

# Compensation de Mouvement par Maillage : Apprentissage et Maintien de la Connectivité

Pascaline Parisot<sup>1</sup>

Vincent Charvillat<sup>1</sup>

Géraldine Morin<sup>1</sup>

<sup>1</sup> LIMA – IRIT – UMR CNRS 5505

ENSEEIH, 2 rue Camichel, 31071 Toulouse - France

prenom.nom@enseeiht.fr

## Résumé

Dans cet article, nous présentons une méthode simple de suivi par maillage triangulaire qui maintient la connectivité du maillage tout au long du suivi. Cette méthode généralise la méthode de suivi d'un motif rigide par apprentissage de Jurie et Dhome [1] au cas d'un motif déformable. La méthode proposée est plus simple à implanter et plus rapide que les méthodes de suivi par maillage comme l'Hexagonal Matching [10], les méthodes d'Altunbasak et al. [2] ou les méthodes globales [12, 11]. En effet, grâce à une phase d'apprentissage qui peut être faite hors ligne pour toute une séquence vidéo, la phase de suivi est très rapide. Dans cet article, nous détaillerons notre méthode et montrerons qu'elle donne des résultats similaires à l'Hexagonal Matching.

## Mots clefs

triangulation, transformation locale, apprentissage, suivi temps-réel, relaxation.

## 1 Introduction

Les maillages 2-D dynamiques ont été étudiés dans différents travaux de recherche : le maillage hiérarchique et basé contenu [3, 4, 5], l'estimation de mouvement hiérarchique, le codage et la transmission permettant la mise à l'échelle, les modèles d'occultation et le remaillage. Historiquement destiné à la compression vidéo bas débit [6], le suivi par maillage 2-D est aussi un outil intéressant pour la segmentation d'objets, le codage ou la génération de mouvement [7, 8, 5].

La compensation de mouvement par maillage 2-D est une des solutions alternatives aux techniques classiques basées blocs.

Notre méthode de compensation de mouvement par maillage généralise les méthodes de suivi de motif rigide par apprentissage [9, 1] au cas de motif déformable (une triangulation) en assurant le maintien de la connectivité du maillage [2] (c'est-à-dire pas de "déchirement"). Cette méthode pourra être comparée aux méthodes de suivi par maillage comme l'Hexagonal Matching [10] ou les méthodes d'Altunbasak et Tekalp [2].

Dans la partie 2, nous établissons un bref état de l'art relatif aux travaux les plus proches de notre approche, présentée dans la partie 3, et comparables à celle-ci. Dans la partie 4, nous comparons notre méthode à l'Hexagonal Matching. Enfin, nous concluons et offrons quelques perspectives.

## 2 Etat de l'art

Les méthodes de suivi ont pour but d'estimer un champ dense de mouvement entre deux images successives  $I_{t-1}$  et  $I_t$  d'une séquence vidéo. Estimer un champ dense de mouvement revient à déterminer pour tout point  $p_{t-1}$  de  $I_{t-1}$  image d'un point 3-D  $P$  d'un objet réel visible sur  $I_{t-1}$  et  $I_t$ , le vecteur  $\overrightarrow{p_{t-1}p_t}$ , où  $p_t$  est l'image du même point 3-D  $P$  sur  $I_t$ .

Nous présentons ici les méthodes qui, comme celle que nous proposons, se basent sur le critère de la différence d'image déplacée (voir [2, 11, 12, 1, 9, 10]). Des critères basés sur le flot optique sont aussi utilisés (par exemple [13] et [2]). Ces méthodes font l'hypothèse de la conservation des niveaux de gris, c'est-à-dire que les niveaux de gris associés aux points  $p_{t-1}$  et  $p_t$ , selon les notations précédentes, sont invariants.

