

Suivi par triangulation d'objets déformables : méthode par apprentissage

Learning based method for tracking non-rigid objects

P. Parisot¹

V. Charvillat¹

G. Morin¹

¹ IRIT – UMR CNRS 5505 site de l'ENSEEIH

2 rue Camichel, 31071 TOULOUSE
parisot,charvi,morin@enseeiht.fr

Résumé

Dans cet article, nous présentons une nouvelle méthode de suivi par maillage triangulaire qui maintient la connectivité du maillage lors du suivi. Cette méthode est basée sur la méthode de suivi de motif rigide, proposée par Jurie et Dhome [8]. La compensation de mouvement par maillage est une technique de suivi intéressante, en particulier pour le codage très bas débit.

Notre méthode est plus simple à implémenter et moins coûteuse que d'autres méthodes de suivi par maillage comme l'appariement hexagonal [13] ou les méthodes d'Altunbasak et al. [1]. En effet, grâce à une phase d'apprentissage qui peut être faite hors-ligne, la phase de suivi est simplifiée. En particulier, nous montrons que les transformations du maillage à apprendre ne sont pas intuitives et justifions l'intérêt de les paramétrer par les coordonnées barycentriques généralisées.

Mots Clefs

triangulation, transformation nodale, transformation locale, apprentissage, suivi temps-réel, relaxation, maintien de la connectivité du maillage, coordonnées barycentriques généralisées.

Abstract

In this article, we present a new triangular mesh tracking method preserving the mesh connectivity during the tracking. This method is based on the rigid template tracking method, proposed by Jurie and Dhome [8]. The mesh motion estimation is an interesting tracking technique, in particular, for very low bitrate coding.

Our method is simpler and of lower complexity than other comparable mesh tracking methods: Hexagonal Matching [13], Altunbasak et al. [1] method. In fact, thanks to a learning step that can be performed off-line, the tracking step is simpler. In particular, we show that the mesh transformations to learn are not intuitive and propose and justify a

parametrization using the generalized barycentric coordinates.

Keywords

triangular mesh, nodal transformation, local transformation, learning, real-time tracking, relaxation, mesh connectivity preserving, generalized barycentric coordinates.

1 Introduction

Le suivi de mouvement dans des séquences d'images se fait généralement par l'intermédiaire des différentes approches suivantes :

- la minimisation d'un critère fondé sur la différence d'image déplacée,
- l'équation du flot optique,
- l'appariement par blocs ou autres indices visuels.

Ces approches permettent d'estimer un modèle de mouvement choisi (rigide ou non-rigide).

Dans le contexte du suivi d'objets plans rigides et de la minimisation du critère fondé sur la différence d'image déplacée, Hager et Belhumeur [7] ont travaillé sur la linéarisation du critère et sur la modélisation des mouvements afin d'obtenir un algorithme de suivi efficace. Jurie et Dhome [8] ont rendu plus robuste la convergence de cette méthode. De plus, l'introduction d'une étape d'apprentissage leur permet d'effectuer le suivi en temps réel.

Dans le cadre du suivi d'objets non-rigides, Bookstein [3] propose une manière de trouver les paramètres d'un modèle de déformations non-rigides combinant une transformation affine et des non-rigidités (plus ou moins complexes) centrées en un nombre (plus ou moins grand) de points. Gleicher [6], quant à lui, propose une nouvelle méthode capable de mieux suivre des petites régions d'une image se déformant avec des transformations non-linéaires. Il introduit des transformations projectives par morceaux pour modéliser le suivi d'objets non-plans. Pour notre part, nous

nous intéressons à l'utilisation des maillages triangulaires modélisant des transformations affines par morceaux.

Les maillages triangulaires ont été étudiés dans différents travaux de recherche : le maillage hiérarchique et basé contenu [18, 2, 15], l'estimation de mouvement hiérarchique [10], le codage et la transmission permettant la mise à l'échelle [10], les modèles d'occultation [14]. Le suivi par maillage bidimensionnel, historiquement destiné à la compression vidéo bas débit [4], est un outil également très intéressant pour la segmentation d'objets, le codage ou la génération de mouvement [16, 14, 15].

Nous exposons dans cet article comment adapter la méthode de Jurie et Dhome [8] au cas du suivi d'objets déformables grâce à une triangulation tout en maintenant la connectivité du maillage.

Dans une première partie, nous présentons les transformations du maillage auxquelles nous nous intéressons. Dans une deuxième partie, nous posons le problème à résoudre, en l'occurrence nos hypothèses et le critère à minimiser. Dans une troisième partie, nous donnons un état de l'art sur la résolution de ce problème. Ensuite, dans une quatrième partie, nous proposons notre méthode pour le résoudre. Dans une dernière partie, nous donnons quelques résultats de validation. Finalement, nous concluons.

2 Transformations et transformations inverses

Dans ce paragraphe, nous nous intéressons aux déformations du maillage ainsi qu'à leur paramétrisation. Nous présentons le cas particulier des transformations nodales. Notre méthode utilisant aussi les transformations inverses, nous les définissons donc.

2.1 Transformations du maillage

Un maillage est déformé si au moins un de ses nœuds bouge (voir figure 1). Tandis que Nakaya et al. [13] et Altunbasak et al. [1] ont choisi de paramétrer cette déformation sur chaque triangle par les six paramètres de leur transformation affine, Lechat [9] et Marquant [10, 11] ont, quant à eux, choisi de la paramétrer en chacun des nœuds par les deux paramètres de leur déplacement cartésien. Nous avons opté pour cette dernière paramétrisation. En effet, tous les paramètres représentant une même donnée physique (un déplacement dans une des deux dimensions), ils ont la même sensibilité à une même perturbation. Ceci n'est pas le cas, par exemple, pour les six paramètres d'une transformation affine. En vue du suivi, si le maillage a N nœuds, nous avons donc $2N$ paramètres à estimer. Nous les regroupons dans un vecteur μ .

