

Estimation de Mouvement par Treillis actif: Approche globale et Approche locale

* Marie Dudon, ** Alain Le Méhauté, ** Christian Roux

* CNET (France Télécom) PAB/STC/SGV,

38-40, rue du Général Leclerc, 92131 Issy-les-Moulineaux, France.

** ENST Br., Dép. ITI, BP 832, 29285 Brest CEDEX, France

Email: marie.dudon@issy.cnet.fr

RESUME

Cet article présente une méthode d'estimation de mouvement fondée sur une décomposition triangulaire irrégulière de l'image. Le modèle de mouvement sous-jacent assimile les images à des surfaces déformables. Les images sont représentées par un ensemble de triangles adjacents dont l'évolution individuelle correspond à une transformation affine. On présente et on illustre deux approches, une locale et l'autre globale pour l'estimation des paramètres de mouvement.

1 INTRODUCTION

La plupart des codeurs à bas débits comprennent un algorithme d'estimation et de compensation de mouvement par blocs. Ce type de schéma prédictif produit des distorsions de l'image décodée du type effets de blocs, qui deviennent gênantes dans le cas de mouvements importants ou complexes [1].

La méthode présentée consiste en une estimation de la déformation de la surface de l'image. Pour cela, le modèle sous-jacent assimile la surface de l'image à une grille triangulaire irrégulière dont l'évolution est représentée localement par une transformation affine par morceaux. L'estimation de la déformation est réalisée en deux étapes : l'estimation de la structure du treillis, suivie de celle de la transformation de cette grille.

Différentes approches d'estimation de ces déformations sont développées : des approches globales utilisant soit le formalisme de la méthode des éléments finis soit les algorithmes génétiques ainsi qu'une approche locale récursive.

Les différentes applications de cette méthode sont la prédiction et l'interpolation temporelle d'images dans le cadre du codage d'image, le recalage élastique dans le contexte de l'imagerie médicale, mais aussi l'interpolation spatiale pour l'imagerie stéréoscopique.

ABSTRACT

This article presents a motion estimation algorithm based on a irregular triangular decomposition of the image. The underlying motion model compares images to deformable surfaces. Images are represented by a set of adjacent triangles and each triangle evolution corresponds to an affine transform. Two approaches are presented and illustrated, a local one, and a global one, for motion parameter estimation.

2 MODELE DE DEFORMATION

Définir un modèle de déformation ou de mouvement consiste à formuler des hypothèses sur la nature de cette déformation. Moins ces hypothèses sont restrictives, plus le modèle sera capable d'appréhender des situations de natures différentes. Néanmoins, des hypothèses peu restrictives impliquent généralement une complexité accrue du modèle.

Dans notre cas, nous avons choisi de représenter les déformations par une transformation affine par morceaux (triangles). Ce type de transformation représente un bon compromis entre représentativité et complexité. Il peut en effet rendre compte d'une large gamme de mouvements 2D tels que des translations, rotations, déformations. Néanmoins, des changements dus à des zones de fond découvert ne seront pas appréhendés par ce modèle.

L'évolution de chaque triangle est représentée par une transformation affine. Pour chaque point (x,y) de la région considérée, la position transformée (x',y') correspondante est donnée par la relation suivante :

$$\begin{cases} u(x,y) = x' - x = a_1x + a_2y + a_3 \\ v(x,y) = y' - y = a_4x + a_5y + a_6 \end{cases} \quad (1)$$

où les coefficients a_i sont déterminés pour chaque triangle.

La dynamique des coefficients a_i ainsi que leur quantification étant difficilement maîtrisable, une description



indirecte de ces coefficients peut être obtenue au travers de la description de la position transformée d'un ensemble de sites de référence :

$$\vec{d}(x, y) = \sum_{k=1}^3 \varphi_k(x, y) \vec{d}_k \quad (2)$$

où $\varphi_k(x, y)$ est une fonction d'interpolation linéaire qui dépend des coordonnées des nœuds k .

Dans notre cas, les sites de référence considérés sont les sommets de chaque triangle. Les triangles étant adjacents, la transformation est continue et représentée par les déplacements nodaux $\vec{d}_k = \vec{d}(x_k, y_k)$.