L'image est un ensemble fini de pixels ; le critère basé sur la différence d'image déplacée s'exprime donc dans le domaine discret et est donné par l'équation (1). La nullité de ce critère signifie que, pour tous les pixels  $p$  appartenant à un domaine  $\Omega$ , le niveau de gris de ce pixel sur l'image de référence  $I_{ref}$  est celui de la nouvelle position du pixel  $p$  ayant subi la transformation  $f$  paramétrée par le vecteur  $\mu$  sur l'image  $I_t$ .

$$E_t(\mu) = \sum_{p \in \Omega} (I_t(f(p; \mu)) - I_{ref}(p))^2 \quad (1)$$

Estimer le champ dense de mouvement consiste à minimiser ce critère non linéaire (1) ; c'est un problème d'estimation aux moindres carrés non linéaire.

Les méthodes relatives à notre approche peuvent être classées selon les choix, d'une part, des transformations  $f(\cdot; \mu)$  et de leur paramétrisation  $\mu$ , d'autre part, du domaine  $\Omega$  et, enfin, de l'image de référence  $I_{ref}$ .

Parmi toutes les méthodes basées sur le critère  $E_t$  (équation (1)), nous choisissons de présenter celles sur lesquelles notre méthode se base ainsi que celles auxquelles nous nous comparons. Nous présentons tout d’abord des méthodes récentes de suivi par apprentissage [9, 1] et ensuite des méthodes de suivi par maillage [2, 11, 12, 10].

## 2.1 Méthodes par apprentissage

Hager et Belhumeur [9] ont récemment proposé un cadre général pour le suivi d’objets plans texturés dans les séquences vidéo. Cette méthode se base sur le critère  $E_t$  (équation (1)) pour lequel  $\Omega$  est un motif d’échantillonnage (un ensemble de pixels épars). La figure 1 (a-haut) donne un exemple de motif dans un triangle. L’image de référence est  $I_0$ . Le critère  $E_t$  peut être réécrit ainsi :

$$E_t(\mu) = \| I_t(\mu) - I_0(0) \|^2,$$

où  $I_t(\mu)$  est le vecteur  $(I_t(f(p; \mu)))_{p \in \Omega}$  et  $I_0(0) = (I_0(f(p; 0)))_{p \in \Omega} = (I_0(p))_{p \in \Omega}$ . En supposant une petite variation  $\Delta\mu$  entre deux images successives  $I_t$  et  $I_{t+\tau}$  ( $\tau = 1$ , en général) et en effectuant un développement du premier ordre de  $I_{t+\tau}(\mu + \Delta\mu)$ , on obtient :

$$I_{t+\tau}(\mu + \Delta\mu) \approx I_t(\mu) + M(\mu, t)\Delta\mu + \tau \nabla_t I_t(\mu),$$

où la matrice  $M(\mu, t)$  (notée  $M$  par la suite) est la matrice jacobienne de l’image. En utilisant l’approximation  $\tau \nabla_t I_t(\mu) \approx I_{t+\tau}(\mu) - I_t(\mu)$ , on a :

$$E_t(\Delta\mu) \approx \| M(\mu, t)\Delta\mu + I_{t+\tau}(\mu) - I_0(0) \|^2.$$

En résolvant  $\nabla E_t(\Delta\mu) = 0$  et en posant  $\Delta I = I_{t+\tau}(\mu) - I_0(0)$ , on a :

$$\Delta\mu = -(M^t M)^{-1} M^t \Delta I.$$

On obtient donc une relation linéaire  $\Delta\mu = A_t \Delta I$  entre  $\Delta\mu$ , une petite variation des paramètres  $\mu$ , et  $\Delta I$ , une différence d’intensité lumineuse entre les pixels du motif déformé sur  $I_{t+\tau}$  et ceux du motif initial sur  $I_0$ . Hager et Belhumeur proposent d’estimer  $A_t$  grâce à l’inverse de la jacobienne  $M$  de l’image. Ils montrent qu’en supposant  $f$  inversible, dérivable et séparable en  $p$  et en  $\mu$ , cette matrice se décompose en le produit de deux matrices : l’une ne dépendant que de  $p$  qui peut donc être calculée une seule fois hors ligne, l’autre ne dépendant que de  $\mu$  qui est à calculer pour chaque image. Le suivi consiste à utiliser la dernière estimation des paramètres  $\mu$  pour calculer la différence  $\Delta I$  entre les niveaux de gris du dernier motif estimé et du motif de référence ainsi que la partie de  $A_t$  dépendant de  $\mu$ . Enfin, on obtient  $\Delta\mu$  par la relation  $\Delta\mu = A_t \Delta I$ .