La figure 1 illustre une transformation du maillage. Toute déformation du maillage est donc une transformation affine par morceaux (les morceaux étant les triangles en traits pleins). La déformation inverse est elle-aussi une transformation affine par morceaux (les morceaux étant, cette fois-ci, les triangles en traits discontinus) : l'image du maillage en traits discontinus est le maillage en traits pleins.

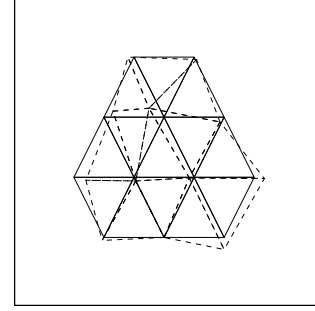


FIG. 1 – Exemple de maillage déformé : en traits pleins, le maillage initial ; en traits discontinus, le maillage déformé.

2.2 Transformations nodales

Toute transformation du maillage peut se décomposer en une succession de transformations nodales T_{nod} qui sont caractérisées par le déplacement d'un seul nœud du maillage (les autres nœuds ne bougeant pas) (voir figures 2 et 3-haut). Ces transformations sont donc les plus simples que l'on puisse appliquer à un maillage. C'est la raison pour laquelle nous nous y intéressons.

Nous supposons que la transformation nodale du maillage est un modèle de déformation localement exact, ce qui implique que l'image se déforme localement selon une transformation affine par morceaux (voir figure 2) définie sur chacun des triangles Δ_i de son support polygonal qui est délimité par l'ensemble des nœuds adjacents au nœud d'intérêt (voir figure 4).

Sur chacun des triangles Δ_i , la transformation affine Aff_i est celle qui fait correspondre le triangle à un triangle Δ'_i où seul le sommet C a bougé en une nouvelle position C' .

$$T_{nod} : \begin{array}{l} \Delta_i \rightarrow \Delta'_i \\ p \mapsto T_{nod}(p) = Aff_i(p) \end{array}$$

De manière naïve, on peut penser que la transformation nodale inverse de la transformation nodale qui déplace le nœud C d'un déplacement d est la transformation qui déplace le nœud C par le déplacement $-d$. Cette transformation $T_{naïve}^{-1}$ est illustrée sur la figure 3-a. Nous étudierons cette approche inexacte dans la partie validation (voir paragraphe 6).

De manière exacte, la transformation nodale inverse T_{nod}^{-1} est la transformation affine par morceaux qui ramène donc le nœud C' en la position C (voir figure 3-b).

$$T_{nod}^{-1} : \begin{array}{l} \Delta'_i \rightarrow \Delta_i \\ p \mapsto T_{nod}^{-1}(p) = Aff_i^{-1}(p) \end{array}$$

La transformation inverse naïve $T_{naïve}^{-1}$ diffère donc naturellement de la transformation inverse exacte T_{nod}^{-1} .

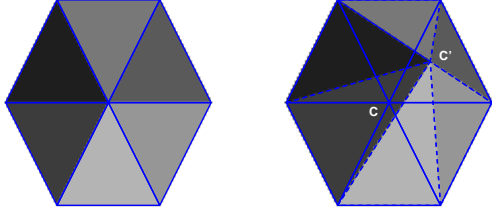


FIG. 2 – Transformation nodale : maillage initial en traits continus (à gauche et à droite), maillage ayant subi une transformation nodale en traits discontinus (à droite). À gauche : image initiale, à droite : image ayant subi une transformation nodale (c'est-à-dire une transformation affine par morceaux sur chacun des triangles du maillage).

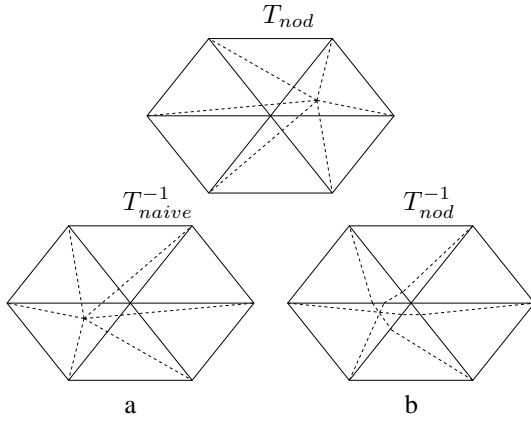


FIG. 3 – En traits pleins, le maillage initial. En traits discontinus, la transformation nodale T_{nod} qui maintient la connectivité du maillage. En bas : à gauche, en traits discontinus, la transformation nodale inverse naïve T_{naive}^{-1} du maillage ; à droite, en traits discontinus, la transformation nodale inverse exacte T_{nod}^{-1} du maillage.

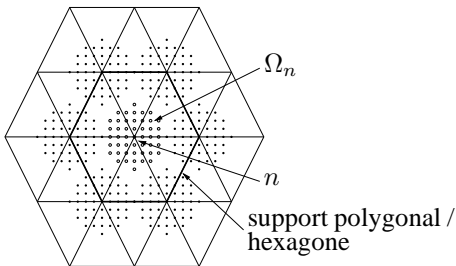


FIG. 4 – Exemple de motif d'échantillonnage. Ω_n : le motif d'échantillonnage associé au nœud n .

3 Position du problème

3.1 Nos hypothèses

Dans le cadre de notre méthode de suivi, nous faisons les hypothèses de travail suivantes :

- nous supposons que les déplacements à suivre d'une image I_{t-1} à la suivante I_t sont petits ;
- nous faisons l'hypothèse de l'invariance des niveaux de gris, c'est-à-dire que les niveaux de gris associés aux images d'un même point tridimensionnel sur les images I_{t-1} et I_t sont les mêmes ;
- nous supposons que chaque support polygonal (voir figure 4) a subi une déformation affine.

