  et la relaxation déterministe. La phase d'initialisation joue un rôle important puisque les algorithmes du gradient sont très sensibles aux conditions initiales ; il convient donc de procéder à un choix optimal des vecteurs de mouvement initialisant l'estimation, en les choisissant non nuls. L'objectif de la deuxième phase qui est la relaxation déterministe est d'estimer le vecteur de paramètres  $\Theta_s$  optimum (au sens de la DFD). Grâce à un enchaînement estimation-relaxation, il est possible de parvenir à une minimisation en évitant des minima locaux.

### 3 Suivi court-terme

#### notations

- $\Theta_{j,t}^-$ : vecteur de paramètres associé à la région  $j$  de l'instant  $t$  vers  $t - 1$
- $\Theta_{j,t}^+$ : vecteur de paramètres associé à la région  $j$  de l'instant  $t$  vers  $t + 1$
- $\Theta_{j,t}^0$ : vecteur de paramètres initial associé à la région  $j$  de  $t$  vers  $t - 1$ .

#### hypothèse de départ

à l'instant  $t - 1$  on dispose :

- d'une segmentation notée  $S_{t-1}$  qui est fournie soit au travers d'une carte dense de labels soit par les approximations polygonales (liste de sommets) des régions.
- d'un champ de vecteurs de paramètres globaux optimaux :  $\{\Theta_{j,t-1}^-\}_{j=1..N}$ .

La phase de suivi court-terme (voir schéma 6) est constituée des trois étapes suivantes :

#### 3.1 Prédiction-projection de $t - 1 \rightarrow t$

Nous supposons donc préalablement estimés les vecteurs optimaux  $\{\Theta_{j,t-1}^-\}_{j=1..N}$  sur la carte de segmentation à l'instant  $t$ :  $S_{t-1}$ ; il est raisonnable d'utiliser ces vecteurs pour prédire la carte de segmentation en  $t$  notée :  $\{\hat{S}_t\}_{j'=1..M}$ . Le problème crucial qui se pose est donc que seuls les vecteurs de paramètres calculés de  $t - 1$  vers  $t - 2$  sont disponibles. Il

faut estimer les vecteurs  $\{\Theta_{j,t-2}^+\}_{j=1..N}$  qui nous permettent de prédire  $\hat{S}_t$ . Nous rappelons les deux possibilités d'obtention du champ de descripteurs  $\{\Theta_{j,t-2}^+\}_{j=1..N}$  à partir des  $\{\Theta_{j,t-1}^-\}_{j=1..N}$

#### 1. Prédiction de $\hat{S}_t$ par inversion du champ $\vec{d}_{t-1}^-$

Comme on connaît le vecteur  $\Theta_{j,t-1}^-$  on peut dériver le champ de vitesse apparentes 2D sur chaque point  $p(x, y)$  de la région  $R_j(t - 1)$  noté :  $\vec{d}_{t-1}^- = (dx, dy)_{t-1}$  alors la projection vers  $t$  est obtenue par inversion de ce champ en chaque point-pixel ou bien par inversion du champ en chaque point-sommet de la région si celle-ci est approximée polygonalement:

$$\vec{d}_{t-1}^+ = (dx, dy)_{t-1}^+ = -\vec{d}_{t-1}^- = (-dx, -dy)_{t-1}^- \quad (2)$$

#### 2. Prédiction de $\hat{S}_t$ par dérivation du vecteur $\Theta_{j,t-2}^+$

Puisque on a le vecteur  $\Theta_{j,t-1}^-$  donc on connaît pour chaque point-pixel  $p(x, y, t - 1)$  de la région  $R_j$  à l'instant  $t - 1$  son déplacement  $\vec{d}_{t-1}^-$  vers le point  $p(x, y, t - 2)$  à l'instant  $t - 2$ , que dans le cas du modèle MLS est donné par :