Jurie et Dhome [1] ont fortement amélioré cette approche en proposant une méthode de suivi d’objets plans en temps réel, consistant à estimer les huit paramètres  $\mu$  d’une homographie. Ils proposent d’estimer, hors ligne, une unique matrice  $A = A_t$  (appelée matrice d’interaction) pour toutes

les images en résolvant le problème au sens des moindres carrés. Cette étape constitue l’apprentissage de  $A$  sur une image de référence. La deuxième étape, le suivi, détaillé dans le paragraphe 3, ne consistant qu’à faire un produit “matrice-vecteur”  $A \Delta I$ , est donc très simple. L’étape d’apprentissage permet ainsi d’effectuer le suivi en temps réel. Nous généralisons cette approche.

## 2.2 Méthodes par maillage triangulaire

Les méthodes de suivi par maillage consistent à compenser le mouvement par la déformation du maillage. Les déformations du maillage peuvent être paramétrées soit par un vecteur déplacement en chacun des nœuds du maillage (cas 1), soit par les six paramètres d’une transformation affine sur chaque triangle (cas 2). Dans le cas 2, des contraintes pour assurer la connectivité du maillage doivent être ajoutées, ceci est beaucoup plus coûteux. Dans le cas 1, le champ dense de mouvement est obtenu par interpolation des vecteurs de mouvement des nœuds du maillage. L’estimation de ces mouvements peut être faite de façon globale (tous les vecteurs déplacements sont estimés en même temps) ou de façon locale (les vecteurs déplacements sont estimés les uns après les autres en utilisant un algorithme de relaxation). Dans ce qui suit, l’image de référence  $I_{ref}$  est  $I_{t-1}$ .

**Méthodes globales.** Ici, on est dans le cas 1 où le domaine  $\Omega$  du critère  $E_t$  (équation (1)) est l’image entière. Dans ces méthodes globales, on s’intéresse à la minimisation itérative directe du critère non linéaire. Les méthodes itératives du premier ordre (basées sur une approximation du premier ordre de la fonctionnelle  $E_t$  à minimiser) ne sont plus utilisées car leur convergence est trop lente [12]. Les méthodes du second ordre (basées sur un développement de Taylor du second ordre) convergent plus rapidement [12, 11]. Néanmoins, ces approches globales sont coûteuses en calcul car elles nécessitent à chaque itération de résoudre un système de grande dimension ( $2N * 2N$  où  $N$  est le nombre de nœuds du maillage).

**Méthodes locales basées nœud.** La méthode d’*Hexagonal Matching* [10] (cas 1) est une approche locale par nœud. Elle considère les transformations nodales  $T_{mail}$  (figure 1 (b-bas)) définies ainsi : une transformation nodale est une transformation élémentaire d’un maillage qui consiste à ne bouger qu’un seul nœud du maillage (les autres étant fixes). Sur chaque triangle du maillage, la transformation nodale est une transformation affine : sur les triangles non adjacents au nœud déplacé, la transformation affine est simplement l’identité ; sur les triangles adjacents au nœud déplacé, la transformation affine est celle qui fait correspondre le triangle à un triangle où seul un sommet a bougé en une nouvelle position (figure 1 (b-bas)). L’*Hexagonal Matching* consiste à faire du *Block Matching* autour d’un nœud du maillage pour trouver un vecteur de mouvement initial et à affiner l’estimation du mouvement par une recherche exhaustive dans une fenêtre de recherche. Cette recherche minimise le critère

$E_t$  (équation (1)) où  $\Omega$  est le polygone support d'un nœud. L'utilisation des transformations  $T_{maill}$  pendant l'étape d'affinement impose un algorithme de relaxation. A cause de l'exhaustivité des recherches, cette méthode est très coûteuse. Elle fait cependant partie des plus performantes en terme de qualité du résultat [2].