Une fois le modèle choisi, il faut en estimer les paramètres : ceux de la structure et ceux de mouvement de la structure. L'estimation de mouvement consiste à minimiser un critère d'erreur de prédiction. Si l'on choisit comme critère l'erreur quadratique moyenne sur tous les points (x, y) du domaine \mathcal{D} de l'image, il faut alors minimiser :

$$\begin{aligned} E_{\mathcal{D}} &= \sum_{(x, y) \in \mathcal{D}} (DFD(x, y, u, v))^2 \\ &= \sum_{(x, y) \in \mathcal{D}} (I_{n+1}(x + u, y + v) - I_n(x, y))^2 \end{aligned} \quad (3)$$

où $I_n(x, y)$ est la luminance au point (x, y) de l'image n .

On distingue les méthodes globales des méthodes locales de par le fait que la minimisation de l'erreur de prédiction se fait, dans le cas des premières, pour l'ensemble des vecteurs simultanément, et dans le cas des méthodes locales, vecteur après vecteur.

3 GENERATION DE LA GRILLE

De nombreuses techniques de génération de maillage ont été proposées dans la littérature [2]. Le but de telles méthodes consiste à regrouper dans un même élément des points voisins ayant des caractéristiques homogènes.

L'ensemble de ces techniques se divise essentiellement en deux familles :

- d'une part, les techniques fondées sur le contenu qui sont similaires à des processus de division et fusion ("split and merge"). Dans ce cas, la décomposition démarre d'une certaine partition de l'image [3], qui est modifiée tant qu'il existe des zones non homogènes, ou tant que certaines conditions ne sont pas atteintes,

- d'autre part, les techniques par placement de nœuds, où les nœuds sont d'abord détectés, puis sont ensuite reliés pour former une partition triangulaire de l'image.

Dans le contexte du codage d'images, la première famille nécessite la transmission d'informations concernant les décisions de partage et de fusion qui ont lieu au cours du processus. Si les points caractéristiques peuvent être retrouvés au décodeur à partir des informations

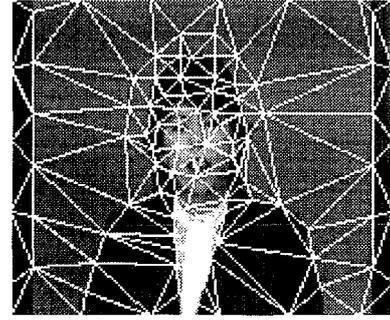


Figure 1: résultat de la triangulation (Claire).

en mémoire, la seconde ne nécessite pas de transmission d'information concernant la position des nœuds.

L'information à représenter étant le mouvement, les nœuds sont placés de préférence à proximité de points ayant des caractéristiques pertinentes pour l'estimation de mouvement, tels que des points de contours et des points d'erreur de prédiction importante.

Le processus de triangulation, quant à lui, peut être choisi de manière à maximiser les angles minimaux (triangulation de Delaunay (Figure 1)), à minimiser l'erreur de prédiction, où plus généralement, à optimiser des caractéristiques de la surface reconstruite.

4 METHODES GLOBALES

Deux méthodes globales d'estimation de mouvement sont présentées, la première basée sur le principe de la sélection naturelle, la seconde utilisant le formalisme des éléments finis.

4.1 Les algorithmes génétiques

Les algorithmes génétiques (AGs) sont des techniques de recherche stochastiques itératives, fondées sur le principe de la sélection naturelle [6]. Ils consistent à faire évoluer une population d'individus représentés par des chromosomes correspondant à des solutions du problème.

Cette évolution (Figure 2) se fait en croisant aléatoirement des couples de chromosomes et en mutant toujours aléatoirement des gènes de certains chromosomes.

Au cours des générations les individus les plus performants sont sélectionnés pour être recombinaisonnés.

La performance d'un individu est déterminée par le biais d'une fonction d'évaluation qui est entièrement liée à la nature du problème à résoudre. Dans le cas de l'estimation de mouvement, la fonction d'évaluation détermine l'erreur quadratique moyenne de prédiction sur un ensemble de points choisis aléatoirement dans l'image. La performance sera d'autant plus élevée que l'erreur de prédiction sera faible.