$$\vec{d}_{t-1}^- = \begin{pmatrix} t_x \\ t_y \end{pmatrix} + \begin{pmatrix} k & -\theta \\ \theta & k \end{pmatrix} \cdot \begin{pmatrix} x - x_g \\ y - y_g \end{pmatrix}$$

ou sous forme matricielle :

$$\vec{d}_{t-1}^- = p(x, y, t - 1) - p(x, y, t - 2) = T + Mp(x, y, t - 2)$$

alors le vecteur  $\vec{d}_{t-2}^+$  est donné comme :

$$\begin{aligned} \vec{d}_{t-2}^+ &= -(M + I)^{-1}T + ((M + I)^{-1} - I)p(x, y, t - 1) \\ &= T' + M'p(x, y, t - 2) \end{aligned} \quad (3)$$

Comme dans le cas précédant la projection est obtenue en appliquant le vecteur  $\vec{d}_{t-2}^+$  à chaque point-pixel ou point-sommet.

#### 3.2 Ajustement à $t$

Dans ce cas nous avons utilisé l'approche markovienne décrite dans [4], cette technique de fusion ascendante de points ("sites") en régions spatiotemporelles est basée essentiellement sur un critère de fusion minimisant l'équation de contrainte du mouvement (projection du vecteur de vitesse apparente sur les gradients spatiotemporels locaux de la fonction Intensité). Cette fusion ne porte que sur les points où il y a ambiguïté après la phase précédente de projection c'est-à-dire les points de l'image ( $t$ ) ayant aucun ou plus d'un antécédent dans l'image précédente ( $t - 1$ ).

#### 3.3 Estimation de $t \rightarrow t - 1$

L'algorithme d'estimation utilisée est celui décrit au paragraphe 2. Seule la phase d'initialisation va varier selon le résultat obtenu à la issue des phases de prédiction et ajustement. Une fois que la segmentation  $\hat{S}_t$  est obtenue, au niveau de l'initialisation de  $\Theta_t$ , deux cas sont à considérer :

- pour toutes les régions  $j$  de  $S_{t-1}$  retrouvées en  $\hat{S}_t$  on initialisera par :

$$\{\Theta_{j,t}^0\}_{j=1..N} = \{\Theta_{j,t-1}^-\}_{j'=1..N}$$

Ceci consiste en une simple duplication région à région des champs de descripteurs de mouvement de ( $t - 1$ ) vers ( $t$ )

- si des nouvelles régions en  $\hat{S}_t$  sont trouvées, puisque l'on ne dispose pas d'information (dans le cas du suivi court-terme) sur celles-ci, alors on initialisera par :

$$\{\Theta_{j,t}^{-0}\}_{j' > N} = \vec{0}$$

#### 4 Suivi long-terme

Quelques premières études ont été menées concernant l'introduction dans un tel schéma d'une prédiction long-terme. Dans ce cas le suivi long-terme de la segmentation spatio-temporelle est effectué par une combinaison d'un filtrage de Kalman [3], des paramètres  $\Theta_{j,t-1}^+$  du mouvement et d'un ajustement de la prédiction par Markov comme dans le cas précédent. Les différents paramètres de mouvement sont filtrés de façon indépendante, en utilisant un modèle de mouvement à accélération constante. Les avantages d'un suivi long-terme sont : une meilleure initialisation des paramètres de mouvement lors de la phase d'estimation (voir schéma 6) ; le suivi long-terme de la segmentation spatio-temporelle devrait permettre de donner des indications pour la gestion des zones de recouvrement-découvrement.

#### 5 Résultats expérimentaux

**Description de la séquence de test** La séquence utilisée, lors des expérimentations des algorithmes développés, est constituée de plusieurs images synthétiques <sup>2</sup> de taille 256x256 qui sont composés de 5 régions dont les textures sont différents les unes des autres (voir figure 1). Trois types de mouvement ont été synthétisés pour construire cette séquence : une composition d'une translation et rotation effectuée sur le grand et le petit rectangle et deux types de mouvements purs, divergence et rotation sur le rond et le carré. Par le mouvement des deux rectangles et du rond, la séquence présente de grandes zones couvertes et découvertes.