Altunbasak et Tekalp [2] (cas 2) ont cherché à améliorer l'*Hexagonal Matching* en proposant des solutions analytiques localement optimales. Ils estiment pour chaque nœud les six paramètres de la transformation affine de chacun des  $T$  triangles dont il est sommet. Le paramètre  $\mu$  a pour dimension  $6T$  et  $\Omega$  est le polygone support. Pour maintenir la connectivité du maillage, ils rajoutent au problème de minimisation du critère  $E_t$  (équation (1)) des contraintes de connectivité qui consistent à dire qu'en un nœud, les transformations affines de tous les triangles dont il est sommet donnent le même vecteur déplacement. Le problème est résolu grâce au Lagrangien (condition du premier ordre). Cette méthode est moins coûteuse mais aussi moins performante que l'*Hexagonal Matching*.

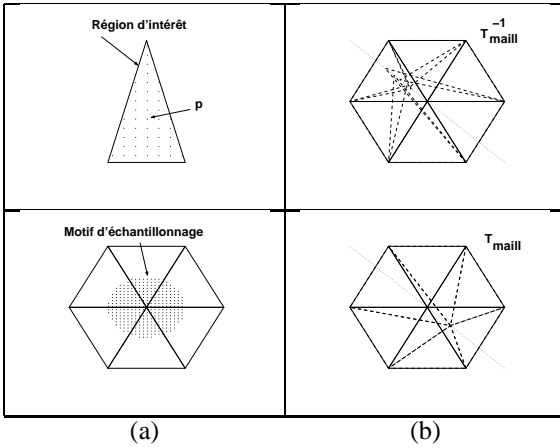


Figure 1 – (a) Exemples de motifs d'échantillonnage : (a-haut) d'une région d'intérêt, (a-bas) associé à un nœud du maillage ; (b) Transformation nodale  $T_{maill}$  : (b-haut) : apprentissage, (b-bas) : suivi.

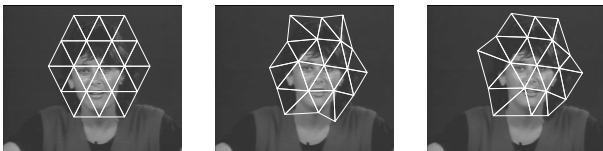


Figure 2 – Séquence vidéo "Miss America". Gauche : image de référence (image 1) et son maillage d'apprentissage. Milieu : image 78 et son maillage suivi par la méthode d'Hexagonal Matching. PSNR de 31.41 dB. Droite : image 78 et son maillage suivi par notre méthode. PSNR de 30.64 dB.

### 3 Notre méthode

Notre méthode est une méthode locale de suivi (cas 1). Elle généralise la méthode de suivi de Jurie et Dhome [1] au cas du suivi par maillage triangulaire 2-D en assurant le maintien de la connectivité du maillage. Pour cela, nous estimons un vecteur déplacement en chacun des nœuds du maillage en effectuant l'apprentissage d'une matrice d'interaction  $A_n$  par nœud  $n$ . Pour effectuer cet apprentissage, nous nous intéressons aux transformations nodales  $T_{maill}$  proposées dans l'*Hexagonal Matching* (paragraphe 2.2).

Notre méthode se base sur le critère  $E_t$  (équation (1)) où  $\Omega$  est un ensemble de points d'intérêt situés autour des nœuds du maillage (figure 1 (a-bas)),  $\mu$  est un vecteur déplacement (exprimé en coordonnées cartésiennes) par nœud et l'image de référence  $I_{ref}$  est  $I_0$ .

Comme la méthode de Jurie et Dhome [1], notre méthode se décompose en deux étapes. Pendant l'étape d'apprentissage, une matrice d'interaction par nœud est apprise et pendant l'étape de suivi, la déformation du maillage est estimée.

