

Segmentation vidéo et suivi d'objets par maillage déformable : gestion des discontinuités de mouvement

Sébastien Valette, Isabelle Magnin et Rémy Prost

CREATIS, UMR CNRS 5515, affilié à l'INSERM

INSA, Bâtiment Blaise Pascal, Villeurbanne, France

{sebastien.valette, isabelle.magnin, remy.prost}@creatis.insa-lyon.fr

Résumé - Nous proposons un nouvel algorithme de maillage actif qui place les arêtes du maillage sur les discontinuités de mouvement. Le mouvement intérieur à chaque cellule est estimé par mise en compétition des estimations avant et arrière. Cette approche permet le déplacement des nœuds du maillage sans avoir à remailler les zones occultées ou découvertes.

Abstract - In this paper, we propose a novel mesh-based video segmentation and objects tracking algorithm with both robust motion estimation and modeling of motion discontinuities. In addition, occlusions and uncovered regions are well managed. This allows the mesh deformation without remeshing in motion occlusion regions.

1. Introduction

Les futurs services vidéo utiliseront les standards MPEG 4 et MPEG 7. En opposition aux anciens standards qui gèrent les séquences vidéos sans se soucier de leur contenu, ces standards sont basés sur la gestion des objets vidéo (OV). La mise en place de tels standards nécessite donc la création d'outils permettant la segmentation des objets vidéos dans les séquences, créant ainsi des plans d'objets vidéo. Cette opération reste encore un problème ouvert. Les maillages actifs, utilisés initialement pour la compression vidéo à bas débit [1][6][9][10], sont de bons candidats pour cette tâche [2][4][7].

L'étape de segmentation de mouvement consiste en la séparation de chaque image en plusieurs parties, formant ainsi la carte de segmentation qui donne la position spatiale de chaque objet durant la séquence à traiter. L'intervention humaine pour cette opération semble inévitable, car dans le cas général le nombre d'OVs dans une séquence ne peut être déterminé automatiquement. En prenant l'exemple d'un camion se déplaçant sur une route, on peut considérer le camion en tant qu'un seul OV, mais l'arrière plan immobile peut être un seul OV ou peut être découpé en objets plus petits : la route, les arbres, le ciel. Partant de ce constat, nous avons fait l'hypothèse que la segmentation de la première image de la séquence à traiter est semi-automatique ou manuelle.

Des travaux antérieurs ont combiné segmentation spatiale et estimation de mouvement pour effectuer le suivi des OV dans une séquence [5]: chaque image subit une sur-segmentation spatiale, la découpant en petites régions homogènes au sens de la luminance. Ensuite, le mouvement de chaque région par rapport à l'image précédente, est estimé afin de propager la segmentation initiale des OV le long de la séquence. L'estimation peut être de type avant ou arrière.

Dans la plupart des cas, l'estimation de mouvement des différentes régions est effectuée par la méthode de mise en correspondance, basée sur la minimisation d'une erreur de compensation de mouvement entre les images successives. Une difficulté apparaît alors : ce type d'estimation de mouvement conduit à de bons résultats lorsqu'il est appliqué à des régions hétérogènes au sens de la luminance, mais devient moins efficace pour les régions homogènes. Ceci pénalise les méthodes qui effectuent séquentiellement une segmentation spatiale et une estimation de mouvement. Contrairement à ces travaux, notre algorithme ne forme pas les régions en les rendant homogènes en terme de luminance, mais estime le mouvement de cellules pouvant être fortement hétérogènes. Dans la deuxième section nous expliquons le fonctionnement général de l'algorithme. Les deux sections suivantes apportent des détails supplémentaires sur la déformation du maillage et son raffinement. La section 5 illustre cette approche par quelques résultats expérimentaux et la conclusion présente des perspectives d'amélioration de l'algorithme.