**Suivi court-terme** : la figure 2 montre la segmentation spatio-temporelle parfaite idéale de l'image originale trame 2 (voir figure 1), dans ce cas l'EQM obtenue est égale à : 43.8

- **Prédiction** : la figure 3 montre les ambiguïtés trouvées lors de la projection des pixels par dérivation du vecteur  $\Theta_{j,t-2}^+$ .
- **Ajustement et estimation** : la figure 4 montre le résultat après la phase d'ajustement et estimation de la segmentation (voir le schéma général 6). Si l'on regarde la zone d'ambiguïtés à droite du grand rectangle de la figure 3 et si l'on regarde la carte de segmentation après ajustement et segmentation on peut constater qu'une partie de cette zone a été bien affectée au petit rectangle.

**Suivi long-terme** : la figure 5 montre l'allure de la segmentation trame 2 projetée et ajustée jusqu'à la trame 6 la table 1 fournit les EQM au long de la séquence.

#### 6 Conclusion

Cette étude vise à la conception d'une méthode de compensation de mouvement dont la phase d'estimation de mouvement associée est orientée objets (régions). Par suivi temporel court-terme il est démontré sur séquences synthétiques comme naturelles, l'intérêt d'introduire une phase de prédiction projection dans le sens du mouvement estimé avant ré estimation. Une phase d'ajustement est néanmoins nécessaire afin de s'approcher au mieux des contours instantanés

EQM des trames 2,3 et 6			
	trames		
	2	3	6
DF	72.9	78.5	89.5
EQM suivi-court	43.8	33.1	48.7
EQM suivi-long	-	33.1	40.8

Table 1 : Evolution de l'EQM au long de la séquence

des frontières d'objets. Ce travail se poursuit par la conception efficace de gestion des découvrément et recouvrement de zones (par ajustement spatial ou par la prise en compte de mesures locales de mouvement [1]) et par une prédiction long-terme filtrée pour assurer ces tâches de suivi temporel de segmentation.

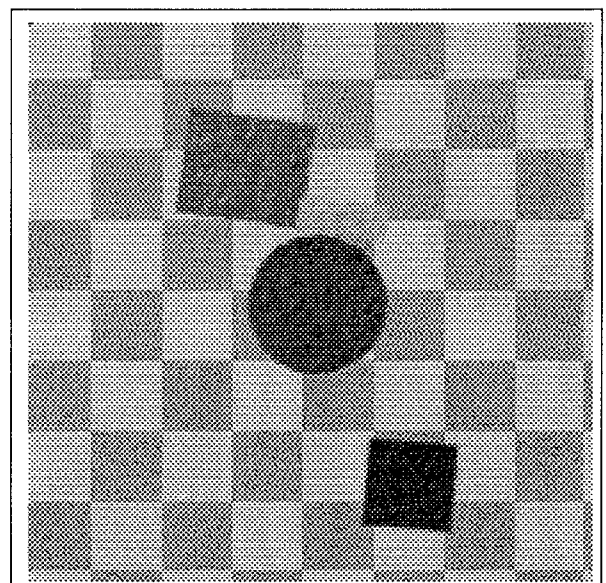


Figure 1 : Image synthétique trame 2

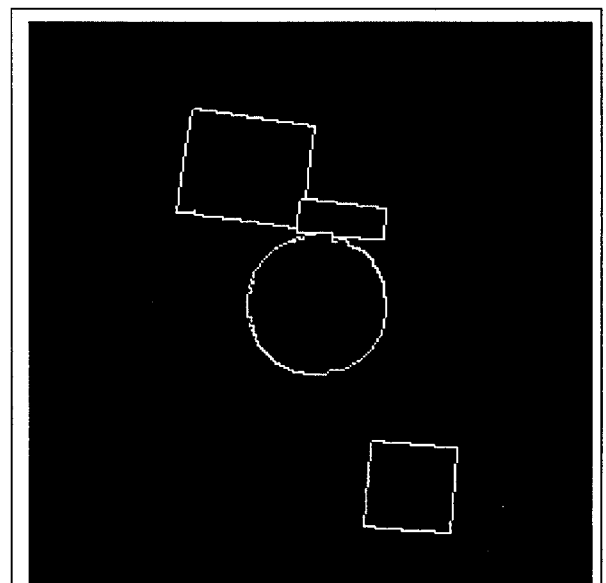


Figure 2 : Segmentation spatio-temporelle (trame 2, 5 régions) parfaite "idéale".