3.2 Les critères

Pour estimer les transformations nodales du maillage, nous nous intéressons au critère basé sur la différence d'image déplacée exprimé dans le domaine discret :

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

La nullité de ce critère signifie que pour tous les points p appartenant à un domaine Ω , le niveau de gris de ce point¹ sur l'image de référence I_{ref} (en général, $I_{ref} = I_0$ ou I_{t-1}) est celui de la nouvelle position du point p ayant subi la transformation f paramétrée par le vecteur μ sur l'image I_t .

Nous cherchons le vecteur μ_t qui minimise ce critère.

$$\mu_t = \operatorname{argmin}_{\mu} E_t(\mu) \quad (2)$$

Ce problème est un problème aux moindres carrés non-linéaire. Le vecteur μ_t contient les paramètres de la déformation du maillage entre les images I_{ref} et I_t .

La fonction f est une transformation affine par morceaux : affine sur chacun des triangles du maillage. Le déplacement de n'importe quel point p se trouvant à l'intérieur d'un triangle est obtenu par interpolation linéaire de Lagrange des déplacements des sommets du triangle.

Le critère global peut se décomposer en un critère local associé à chaque nœud n du maillage :

$$\forall n \in 1..N, E_t(\mu, n) = \sum_{p \in \Omega_n} (I_t(f(p; \mu)) - I_{ref}(p))^2, \quad (3)$$

où p est un point du domaine Ω_n , non dense dans notre cas, associé au nœud n (voir figure 4) et N est le nombre de nœuds du maillage. Le domaine Ω_n est défini comme étant l'ensemble des points régulièrement répartis à l'intérieur du cercle centré sur le nœud n et de diamètre la longueur de la plus petite arête partant de ce nœud.

La déformation d'un maillage pouvant être obtenue par composition de transformations nodales, nous allons dans un premier temps nous intéresser à ces transformations et

¹ Pour déterminer le niveau de gris en n'importe quel point de coordonnées non-entières, une interpolation bilinéaire entre les niveaux de gris des quatre plus proches pixels est utilisée.

à leurs critères locaux $E_t(\mu, n)$. Le vecteur de paramètres μ peut être réduit à seulement deux paramètres : les coordonnées cartésiennes du déplacement du nœud n . Nous reviendrons au critère global de l'équation (1) au paragraphe 5.4.

4 État de l'art de la résolution

Dans le cadre du maillage, parmi les méthodes existantes pour résoudre le problème (équation (2)), nous présentons dans ce paragraphe celles auxquelles nous pourrions nous comparer : la méthode d'appariement hexagonal [13] pour les transformations nodales, la méthode d'Altunbasak et al. [1] pour le maintien de la connectivité et les méthodes globales [9, 10, 11] pour le choix des paramètres. Ces méthodes peuvent être classées en deux catégories : les méthodes globales pour lesquelles tous les paramètres (c'est-à-dire tous les déplacements pour tous les nœuds) sont estimés en même temps et les méthodes locales pour lesquelles seul un sous-ensemble de paramètres est estimé localement.

4.1 Méthodes locales

La méthode d'appariement hexagonal [13] est une méthode locale basée sur le déplacement d'un nœud, décomposée en deux étapes, qui optimise les vecteurs mouvement de chaque nœud du maillage par une compensation de mouvement affine sur chaque triangle. La première étape de la méthode d'appariement hexagonal consiste à faire de l'appariement par bloc autour du nœud n considéré pour trouver un vecteur de mouvement initial. La seconde étape consiste à affiner cette estimation du mouvement par une recherche exhaustive à l'intérieur d'une fenêtre centrée sur le nœud. Pendant cette étape, les nœuds de l'hexagone support de n sont immobiles. Seul le nœud n peut bouger et ainsi engendrer des transformations affines sur chaque triangle. Le vecteur de mouvement du nœud n est perturbé dans la fenêtre de recherche. A chaque perturbation, les six paramètres affines de chacun des six triangles de l'hexagone sont recalculés ainsi que l'erreur de compensation de mouvement. Cette erreur peut être basée sur le critère E_t de la différence d'image déplacée défini par l'équation (1), où μ est le vecteur des trente-six paramètres à estimer, Ω est l'ensemble des pixels de l'hexagone considéré (voir figure 4) et l'image de référence est I_{t-1} . La perturbation qui minimise le plus ce critère est acceptée comme affinement du vecteur de mouvement.

À cause de l'exhaustivité des recherches, cette méthode est très coûteuse en temps de calcul. Par contre, elle fait partie des plus performantes en terme de qualité des résultats. Elle doit mettre en place un algorithme de relaxation pour prendre en compte tous les nœuds du maillage. Il est de plus possible de traiter simultanément plusieurs nœuds grâce à une partition des nœuds en trois sous-ensembles indépendants. À l'intérieur d'un sous-ensemble, le mouvement d'un nœud n'influe pas sur le mouvement des autres.

Altunbasak et al. [1] proposent des méthodes locales. Ils établissent des solutions analytiques localement optimales

pour l'estimation de mouvement. Ils proposent deux approches : une approche basée triangle dont nous ne parlerons pas ici et une approche basée nœud qui nous intéressera. Leur méthode n'est pas limitée aux hexagones. Elle fonctionne pour n'importe quel polygone. Ils s'intéressent à un nœud n et à son support polygonal ainsi qu'à ses sommets. Dans cette approche, ils cherchent à déterminer les paramètres des transformations affines de chacun des triangles du polygone. Pour cela, ils supposent connaître un champ dense de mouvement et ils cherchent à "traduire" ce champ en transformations affines par triangle tout en conservant la connectivité entre les triangles. Ces contraintes de connectivité traduisent donc le fait que les vecteurs déplacement du nœud considéré, obtenus par chacune des transformations affines des triangles du polygone, sont les mêmes. De même, le déplacement du sommet commun entre deux triangles du polygone est le même. Ce problème de minimisation avec contraintes est résolu grâce au Lagrangien.

Leur méthode a l'avantage de proposer une solution analytique. Elle est moins coûteuse que la méthode d'appariement hexagonal mais donne des résultats un peu moins bons.