2. Création du maillage

Créer un nouveau maillage pour chaque image est inutile : il suffit de le créer sur la première image puis de le déformer et le raffiner (en découpant certaines cellules) au cours de la séquence. Ainsi, nous pouvons considérer l'étape initiale de création du maillage comme un cas particulier du processus de déformation et de raffinement : en commençant avec un maillage grossier et régulier, notre algorithme le déforme pour créer la carte de segmentation i.e. il déplace les nœuds du maillage pour faire coïncider ses arêtes avec les frontières entre les objets. Ensuite les cellules nécessitant plus de détail sont subdivisées dans un schéma quadtree. Ces deux étapes sont répétées jusqu'à ce que la segmentation soit suffisamment précise. La figure 1 illustre la création d'un maillage actif.

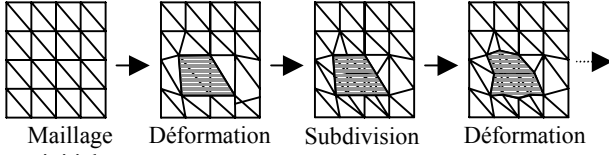


FIG. 1 : création du maillage

Comme le maillage est déformé durant toute la séquence, des remaillages doivent être effectués pour empêcher que certaines cellules soient trop petites ou dégénérées. Deux types de remaillage sont pris en compte ici : la fusion de deux sommets (figure 2.a) ou la permutation d'une arête (figure 2.b).

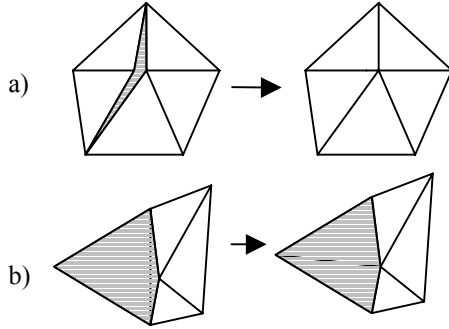


FIG. 2 : remaillage

- a) fusion de sommet
- b) permutation d'une arête

3. Déformation du maillage

3.1 Fonctionnelle d'énergie

L'étape de déformation du maillage est l'étape cruciale de notre algorithme. Nous l'avons élaborée en considérant les observations suivantes :

- Chaque cellule doit contenir une région de l'image homogène au sens du mouvement afin qu'aucune d'entre elles ne contienne deux objets différents.
- Au temps t , la carte de segmentation du temps $t-1$ a déjà été construite. Une projection arrière de chaque cellule du temps t sur cette carte peut déterminer si le maillage a été déformé correctement au regard de la segmentation précédente.
- Les arêtes frontières entre les OV doivent être situées sur des zones à fort gradient spatial.

En prenant en compte ces trois observations, le maillage est déformé de manière à minimiser l'énergie E définie ainsi:

$$E(t) = E_m(t) + \alpha.E_b(t) + \beta.E_s(t) \quad (1)$$

Cette énergie globale se décompose en trois termes : l'énergie de mouvement $E_m(t)$, l'énergie de projection arrière $E_b(t)$ et l'énergie spatiale de frontière $E_s(t)$, les paramètres α et β permettent l'équilibrage entre ces énergies.

3.2 Energie de mouvement

L'énergie de mouvement est définie par :

$$E_m(t) = \frac{\sum_i \left(\sum_{p \in c_i} [i(\tau_i(p), t + d_i) - i(p, t)]^2 \right)}{\sum_i a_i} \quad (2)$$

où $i(p, t)$ est l'intensité du pixel p au temps t et a_i est la surface de la cellule c_i .

$E_m(t)$ est l'erreur quadratique de compensation de mouvement pour l'image t , avec la particularité que le mouvement de chaque cellule c_i est estimé par la méthode avant ou arrière, afin de diminuer les problèmes d'estimation dans les zones occultées ou découvertes. Une zone occultée conviendra mieux à une estimation arrière, tandis qu'une zone découverte verra son mouvement correctement estimé uniquement par estimation avant.