<sup>2</sup>images fournies par le laboratoire IRESTE de Nantes



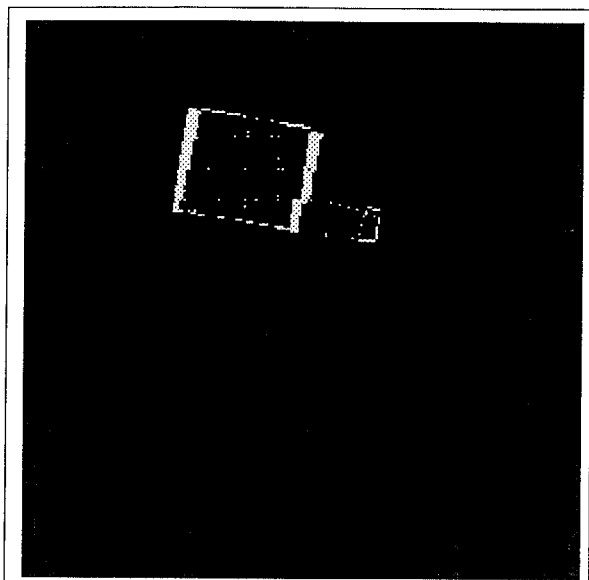


Figure 3 : En blanc les zones des ambiguïtés lors de la projection de pixels de la segmentation de la séquence synthétique (trame 2, 5 régions) par dérivation de  $\Theta_{j,t-2}^+$ .

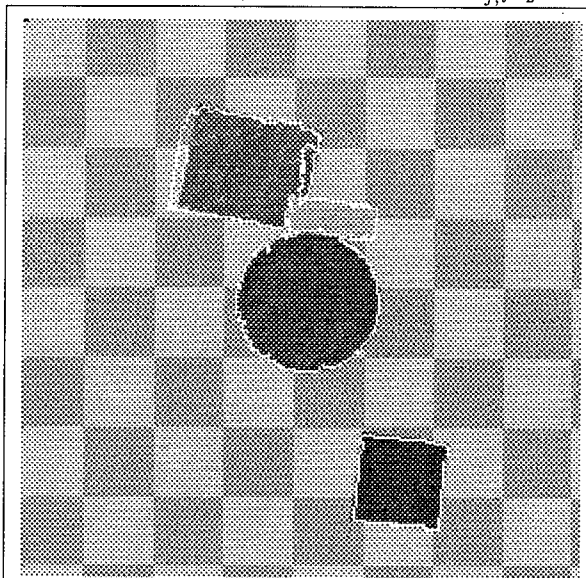


Figure 4 : Composition de la segmentation trame 2 projetée et ajustée sur l'image synthétique originale trame 3.

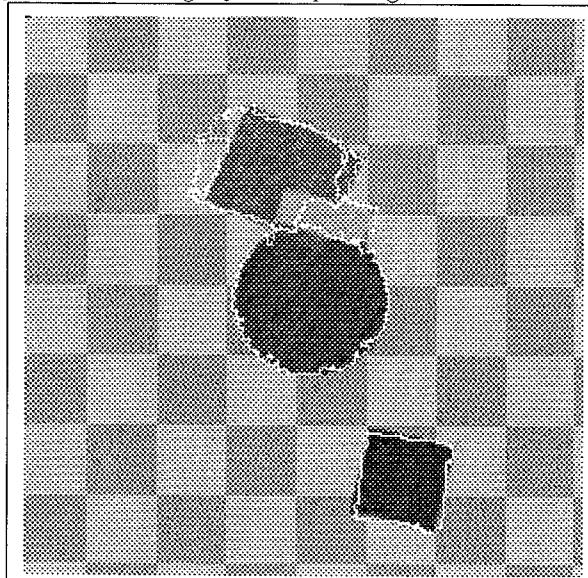


Figure 5 : Suivi long-terme, composition de la segmentation trame 6 projetée et ajustée sur l'image synthétique originale trame 6.