Dans [14], nous montrons que pour pouvoir suivre un déplacement  $d$  d'un nœud, il ne suffit pas d'apprendre le déplacement inverse  $d' = -d$  du nœud mais, pour chacun des triangles  $T$  dont ce nœud est sommet, il faut apprendre un déplacement  $d' = -k_T d$  où  $k_T$  est un coefficient dépendant du triangle  $T$ . Donc pour pouvoir suivre la transformation  $T_{maill}$  (figure 1 (b-bas)), il faut apprendre la transformation  $T_{maill}^{-1}$  (figure 1 (b-haut)).

La phase d'apprentissage est effectuée uniquement sur l'image de référence  $I_0$  sur laquelle la région d'intérêt à suivre est échantillonnée.

Afin de suivre une transformation nodale  $T_{maill}$ , le motif d'échantillonnage sur le polygone support (figure 1 (a-bas)) est perturbé (figure 1 (b-haut)) pour recueillir des différences d'intensité  $\Delta I$  entre le niveau de gris aux points du motif et celui aux points perturbés.

A partir de nombreuses expériences  $k$  ( $k \in [1..K]$ , où  $K \approx 400$ ), consistant à appliquer des perturbations  $(T_{maill}^{-1})^k$  puis à mesurer les  $\Delta I^k$  induits, on peut estimer au sens des moindres carrés, pour chaque nœud  $n$ , la matrice d'interaction  $A_n$  (grâce à la relation linéaire proposée par Jurie et Dhome [1]) telle que  $A_n \Delta I^k \approx \mu^k$ , pour tout  $k$ , où  $\mu^k$  est le vecteur des paramètres de la transformation nodale  $T_{maill}^k$  à suivre.

Pendant la phase de suivi entre deux images  $I_{t-1}$  et  $I_t$ , la connaissance de la position du motif sur  $I_{t-1}$  et des matrices d'interaction en chacun des nœuds suffit pour estimer le mouvement. Le simple calcul des différences de niveaux de gris  $\Delta I_t$  aux points du motif sur les images  $I_{t-1}$  et  $I_t$  et la multiplication "matrice-vecteur"  $A_n \Delta I_t$  nous donnent les paramètres de la transformation utiles au recalage. Cette phase de suivi, ne nécessitant aucun appariement, peut être effectuée en temps réel. Pour prendre en compte tous les nœuds, un algorithme de relaxation est mis en œuvre.

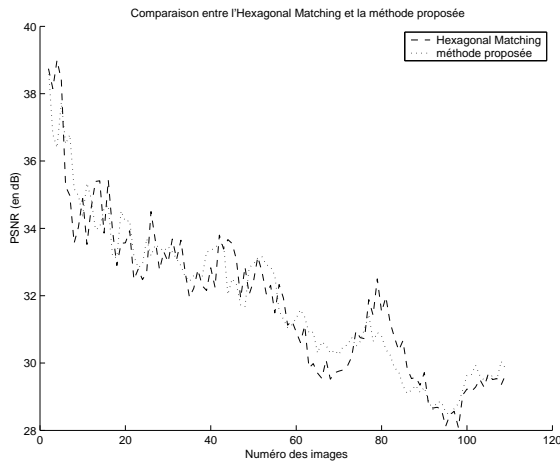


Figure 3 – Comparaison entre l'Hexagonal Matching et la méthode proposée sur la séquence vidéo "Miss America".

## 4 Résultats

Expérimentalement, nous avons montré que, localement, l'apprentissage théorique qui consiste à apprendre un déplacement du nœud différent pour chaque triangle donne de meilleurs résultats que l'apprentissage qui consiste à apprendre l'opposé du déplacement que l'on veut suivre (voir article de GTMG [14], url : <http://www.enseeiht.fr/~parisot>).

La figure 2 montre l'image de référence avec son maillage d'apprentissage ainsi que les maillages estimés sur la 78<sup>ième</sup> image de la séquence vidéo *Miss America* par la méthode de l'*Hexagonal Matching* et par notre méthode.