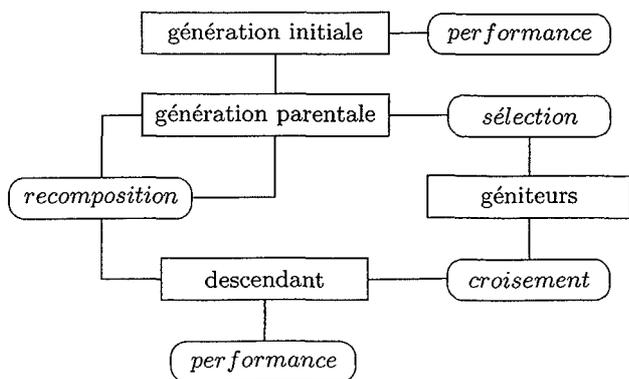


Figure 2: représentation de l'algorithme génétique.

Cette technique s'apparente, dans notre cas, à une approche globale dans la mesure où chaque individu représente un champ de vecteurs complet :

$$\begin{array}{|c|c|c|c|c|c|} \hline u_1 & \dots & u_n & v_1 & \dots & v_n \\ \hline \end{array}$$

Si on considère par exemple, qu'un paramètre est codé sur quatre bits, ce qui correspond à une dynamique de ± 7 pixels, chaque individu ou champ de vecteurs est représenté par une chaîne binaire du type :

$$\begin{array}{|c|c|c|c|c|c|} \hline 0011 & \dots & 0001 & 0000 & \dots & 0110 \\ \hline \end{array}$$

De tels algorithmes permettent de limiter l'ensemble des solutions testées, tout en conservant, grâce aux processus de croisement et de mutation, une probabilité non nulle que chaque solution soit évaluée.

4.2 La méthode des éléments finis

Le modèle discret choisi pour le mouvement convient pour l'utilisation de la méthode des éléments finis (M.E.F.). En effet, le modèle étudié dans le cadre de la M.E.F. est basé sur une subdivision du domaine continu en sous domaines de forme géométrique simple qui constituent les éléments finis [4], interconnectés en des points remarquables : les nœuds. De plus, on définit dans chaque élément une approximation de la solution permettant de résoudre le problème en fonction uniquement des valeurs de la solution aux nœuds. Ces valeurs nodales qui définissent la meilleure approximation sont déterminées en minimisant un critère d'erreur.

En formulant l'hypothèse de continuité et de dérivabilité de la luminance, on peut écrire que pour de faibles déplacements $\vec{d} = (u, v)^T$:

$$I(x+u, y+v, t+\delta t) = I(x, y, t) + u \frac{\partial I(x, y, t)}{\partial x} + v \frac{\partial I(x, y, t)}{\partial y} + \delta t \frac{\partial I(x, y, t)}{\partial t} + \epsilon \quad (4)$$

soit en discrétisant temporellement :

$$I_{n+1}(x+u, y+v) = I_n(x, y)$$

$$+ u \frac{\partial I_n(x, y)}{\partial x} + v \frac{\partial I_n(x, y)}{\partial y} + FD(x, y) + \epsilon \quad (5)$$

$$\text{avec } FD(x, y) = I_{n+1}(x, y) - I_n(x, y).$$

Dans ce cas, l'erreur de prédiction se ramène à :

$$E_{\mathcal{D}} = \sum_{(x, y) \in \mathcal{D}} (u \cdot g_x + v \cdot g_y + FD(x, y))^2$$

$$\text{avec } g_x = \frac{\partial I_n(x, y)}{\partial x}, g_y = \frac{\partial I_n(x, y)}{\partial y},$$

En passant à une partition triangulaire de l'image, et en formulant le vecteur $(u, v)^T$ selon l'équation (2), on obtient :

$$E_{\mathcal{D}} = \sum_{e \in \mathcal{D}} \sum_{\vec{r} \in e} \left(\left(\sum_{k \in \text{som}(e)} \varphi_k^e(\vec{r}) u_k \cdot g_x \right) + \left(\sum_{k \in \text{som}(e)} \varphi_k^e(\vec{r}) v_k \cdot g_y \right) + FD(\vec{r}) \right)^2 \quad (6)$$

où $\text{som}(e)$ correspond à l'ensemble des indices des sommets de l'élément e et $\vec{r} = (x, y)^T$.