La comparaison entre les deux erreurs de compensation donnera la direction d_i de l'estimation de mouvement de la cellule considérée. Si la plus petite erreur est obtenue par estimation arrière alors $d_i = -1$ et $\tau_i(p) = \tau_i^-(p)$, sinon $d_i = 1$ et $\tau_i(p) = \tau_i^+(p)$. Le modèle de mouvement choisi est associé à τ_i . Nous avons expérimenté plusieurs modèles de mouvement :

- Le modèle de translation (2 paramètres par cellule) : on considère uniquement un mouvement de translation des cellules.
- Un modèle affine (6 paramètres par cellule) : un vecteur déplacement est affecté à chacun des trois nœuds de la cellule. Chaque nœud se verra donc attribuer autant de vecteurs déplacements qu'il aura de cellules adjacentes.
- Un modèle affine contraint (6 paramètres par cellule), où les discontinuités de mouvement sont autorisées uniquement aux frontières entre les objets. Cette restriction permet une estimation plus rapide et plus robuste, en faisant l'hypothèse que le mouvement à l'intérieur d'un même objet doit être continu. Il s'ensuit que les nœuds à l'intérieur d'un objet auront un seul vecteur déplacement, alors que les nœuds placés sur les frontières entre les OV en auront deux.
- Un modèle de mouvement représenté par un maillage différent pour chaque objet et estimant aussi un champ continu à l'intérieur des objets, méthode similaire à celle décrite dans [8].

Ce terme d'énergie $E_m(t)$ tendra à rendre les cellules homogènes au sens du mouvement, prenant en compte la gestion des discontinuités de mouvement, comme décrit en [3].

3.3 Energie de projection arrière:

L'énergie de projection arrière est définie par :

$$E_b(t) = \frac{\sum_i m_i}{\sum_i a_i} \quad (3)$$

où m_i est le nombre de pixels mal classés.

Chaque cellule c_i du maillage au temps t est projetée sur la carte de segmentation créée au temps $t-1$ en utilisant la transformation $\tau_i^-(p)$ calculée lors de l'estimation de mouvement. La cellule considérée est associée à l'OV sur lequel elle est projetée. Ensuite, le nombre de pixels mal classés m_i , les pixels qui ont été projetés en dehors de l'objet associé, est calculé. Par exemple, la cellule c_1 de la figure 3 est associée à l'objet 1, et m_1 est égal au nombre de pixels contenus dans la région grisée R.

Cette énergie ajoute une contrainte de régularisation temporelle sur la segmentation.

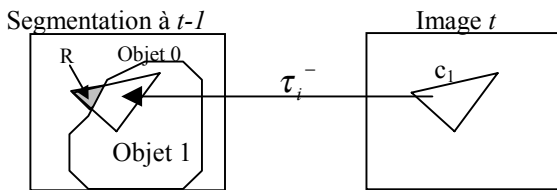


Fig. 3 : calcul de l'énergie de projection arrière

3.4 Energie spatiale de frontière

L'énergie spatiale de frontière est définie par :

$$E_s(t) = \frac{\sum_j \left(r(e_j) \sum_{p \in e_j} dist(p) \right)}{\sum_j l_j} \quad (4)$$

où :

- l_j est la longueur de l'arête e_j .
- $dist(p)$ est la distance du point p au contour détecté le plus proche. Le contour est détecté par un seuillage sur le gradient de l'image en cours (figure 4.a). Une carte de distance est alors créée (figure 4.b)
- $r(e_j) = 1$ si l'arête e_j constitue une frontière entre deux OV. Sinon, $r(e_j) = 0$.

$E_s(t)$ est une énergie de type contour actif, contraignant le contour de chaque OV segmenté à se trouver sur les zones de l'image à fort gradient spatial.