4.2 Méthodes globales

Dans les méthodes globales, on s'intéresse à la minimisation itérative directe du critère non-linéaire E_t (équation (1)) pour lequel Ω représente l'ensemble des pixels de l'image, l'image de référence est I_{t-1} et le vecteur des paramètres μ contient $2N$ paramètres (où N est le nombre de nœuds du maillage). Parmi ces méthodes, les méthodes itératives du premier ordre (basées sur une approximation du premier ordre du critère E_t à minimiser) ne sont plus utilisées car leur convergence est trop lente [9]. Les méthodes du second ordre (basées sur un développement de Taylor du second ordre) convergent plus rapidement [9, 10, 11]. Néanmoins, ces approches globales sont coûteuses en calcul car elles nécessitent à chaque itération de mettre à jour un système de grande dimension ($2N * 2N$). L'utilisation de la méthode de Levenberg-Marquardt a été étudiée pour améliorer la convergence [9].

5 Estimation des transformations nodales par apprentissage

Nous avons vu que nous pouvons nous limiter à l'étude du mouvement d'un seul nœud. Nous allons maintenant étudier ce qui se passe entre l'image de référence I_0 , dans notre cas, et l'image suivante I_1 dans la séquence vidéo. Notre méthode adapte la méthode de Jurie et Dhome [8] pour les motifs plans à des transformations nodales. Elle se décompose en deux étapes : une étape d'apprentissage qui consiste à apprendre des relations entre des différences d'intensité lumineuse dues aux transformations nodales du maillage et les paramètres de cette transformation et une étape de suivi qui consiste à retrouver les paramètres de la transformation du maillage de manière à se recalculer sur

l'image.

Nous présentons tout d'abord le principe de la méthode. Ensuite, nous détaillons le cas, qui nous intéresse plus particulièrement, de l'apprentissage et du suivi des transformations nodales.

5.1 Suivi par apprentissage : le principe

A l'instar d'Hager et Belhumeur [7], notre méthode se base sur un critère $E_t(\mu, n)$ (équation (3)) pour lequel Ω_n représente un motif d'échantillonnage (c'est-à-dire un ensemble de pixels épars) (voir figure 4). L'image de référence est I_0 . Le critère $E_t(\mu, n)$ peut être réécrit ainsi :

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

où $I_t(\mu)$ est le vecteur $[I_t(f(p; \mu))]_{p \in \Omega_n}$ et $I_0(0) = [I_0(f(p; 0))]_{p \in \Omega_n} = [I_0(p)]_{p \in \Omega_n}$. 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 dans 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 :

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

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

$$\Delta\mu = -(M^T M)^{-1} M^T \Delta I,$$

où M^T est la matrice transposée de M . On obtient donc une relation linéaire $\Delta\mu = A^t \Delta I$ entre $\Delta\mu$, une petite variation des paramètres μ , et ΔI , une différence d'intensité lumineuse entre les pixels du motif déformé sur $I_{t+\tau}$ et ceux du motif initial sur I_0 . Les auteurs estiment A^t par la pseudo-inverse de la jacobienne M de l'image. Le suivi consiste à utiliser la dernière estimation des paramètres μ pour calculer la différence ΔI entre les niveaux de gris du dernier motif estimé et du motif de référence ainsi que A^t . On obtient donc $\Delta\mu$ par la relation $\Delta\mu = A^t \Delta I$.

Tout comme Jurie et Dhome [8], nous supposons que nous pouvons nous ramener à tout instant à l'image de référence, puis revenir à l'image courante. Alors que Jurie et Dhome utilise la transformation entre l'image de référence et l'image I_{t-1} , c'est-à-dire $f(\cdot; \mu_{t-1})$, nous avons choisi de substituer/remplacer cette transformation par l'utilisation des coordonnées barycentriques généralisées (voir paragraphe 5.2). Il est donc possible d'estimer une seule fois la matrice A sur l'image de référence, à partir de perturbations du motif d'échantillonnage sur celle-ci.

Dans le paragraphe 5.2, nous détaillons comment cette matrice est estimée dans le cas de la transformation nodale. Pendant le suivi, la connaissance de A , appelée matrice d'apprentissage, et le calcul des ΔI permettent de déterminer par simple produit "matrice-vecteur", les paramètres

de la déformation du maillage sur l'image de référence. Il ne reste plus qu'à revenir à l'image courante. Ce suivi peut donc s'effectuer en temps réel.

5.2 Apprentissage et suivi d'une transformation nodale

L'apprentissage. Nous nous intéressons à l'apprentissage d'une transformation nodale du maillage, par exemple au nœud n . Pour cela, nous utilisons le critère local $E_t(\mu, n)$ (équation (3)). Nous cherchons donc à apprendre la matrice A_n d'apprentissage du nœud n qui lie linéairement la différence d'intensité lumineuse ΔI aux points d'intérêt Ω_n , associés au nœud n , à un déplacement $\mu(n)$ du nœud n . 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_{nod} de l'image, le motif nodal d'échantillonnage Ω_n (voir figure 4) sur le support polygonal est perturbé par la transformation T_{nod}^{-1} pour recueillir des différences d'intensité ΔI entre le niveau de gris aux points du motif nodal et celui aux points perturbés :

$$\Delta I = [I_{ref}(T_{nod}^{-1}(p; \mu(n))) - I_{ref}(p)]_{p \in \Omega_n}$$

A partir de la simulation de nombreuses expériences j ($j \in [1..J]$, où $J \approx 400$), consistant à appliquer des perturbations $(T_{nod}^{-1})^j$ puis à mesurer les ΔI^j induits, nous pouvons estimer, au sens des moindres carrés, la matrice d'apprentissage A_n (qui suit le modèle linéaire proposé par Jurie et Dhome) telle que $A_n \Delta I^j \approx \Delta\mu(n)^j$, pour tout j , où $\Delta\mu(n)^j$ est le vecteur des paramètres de la transformation nodale $(T_{nod})^j$ à suivre.