La minimisation de l'erreur se fait par l'annulation de sa dérivée par rapport aux paramètres recherchés :

$$\frac{\partial E_{\mathcal{D}}}{\partial u_n} = 0 \quad \text{et} \quad \frac{\partial E_{\mathcal{D}}}{\partial v_n} = 0 \quad \forall n$$

soit pour un nœud n :

$$\sum_{e \in \text{supp}(n)} \sum_{k \in \text{ver}(e)} \begin{pmatrix} G_{xx}^k & G_{xy}^k \\ G_{xy}^k & G_{yy}^k \end{pmatrix} \begin{pmatrix} u_k \\ v_k \end{pmatrix} = - \sum_{e \in \text{supp}(n)} \begin{pmatrix} F_x \\ F_y \end{pmatrix} \quad (7)$$

$$\text{où } \begin{cases} G_{ij}^k = \sum_{\vec{r} \in e} \varphi_k^e(\vec{r}) \varphi_n^e(\vec{r}) g_i g_j \\ F_i = \sum_{\vec{r} \in e} FD(\vec{r}) \varphi_n^e(\vec{r}) g_i \end{cases} \quad i, j \in \{x, y\}$$

et où $\text{supp}(n)$ correspond à l'ensemble des éléments ayant le nœud n comme sommet. Ce système d'équations est valable pour tout nœud n . On aboutit à un système global représenté par une matrice $(2n \times 2n)$.

Le coût en temps de calcul est directement lié au nombre de nœuds n de la triangulation. En effet, la résolution du système passe par l'inversion d'une matrice très creuse de taille $(2n \times 2n)$.

La résolution du système obtenu ne conduit néanmoins pas toujours à une réduction d'erreur de prédiction du fait que les déplacements nodaux à estimer n'ont pas toujours une amplitude faible. Il peut alors être intéressant d'utiliser une approche itérative classique en analyse numérique [5]. Cette approche est en cours de développement.

5 METHODES LOCALES

La méthode suivante d'estimation des déplacements nodaux est une approche locale dans la mesure où les vecteurs sont estimés en minimisant un critère d'erreur sur un ensemble de points défini uniquement par les



nœuds voisins. Dans un premier temps, un vecteur initial est estimé pour chaque nœud grâce à un algorithme d'appariement de blocs, puis l'ensemble des paramètres de la transformation affine est ajusté pour chaque triangle de manière itérative [7] (Figure 3).

Cet ajustement ou raffinement est réalisé en déplaçant chaque nœud tout en maintenant les nœuds voisins fixes. Le déplacement qui minimise l'erreur de prédiction à l'intérieur du polygone défini par les nœuds voisins est ajouté à l'ancien vecteur.

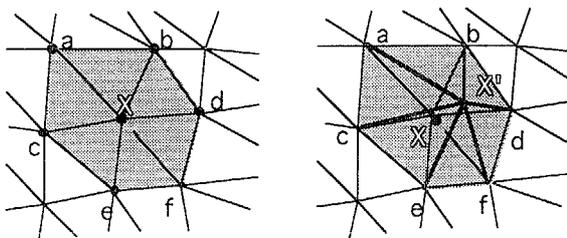


Figure 3: processus de raffinement : X est le nœud à raffiner, X' est sa nouvelle position après raffinement.

Après chaque itération, la différence entre l'image courante et la prédiction décroît ou reste indentique. On constate que l'estimation converge vers un minimum local ou global.

6 RESULTATS EXPERIMENTAUX

Les simulations ont été réalisées à partir de la séquence de test Miss America, au format QCIF (176*144), à la fréquence de 10 Hz. Le nombre de nœuds a été fixé à $n = 100$.