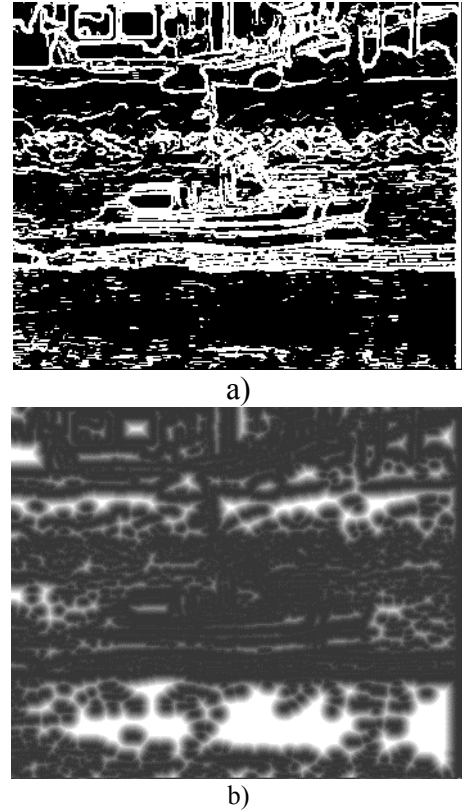


FIG. 4 : carte de distance au contour le plus proche

4. Etape de découpage

Une fois que le maillage est déformé, chaque cellule subit un test : pour une cellule donnée c_i , si le nombre de pixels mal classés lors de la projection arrière m_i est plus grand qu'un seuil fixé, alors la cellule doit être subdivisée.

Dans l'exemple de la figure 3, la cellule c_1 a été associée à l'OV 1 mais sa partie grisée projetée sur l'objet 0 est grande (m_1 grand), et la cellule sera subdivisée. Ce critère de découpage subdivise donc les cellules en fonction de la précision de la segmentation actuelle vis-à-vis de la segmentation effectuée sur l'image précédente.

5. Resultats

La figure 5 montre les résultats obtenus sur la séquence « Coastguard ». Le modèle de mouvement utilisé emploie un maillage mouvement par objet. L'image 5.a est le masque de segmentation initial créé sur l'image 100. Les images 5.b à 5.h montrent les segmentations suivantes sur les images 101,103,105,107,109,111,113. Les contours des deux bateaux sont assez bien segmentés, mais on peut remarquer une imprécision dans les zones découvertes (derrière les deux bateaux).