Algorithme d'apprentissage :

Estimation de la matrice A_n d'apprentissage

- Initialiser les variables

+ $\Delta I = []$

+ $\Delta\mu = []$

- Pour chaque perturbation $\Delta\mu(n)^j$

+ Calculer ΔI^j

+ Construire $\Delta I = [\Delta I \ \Delta I^j]$

+ Construire $\Delta\mu = [\Delta\mu \ \Delta\mu(n)^j]$

Fin Pour

- Résoudre le problème :

$$A_n = \underset{A}{\operatorname{argmin}} \| A \Delta I - \Delta\mu \|^2$$

Le suivi entre I_0 et I_1 . Supposons que I_1 est l'image résultat de l'image I_0 qui a subi une transformation nodale au nœud n . Pendant l'étape de suivi entre ces deux images, la connaissance de la position du maillage initial, donc du motif sur l'image I_0 , et de la matrice d'apprentissage A_n suffit pour estimer le mouvement. Le simple calcul des différences de niveaux de gris $\Delta I(n)$ aux points du motif nodal sur les images I_0 et I_1

$$\Delta I(n) = [I_1(p) - I_0(p)]_{p \in \Omega_n}$$

et la multiplication “matrice-vecteur” $A_n \Delta I(n)$ nous donnent directement les paramètres $\Delta \mu(n)$ de la transformation utiles au recalage du nœud n sur l’image I_1 .

Le suivi entre I_{t-1} et I_t . Supposons que I_t est l’image résultat de l’image I_{t-1} qui a subi une transformation nodale au nœud n . Pendant l’étape de suivi entre ces deux images, l’image I_{t-1} étant quelconque, la connaissance de la position du maillage sur l’image I_{t-1} et de la matrice d’apprentissage A_n nous donne par le simple calcul des différences de niveaux de gris $\Delta I(n)$ aux points du motif nodal sur les images I_0 et I_{t-1}

$$\Delta I(n) = [I_t(f(p; \mu_{t-1})) - I_0(p)]_{p \in \Omega_n},$$

et de la multiplication ”matrice-vecteur” $A_n \Delta I(n)$, les paramètres utiles au recalage du nœud sur l’image de référence. Pour revenir à l’image courante, nous paramétrons ces transformations avec les coordonnées barycentriques généralisées. En effet, elles nous permettent d’exprimer le déplacement non plus par rapport au repère lié à l’image de référence mais par rapport à un repère relatif au support polygonal. Cette paramétrisation permet, sous les hypothèses vues au paragraphe 3.1, de transposer naturellement la situation de l’image de référence à l’image courante.

Dans le paragraphe suivant, nous donnons la définition des coordonnées barycentriques généralisées ainsi que leur utilisation.

5.3 Paramétrisation “affine invariante” des transformations nodales

Coordonnées barycentriques généralisées. *Définition :* Soient P, P_1, \dots, P_k $k + 1$ points du plan avec P_1, \dots, P_k ordonnés dans le sens direct autour de P . Ces k points forment un polygone Π .

On appelle n’importe quel ensemble de fonctions $\lambda_i : \Omega \rightarrow \mathbb{R}$, $i = 1..k$ coordonnées barycentriques si elles satisfont $\forall P \in \Pi$ les trois propriétés :

$$\begin{aligned} (i) \quad & \lambda_i(P) \geq 0, i = 1..k \\ (ii) \quad & \sum_{i=1}^k \lambda_i(P) = 1 \\ (iii) \quad & \sum_{i=1}^k \lambda_i(P) P_i = P \end{aligned} \quad (4)$$

Dans le plan, nous parlons de coordonnées barycentriques généralisées (abrégées en cbg dans la suite), lorsque le nombre de sommets du polygone est supérieur à trois. Il n’y a pas unicité des cbg. Wachspress [17], Floater [5], Meyer et al. [12] se sont intéressés à la définition de telles coordonnées. Toutes ces coordonnées ont des propriétés différentes et particulières. Parmi ces définitions, nous avons choisi les coordonnées de Wachspress [17, 12] pour leurs propriétés : elles sont polynomiales rationnelles de degré minimum et invariantes par transformation affine. Par contre, elles ne sont définies que pour des polygones convexes. Dans notre cas d’utilisation, lors de l’évolution du maillage, nous ne pouvons pas toujours assurer cette contrainte, il faudra donc prévoir comment le traiter.

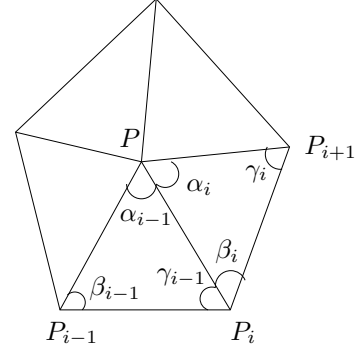


FIG. 5 – Notations pour les coordonnées barycentriques généralisées.

Les k coordonnées barycentriques généralisées de Wachspress associées à un point P appartenant au polygone support du nœud n ont pour expression (pour les notations, voir la figure 5):

$$\lambda(n,i) = \frac{\cotg(\gamma_{i-1}) + \cotg(\beta_i)}{\|P - P_i\|^2}, i = 1..k.$$

Où et comment sont-elles utilisées ? Les cbg de Wachspress étant invariantes par transformation affine, si nous supposons que chaque support polygonal a subi une transformation affine entre l’image de référence et l’image I_{t-1} , alors nous pouvons utiliser les cbg pour exprimer la transformation globale du polygone entre I_0 et I_{t-1} . Les cbg du déplacement cartésien exprimées dans l’image de référence seront les mêmes que celles dans l’image I_t . Nous pourrions donc les traduire en déplacement cartésien dans l’image courante.