La figure 3 compare notre méthode à l'*Hexagonal Matching* sur la séquence vidéo *Miss America*. Nous utilisons la mesure du PSNR entre l'image originale et celle reconstruite à partir de la déformation estimée du maillage et de la texture de la première image de la séquence vidéo sur laquelle l'apprentissage a été effectué. D'après ce graphique, nous constatons que notre méthode donne des résultats similaires à l'*Hexagonal Matching* en ayant l'avantage d'être moins complexe. En effet, alors que pour l'*Hexagonal Matching* les recherches exhaustives sont à effectuer entre chacune des images successives, pour notre méthode, l'apprentissage est fait une seule fois pour la séquence vidéo puis, pour le suivi, il ne reste plus qu'à calculer des produits "matrice-vecteur" pour chaque image.

## 5 Conclusion

Notre méthode de suivi par maillage basé sur un apprentissage maintient la connectivité du maillage. Cette méthode donne des résultats similaires à celle de l'*Hexagonal Matching* qui est reconnue comme faisant partie des plus performantes en terme de qualité du résultat [10, 2], tout en étant moins complexe.

Cette méthode de suivi peut être utilisée pour la représentation d'un objet d'une séquence vidéo, pour la

compensation de mouvement ou encore pour le codage vidéo. Elle pourrait également être utilisée dans le cadre de modèles génératifs de mouvement.

## Références

- [1] F. Jurie et M. Dhome. Hyperplane approximation for template matching. *IEEE Trans on PAMI*, 2002.
- [2] Y. Altunbasak et A.M. Tekalp. Closed-form connectivity-preserving solutions for motion compensation using 2-d meshes. *IEEE Trans. on Image Processing*, 6(9), 1997.
- [3] Y. Wang et O. Lee. Active mesh - a feature seeking and tracking image sequence representation scheme. *IEEE Trans. Image Processing*, 3 :610–624, Septembre 1994.
- [4] Y. Altunbasak et A.M. Tekalp. Occlusion-adaptive content-based 2-d mesh design and forward tracking. *IEEE Trans. Image Processing*, 6(9), 1997.
- [5] S. Valette et I. Magnin et R. Prost. Active mesh for video segmentation and objects tracking. *Proc. IEEE ICIP'2001*, 2, Octobre 2001.
- [6] A. Buisson et N. Laurent. Un codeur hybride h264 - maillage. Dans *Actes de la conférence CORESA'03*, Lyon, Janvier 2003.
- [7] P. van Beek et A.M. Tekalp et N. Zhuang et I. Celasun et M. Xia. Hierarchical 2-d mesh representation, tracking and compression for object-based video. *IEEE Trans. on Circuits Systems for Video Technology*, 1999.
- [8] C. Toklu. Object-based digital video processing using 2-d meshes. *PhD Thesis, Univ of Rochester, USA*, 1998.
- [9] G.D. Hager et P.N. Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *IEEE Trans on PAMI*, 1998.
- [10] Y. Nakaya et H. Harashima. Motion compensation based on spatial transformations. *IEEE Trans. on Circuits Systems for Video Technology*, 4, 1994.
- [11] G. Marquant. Représentation par maillage adaptatif déformable pour la manipulation et la communication d'objets vidéo. *Thèse de Doctorat, Université de Rennes I*, Décembre 2000.
- [12] P. Lechat. Représentation et codage de séquences vidéo par maillages 2d déformables. *Thèse de Doctorat, Université de Rennes I*, Octobre 1999.
- [13] H. Brusewitz. Motion compensation with triangles. *Proc. 3rd International Workshop on 64 kbits/s Coding of Moving Video, free session*, 4, Septembre 1990.
- [14] P. Parisot, G. Morin, et V. Charvillat. Transformations locales d'une triangulation 2d pour la compensation de mouvement par maillage. Dans *Actes de GTMG (Groupe de Travail en Modélisation Géométrique)*, pages 113–122, Lyon, Mars 2004.