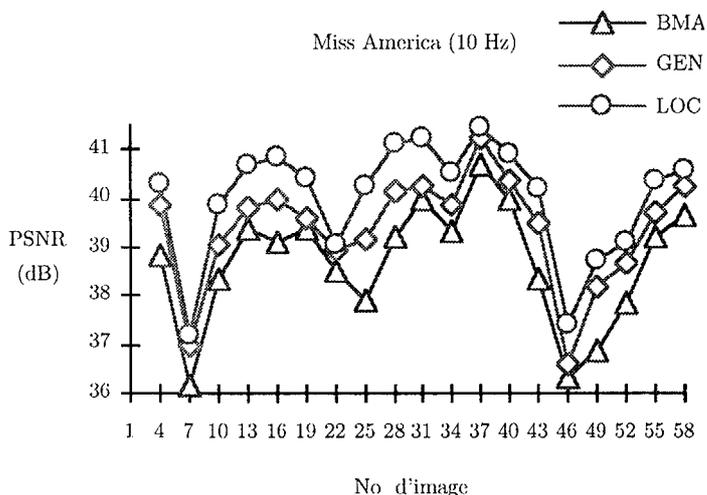


Figure 4: Comparaison d'algorithmes d'estimation des déplacements nodaux : mise en correspondance de blocs (BMA), algorithme génétique (GEN), approche locale (LOC).

Les différentes méthodes sont comparées en terme de qualité de prédiction. L'algorithme génétique a été comparé à l'approche locale itérative pour un coût en temps de calcul équivalent. Dans les deux cas, le champ de

vecteur est initialisé par un champ de déplacements grossiers obtenu grâce à un algorithme de mise en correspondance de blocs. L'étape de raffinement, génétique ou locale, est réalisée pour une dynamique de ± 1 pixel pour chaque paramètre.

On constate (Figure 4) que le champ de vecteurs obtenu par l'approche locale récursive fournit une meilleure prédiction que celui obtenu par l'approche globale génétique. L'approche locale teste pour chaque nœud toutes les nouvelles positions possibles, tandis que l'exploration complète n'est pas assurée dans le cas de l'algorithme génétique. Néanmoins, contrairement aux approches globales, l'approche locale présente l'inconvénient d'être dépendante de l'ordre de balayage des nœuds lors de l'estimation de leur mouvement.

7 CONCLUSION

Différentes méthodes d'estimation de champs de vecteurs ont été proposées pour la représentation de déformations intervenant dans une séquence d'images. Les approches globales permettent de s'affranchir de la dépendance avec l'ordre de considérations des nœuds mais leur convergence s'avèrent moins rapide que celle des méthodes locales.

Les déformations obtenues permettent de réaliser des prédictions ainsi que des rééchantillonnages temporels d'images. D'autre part, il est possible d'envisager dans le cadre du codage, grâce au fait que le nombre de nœuds est paramétrable, une régulation du débit alloué à l'information de mouvement, en faisant varier pour chaque image, le nombre de nœuds du maillage. L'erreur de prédiction croissant lorsque le nombre de nœuds diminue, il peut être intéressant de déterminer le meilleur compromis ou proportion d'information de mouvement à coder par rapport à celle d'erreur de prédiction à un débit donné.

Références

- [1] H.G. Musmann, P. Pirsch, H. J. Grallert, "Advances in picture coding", Proc. IEEE, Vol. 73, p. 523, April 1985.
- [2] K. Ho-Le, "Finite element mesh generation methods: a review and classification", Computer-aided design, Vol. 20, No. 1, jan/feb 1988.
- [3] F. Davoine, J.-M. Chassery, "Compression d'images par fractales", Journées d'études et d'échanges, Rennes, 01/1995, p. 169.
- [4] O. C. Zienkiewicz, "La méthode des éléments finis", 1973, Ediscience, Paris.
- [5] H. Sanson, "Estimation de mouvement basée sur des régions pour le codage de séquences Télévisuelles", 14ème GRETSI, Juan-les-pins, 13-16 sept. 1993.
- [6] J.-J. Jacq, C. Roux, "Recalage monomodalité automatique en imagerie médicale 2D et 3D à l'aide d'un algorithme génétique traditionnel", 9ème congrès de RFIA, AFCET/AFIA, Paris, jan. 1994.
- [7] Y. Nakaya, H. Harashima, "Motion Compensation based on Spatial Transformations", IEEE trans. on CSVT, Vol. 4, No. 3, jun. 1994, p.339.