Connaissant la position des nœuds du maillage sur l’image de référence I_0 et le déplacement cartésien $\Delta \mu_{cart} = d_{ref}$ estimé dans l’image de référence par le suivi, nous pouvons donc exprimer ce déplacement d_{ref} en cbg : $d_{ref} = \Delta \mu_{cbg} = (\lambda_{n,i})_{i \in 1..K}$, où K est le nombre de sommets du support polygonal. Ces sommets doivent être ordonnés. En supposant que le support a subi une transformation affine, alors les paramètres de la transformation exprimés en coordonnées barycentriques généralisées sont les mêmes sur l’image I_t . D’après cette connaissance et celle de la position des nœuds du maillage sur I_{t-1} , nous pouvons calculer le déplacement cartésien d permettant de recalculer le maillage de l’image I_{t-1} sur l’image I_t .

5.4 Apprentissage et suivi d’une déformation quelconque d’un maillage

Nous venons de voir comment faire pour apprendre et suivre une transformation nodale. Nous allons maintenant généraliser ceci au cas de la déformation d’un maillage en utilisant la relaxation.

L’apprentissage. Pour prendre en compte tous les nœuds, pendant l’étape d’apprentissage, une matrice A_n d’apprentissage par nœud n du maillage est estimée.

Le suivi. Pendant l'étape de suivi entre deux images I_{t-1} et I_t , la connaissance de la position du maillage, donc du motif sur I_{t-1} , et des matrices d'apprentissage $\{A_n\}_{n \in 1..N}$ en chacun des nœuds du maillage suffit pour estimer le mouvement. Le calcul des différences de niveaux de gris $\Delta I(n)$ aux points des motifs nodaux sur les images I_0 et I_t (c'est-à-dire $\Delta I(n) = [I_t(f(p; \mu_{t-1}) - I_0(p))]_{p \in \Omega_n}$), la multiplication "matrice-vecteur" $A_n \Delta I(n)$ et l'utilisation des cbg donnent les paramètres de la transformation utiles au recalage d'un nœud sur l'image I_t .

Pour prendre en compte tous les nœuds, un algorithme de relaxation est mis en œuvre (voir ci-dessous).

Algorithme de relaxation :

- Pour chaque nœud n ,
- + Calculer $\Delta I(n)$
- + Déterminer $\mu(n)$, en utilisant les cbg
- + Calculer le critère E_t
- Fin Pour
- Traiter le nœud n qui fait le plus diminuer E_t
- Tant que le gain sur E_t est inférieur à un seuil faire
- + Mettre à jour :
 - les points d'intérêt du polygone support du nœud n
 - les vecteurs déplacement des nœuds du polygone support et du nœud n
 - le critère global des nœuds du polygone support et du nœud n
- + Traiter le nœud n qui fait le plus diminuer E_t
- Fin Tant Que

5.5 Restrictions

Actuellement, nous ne nous intéressons qu'aux nœuds internes N_{int} du maillage. Nous supposons connus et imposons les déplacements des nœuds du bord N_{bord} du maillage. Nous pourrions aussi, par exemple, les estimer par une méthode d'appariement par bloc ou hexagonal.

6 Validations locales et globale

6.1 Protocole expérimental

Dans le schéma de validation, nous nous intéressons tout d'abord à une validation locale de la méthode puis à une validation globale.

Nous avons travaillé sur les deux séquences vidéo "Miss America" et "Foreman" ainsi que sur des séquences vidéo de synthèse construites à partir de deux images de référence : la première image de la séquence "Miss America" ainsi que celle de la séquence "Foreman". La figure 6 montre ces deux images avec leur maillage initial associé. Le motif d'échantillonnage est formé de points régulièrement répartis autour des nœuds du maillage (voir figure 4).

Dans ces séquences de synthèse, nous faisons aussi l'hypothèse que nous connaissons parfaitement le déplacement des nœuds du bord du maillage, donc qu'ils sont parfaitement recalés d'une image à l'autre.

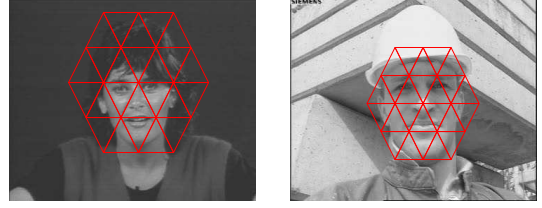


FIG. 6 – Les maillages considérés : à gauche, sur "Miss America"; à droite, sur "Foreman".

Nous comparons les deux apprentissages : l'apprentissage naïf qui consiste à déformer le motif lors de l'apprentissage par l'opposé du déplacement qu'on veut suivre, c'est-à-dire par T_{naive}^{-1} et l'apprentissage exact qui consiste à le déformer par la transformation nodale inverse T_{nod}^{-1} (voir figure 3).

Nous avons testé ces apprentissages sur différentes séquences vidéo de synthèse construites à partir de l'une des deux images de référence. Ces différentes vidéos sont composées d'une succession d'images ayant subi des transformations nodales ou globales (translation, rotation, zoom) ainsi qu'éventuellement d'un mélange ou d'une composition de celles-ci.

Notre critère de comparaison est le nombre d'itérations et l'erreur quadratique entre le déplacement trouvé et à trouver.

Les déplacements appris ont des amplitudes comprises entre -10 et 10 pixels dans toutes les directions.

6.2 Validation locale

Suivi d'une transformation nodale exacte. L'image de référence est l'image de "Miss America". Dans le cadre de cette validation locale, nous supposons que le maillage est parfaitement recalé sur l'image I_{t-1} et que l'image I_t est le résultat de la déformation de l'image I_{t-1} par une transformation nodale à un nœud n .

Nous avons effectué trois séries d'expériences. Dans l'expérience 1, nous supposons qu'entre l'image de référence I_0 et l'image I_{t-1} , le support polygonal et l'image se sont exactement transformés par une transformation affine. Dans l'expérience 2, nous perturbons plus ou moins fortement la position du nœud n à partir de sa position définie par la transformation affine. Enfin, dans l'expérience 3, ce sont les nœuds du support polygonal que nous perturbons.