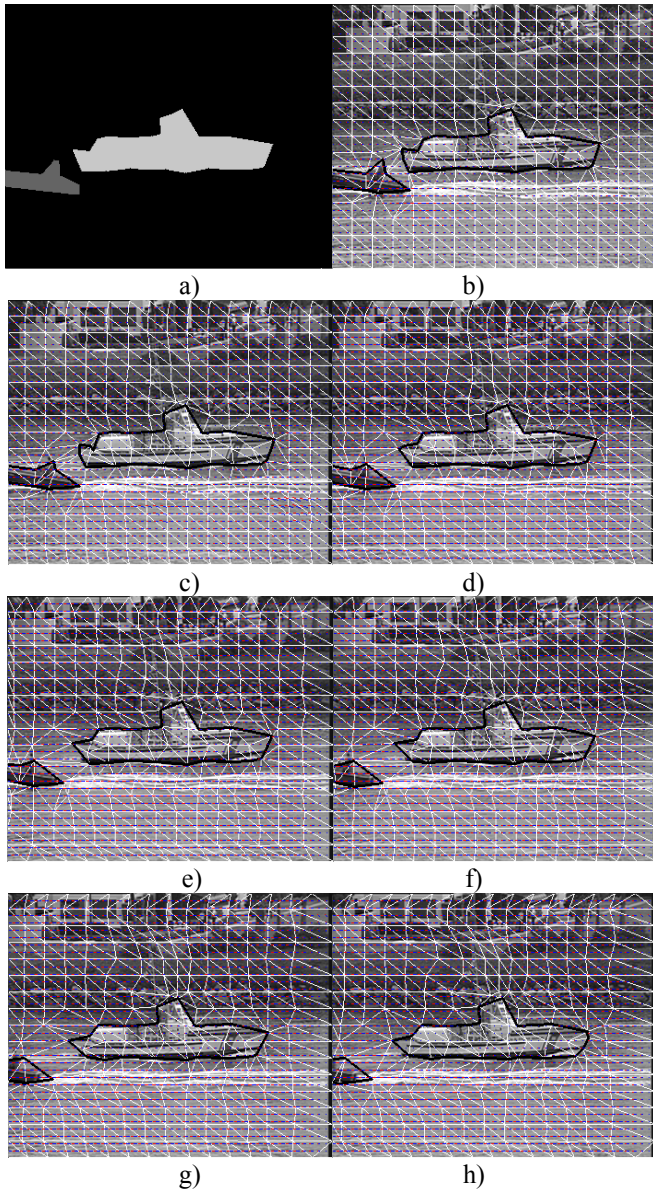


FIG. 5 : résultats sur « coastguard »

6. CONCLUSION

Nous avons proposé un algorithme de construction d'un maillage actif pour la segmentation et le suivi d'objets dans des séquences vidéo. La modélisation utilisée prend en compte les discontinuités de mouvement. Une contrainte spatiale force les frontières des objets segmentés à rester dans les zones à fort gradient, et une contrainte de régularisation temporelle est mise en oeuvre.

Les évolutions envisagées de cette approche concernent: l'utilisation d'une méthode de détection de contours plus évoluée, et une amélioration du traitement des zones découvertes.

REMERCIEMENTS :

Ce travail a été réalisé dans le cadre du projet RNRT OSIAM (Outils de Segmentation d'Images Animées pour MPEG 4-7).

Ce travail de recherche s'inscrit dans les thématiques scientifiques du PRC-GDR ISIS.

- [1] Y. Altunbasak and A. M. Tekalp, Closed-form connectivity-preserving solutions for motion compensation using 2-D meshes, *IEEE Trans. Image Processing*, vol. 6, no 9, pp. 1255-1269, September 1997.
- [2] Y. Altunbasak and A. M. Tekalp, Occlusion-adaptive, content-based mesh design and forward tracking, *IEEE Trans. Image Processing*, vol. 6, no 9, pp. 1270-1280, September 1997.
- [3] A. Chretien-Planat, Estimation de mouvement par maillage actif multiechelle avec prise en compte des discontinuités : Application à l'imagerie cardiaque en Résonance Magnétique, PhD Thesis, no 99ISAL0011, INSA, Lyon, France, 1999.
- [4] M. H. Gökçetekin, M. D. Harmanci, I. Celasun and A. Murat Tekalp, Mesh based segmentation and update for object based video, *IEEE International Conference on Image Processing ICIP 2000*, Vancouver, Canada, October 2000
- [5] G. Gu and M.C. Lee, Semantic video object tracking using region-based classification. *IEEE International Conference on Image Processing ICIP'98*, pp. 643-647, Chicago, USA, 1998
- [6] L. Huang and C.-Y. Hsu, "A new motion compensation method for image sequence coding using hierarchical grid interpolation", *IEEE Trans. on Circuits and Systems for Video Technol.*, vol. 4, pp 72-85, 1994.
- [7] P. Lechat, N. Laurent, H. Sanson, Suivi d'Objets Vidéo par Maillages Hiérarchiques, *GRETSI 1999*, pp. 507-510, Vannes, 13-17 septembre 1999
- [8] G. Marquant, S. Pateux and C. Labit -- Mesh and "Crack lines": application to object-based motion estimation and higher scalability. – *IEEE International Conference on Image Processing ICIP 2000*, Vancouver, Canada, September 2000
- [9] Y. Wang and O. Lee, Active mesh – a feature seeking and tracking image video sequence representation scheme, *IEEE Trans. Image Processing*, vol 3, pp 610-624, september 1994.
- [10] Y. Wang and O. Lee, Use of two-dimensional deformable mesh structures for video coding, part II – the analysis problem and a region-based coder employing an active mesh representation, *IEEE Trans. on Circuits and Systems for Video Technol.*, vol. 6, december 1996, pp. 647-659.