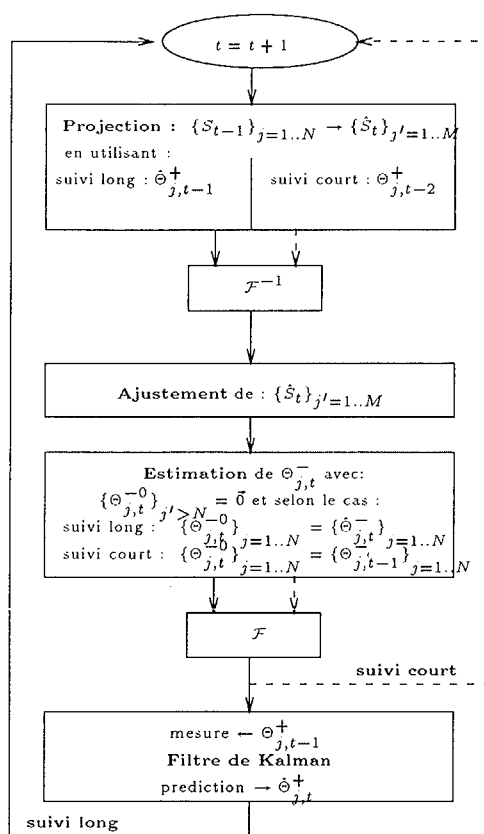


Figure 6 : Schéma général du suivi court-terme et long-terme utilisé

## Bibliographie

- [1] M. J. Black. Combining intensity and motion for incremental segmentation and tracking over long images sequences. In *Proc. of the 2nd European Conf. on Computer Vision (ECCV)*, page , Gènes, Mai 1992.
- [2] J. Benoit D. Barba. Image segmentation by region-contour cooperation for image coding. In *Proc. of IAPR*, page , 1992.
- [3] P. Bouthemy F. Meyer. Region-based tracking in an image sequence. In *Proc. of the 2nd European Conf. on Computer Vision (ECCV)*, page , Gènes, Mai 1992.
- [4] E. Francois. *Interprétation qualitative du mouvement à partir d'une séquence d'images*. Thèse de l'université de Rennes I, Juin 1991.
- [5] Samet H. Quadrees from boundary codes. In *Communications of the ACM*, pages 163-170, Mars 1980.
- [6] C. Labit H. Nicolas, A. Tamtaoui. Estimation des descripteurs de mouvement: application au codage de séquences monoculaires et stéréoscopiques. In *Colloque GRETSI'91*, pages 873-876, Juan-les-Pins, Septembre 1991.
- [7] J. Dugelay H. Sanson. Analyse de mouvements globaux dans le séquences d'images et application au codage de télévision. In *Colloque GRETSI'91*, pages 881-884, Juan-les-Pins, Septembre 1991.
- [8] C. Labit J. Santillana-Rivero, P. Bouthemy. Hierarchical motion-based image segmentation applied to hdtv. In *2nd Int. Workshop on Signal Processing of HDTV*, page , l'Aquila, mars 1988.
- [9] JAIN A.K. Jain J.R. Displacement measurement and its application in interframe image coding. *IEEE Trans. on Communications*, 29(12):1799-1806, 1981.
- [10] H. Nicolas. *Hierarchie de modèles de mouvement et méthodes d'estimation associées. Application au codage de séquences d'images*. Thèse de l'université de Rennes I, Septembre 1992.