Le tableau 1 illustre les résultats de la première expérience et nous permet de conclure que l'apprentissage exact donne des résultats très proches de la solution dès la première itération et que, généralement, une seule itération suffit. Ce n'est pas le cas de l'apprentissage naïf. L'implémentation de l'apprentissage naïf est plus simple que celui de l'apprentissage exact, mais l'apprentissage naïf demande plus d'itérations lors du suivi.

Les tableaux 2 et 3 montrent que plus la perturbation de la position du nœud n ou des nœuds du support polygonal est forte, plus le déplacement trouvé à la première itération est

loin du déplacement à trouver. Ils montrent aussi que si le déplacement est petit, l'utilisation des cbg reste valable. Ces expériences montrent aussi que lorsque le maillage est trop déformé (ce qui est le cas dans les grosses perturbations), il serait peut-être utile de réapprendre.

		app. naïf	app. exact
rotation	nb_ité	7	1
	errdepl_l	0.1515	0.0025
	errdepl_fin	0.0063	0.0025
zoom	nb_ité	2	1
	errdepl_l	0.3898	0.0088
	errdepl_fin	0.0464	0.0088
Taffine	nb_ité	3	3
	errdepl_l	0.2846	0.0270
	errdepl_fin	0.0086	0.0083

TAB. 1 – Expérience 1: Tableau donnant le nombre d'itérations (nb_ité) et les erreurs quadratiques entre les déplacements trouvés et à trouver à la première itération (errdepl_l) et à la dernière (errdepl_fin). Comparaison entre l'apprentissage naïf (app. naïf) et l'apprentissage exact (app. exact) dans le cas où, entre l'image de référence I_0 et l'image I_{t-1} , le support polygonal a subi une transformation affine (rotation de 75 degrés, zoom de 0.9 et Taffine: une combinaison de translation, rotation et zoom). Entre I_{t-1} et I_t l'image s'est déformée selon une transformation nodale.

		perturbation		
		petite	moyenne	grande
Taffine	nb_ité	4	7	4
	errdepl_l	0.0279	0.1773	2.4175
	errdepl_fin	0.0107	0.0201	0.0089

TAB. 2 – Expérience 2: Entre I_0 et I_{t-1} , le support polygonal s'est déformé de façon affine par la transformation Taffine et la position du noeud n de la transformation nodale est perturbée de l'ordre d'un pixel dans le cas d'une petite perturbation, de 5 pixels pour une moyenne et de 30 pixels pour une grande.

Suivi d'une transformation globale. Dans le cadre de cette validation locale, l'image I_{t-1} est déformé par une transformation affine globale. Nous faisons l'hypothèse que le maillage est parfaitement recalé sur l'image I_{t-1} et que tous les nœuds sont parfaitement recalés sur l'image I_t sauf un, le nœud n par exemple. Nous cherchons donc à le recalculer. Nous nous situons donc dans un cas similaire à l'expérience 2.

Le tableau 4 permet de comparer nos deux apprentissages entre eux et avec la résolution directe du problème non-linéaire obtenu par l'utilisation de la fonction "lsqnonlin" de MATLAB qui utilise l'algorithme de Levenberg-Marquardt.

Cette expérience montre que l'apprentissage exact donne

		perturbation		
		petite	moyenne	grande
Taffine	nb_ité	2	8	5
	errdepl_l	0.0452	0.1610	1.0254
	errdepl_fin	0.0361	0.0130	0.0057

TAB. 3 – Expérience 3: Entre I_0 et I_{t-1} , les nœuds du support polygonal ont été tout d'abord transformés par une transformation affine Taffine puis perturbés plus ou moins fortement selon une loi uniforme bidimensionnelle définie sur un disque, centré sur chaque nœud, de rayon 1 pixel pour la petite perturbation, 5 pixels pour la moyenne et 20 pixels pour la grande.

de meilleurs résultats que l'apprentissage naïf et proche de la résolution non-linéaire.

Méthodes	tgr (f8,n4)	tgrz (f9,n5)	tgz (f5,n3)
app. naïf	0.0158	0.0721	0.1040
app. exact	0.0004	0.0411	0.0846
non-linéaire	0.0053	0.0580	0.0235

TAB. 4 – Erreur quadratique entre les déplacements trouvés et à trouver sur les séquences générées à partir de "Foreman". "tg*" indique le nom de la séquence (tg pour transformations globales puis r pour rotation, t pour translation et z pour zoom. Par exemple, la séquence tgrz est donc une séquence obtenue à partir de translation, rotation et zoom des images.), "fi" la ième image et "ni" le ième nœud. Comparaison entre nos deux apprentissages et la résolution non-linéaire du problème.

6.3 Validation globale

La figure 7 montre des résultats obtenus sur la séquence "Miss America" en utilisant l'appariement hexagonal (à gauche) et notre méthode (à droite) (le mouvement des nœuds du bord n'est pas connu). Pour l'appariement hexagonal, l'étape d'appariement par bloc s'effectue avec un bloc de dimension 14x14 pixels, une fenêtre de recherche de dimension 15x15 pixels et un pas de recherche d'un pixel. L'étape de raffinement s'effectue avec une fenêtre de recherche de dimension 5x5 pixels et un pas de 1/4 de pixel et est itérée 3 fois sur chacun des nœuds du maillage. Pour notre méthode, nous apprenons des mouvements des nœuds dans une fenêtre d'apprentissage de dimension 21x21 pixels.

Le calcul du Pic du Rapport Signal à Bruit (PRSB ou PSNR en anglais), entre l'image reconstruite à partir de la position du maillage estimé et l'image correspondante de la séquence vidéo, permet de quantifier le rendu de l'estimation et de comparer notre approche avec d'autres méthodes, en particulier: l'appariement hexagonal [13], les méthodes d'Altunbasak [1] ou des approches non linéaires. La courbe de la figure 8 permet de comparer notre méthode à

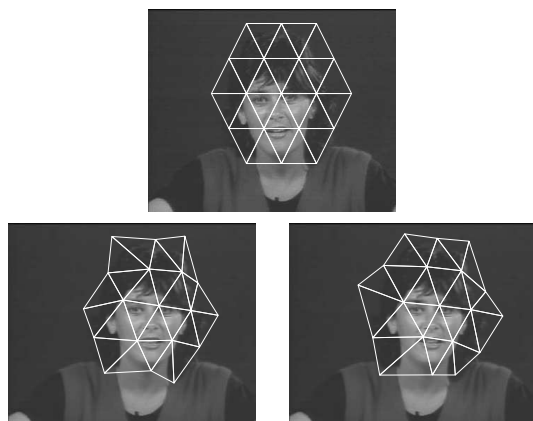


FIG. 7 – Séquence vidéo “Miss America”. En haut : image de référence (image 1) et son maillage d’apprentissage. À gauche : image 78 et son maillage suivi par la méthode d’appariement hexagonal. PRSB 31.41 dB. À droite : image 78 et son maillage suivi par notre méthode. PRSB 30.40 dB.

l’appariement hexagonal sur les 109 images de la séquence vidéo “Miss America”.

7 Conclusion

Notre méthode de suivi par maillage basée sur un apprentissage maintient la connectivité du maillage. Cette méthode donne des résultats similaires à celle de l’appariement hexagonal qui est reconnue comme faisant partie des plus performantes en terme de qualité du résultat [13, 1], tout en étant moins complexe. Nous avons aussi montré dans nos expériences que les hypothèses strictes de l’utilisation des cbg peuvent être relaxées.

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] Y. Altunbasak and A.M. Tekalp, Closed-Form Connectivity-Preserving Solutions for Motion Compensation Using 2-D Meshes, *IEEE Trans. on Image Processing*, vol 6., No 9, 1997
- [2] Y. Altunbasak and A.M. Tekalp, Occlusion-adaptive content-based 2-D mesh design and forward tracking, *IEEE Trans. on Image Processing*, vol. 6, no. 9, 1997.
- [3] F.L. Bookstein, Principal Warps: Thin-Plate Splines and the Decomposition of Deformations, *IEEE Trans. on PAMI*, vol. 11, no 6, juin 1989.
- [4] A. Buisson and N. Laurent, Un codeur hybride H264 - maillage, *Actes de la conférence CORESA’03*, 2003
- [5] M. Floater, Mean value coordinates, *Computer Aided Geometric Design*, Vol.20, pp. 19-27, 2003

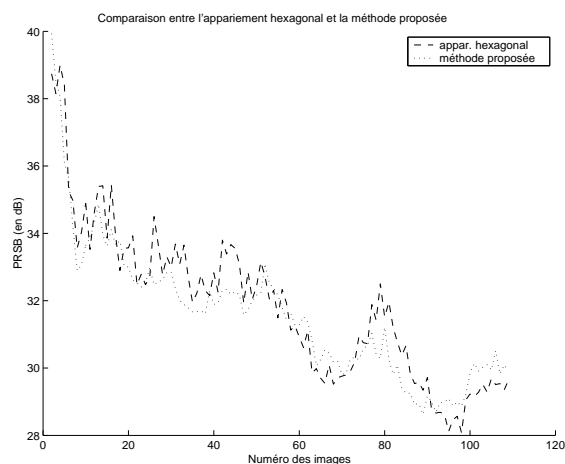


FIG. 8 – Comparaison entre l’appariement hexagonal et la méthode proposée sur la séquence vidéo “Miss America”.

- [6] M. Gleicher, Projective registration with difference decomposition, *Computer Vision and Pattern Recognition*, p. 331-337, 1997.
- [7] G.D. Hager and P.N. Belhumeur, Efficient region tracking with parametric models of geometry and illumination, *IEEE Trans on PAMI*, 20(10),1998.
- [8] F. Jurie and M. Dhome, Hyperplane approximation for template matching *IEEE Trans on PAMI*, 2002.
- [9] P. Lechat, *Représentation et codage de séquences vidéo par maillages 2D déformables*, Thèse de Doctorat, Université de Rennes I, octobre 1999
- [10] G. Marquant, *Représentation par maillage adaptable déformable pour la manipulation et la communication d’objets vidéo*, Thèse de Doctorat, Université de Rennes I, décembre 2000
- [11] G.Marquant, S. Pateux et C. Labit, Mesh-Based Scalable Video Coding with Rate-Distortion Optimization, *VCIP 2000 (Visual Communications and Image Processing 2000)*, Perth, Australia, juin 2000
- [12] M. Meyer and A. Barr and H. Lee and M. Desbrun, Generalized barycentric coordinates on irregular polygons, *Journal of Graphics Tools archive*, Vol. 7, pp. 13-22, 2002
- [13] Y. Nakaya and H. Harashima, Motion compensation based on spatial transformations, *IEEE Trans. on Circuits Systems for Video Technology*, vol. 4, 1994.
- [14] C. Toklu, *Object-based Digital Video Processing using 2-D Meshes* PhD Thesis, Univ of Rochester, USA, 1998.
- [15] S. Valette, I. Magnin and R. Prost, Active mesh for video segmentation and objects tracking, *Proc. IEEE ICIP’2001* (Thessaloniki, Greece), vol II, Oct. 2001.
- [16] P. van Beek, A.M Tekalp, N. Zhuang, I. Celasun and M. Xia, Hierarchical 2-D Mesh representation, tracking and compression for object-based video, *IEEE Trans. on Circuits Systems for Video Technology*, 9(2),1999.
- [17] E. Wachspress, *A rational finite element basis*, Academic Press, 1975
- [18] Y. Wang and O. Lee, Active mesh-a feature seeking and tracking image sequence representation scheme, *IEEE Trans. on Image Processing*, vol. 3, 1994.