Dix-septième colloque GRETSI, Vannes, 13-17 septembre 1999

Suivi d'Objets Vidéo par Maillage Hiérarchique

Patrick LECHAT, Nathalie LAURENT, Henri SANSON

France Télécom - CNET / DIH - CCETT
4, rue du Clos Courtel - 35512 Cesson Sévigné Cedex - FRANCE
{prénom.nom}@cnet.francetelecom.fr

Résumé – La représentation d'objets vidéo par maillages triangulaires a récemment été ajoutée à la norme MPEG-4. Les fonctionnalités attendues sont nombreuses : compression, interpolation spatio-temporelle, réalité augmentée, animation et déformation de textures, indexation basée contenu. Pour qu'elles soient efficaces, des techniques performantes d'estimation de mouvement doivent être mises en œuvre, associées à maillages adaptés. Cet article expose un procédé de suivi d'objets vidéo maillés à partir d'une unique segmentation initiale. S'appuyant sur une représentation par maillage hiérarchique, nous allons successivement estimer le mouvement dominant de l'objet au moyen d'un modèle global affine, puis estimer les mouvements locaux de chacun des noeuds du maillage. Cette dernière étape exploite complètement le caractère hiérarchique de la méthode dans le sens qu'autant d'estimations successives que de niveaux de maillages sont réalisées. Le résultat est un suivi efficace à la fois des mouvement globaux et locaux, ne s'appuyant pas sur les *alpha*-plans des séquences MPEG-4, ni sur une segmentation spatiale associée. Le procédé d'estimation de mouvement est de type différentiel, basé sur la minimisation de la fonction quadratique d'erreur de compensation.

Abstract – Triangular mesh-based representation has recently been added to the MPEG-4 standard which deals with object-based compression and composition. Many functionalities are expected : compression, spatio-temporal interpolation, augmented reality, warping, transfiguration and content-based retrieval of video-objects. To perform them accurately, efficient motion techniques must be employed with adapted meshes. This paper presents a method for tracking video objects with meshes, starting from a single initial segmentation. Our method is based on a mesh hierarchy with a coarse to fine motion estimation : first, coarse mesh motion vectors are initialized with a global 6-parameters affine motion model, then node-based differential motion estimation is performed on successive mesh levels. Result is an efficient tracking for coarse and fine displacements which does not involve neither MPEG-4 *alpha*-planes sequences or any spatial segmentation.

1 Introduction

Les techniques de compression standard (MPEG-1/2) ne permettent pas d'interactivité basé contenu, dues à leur principe de compression trame à trame. Cependant les développements multimédias interactifs récents ont montré la nécessité de manipuler, compresser, indexer les objets vidéos qu'ils soient d'origine naturelle ou synthétique. Développer des outils permettant cette représentation en objets ayant une signification sémantique devient donc indispensable.

Pour représenter de tels objets, leur forme doit être codée au même titre que la texture. Les plans de tranparence ou *alpha*-plans du standard MPEG-4 [5] ont donc été définis et d'importants travaux ont été publiés pour les coder [6] ; en particulier une bibliographie conséquente traite du codage avec ou sans pertes de l'information de contour et texture. Le codage de tels plans de transparence représente un surcoût important aux techniques classiques basées trames. Deux principes se différencient suivant le caractère transparent ou opaque des objets de la scène : l'approche par plan de bits et celle basé contours. Associées à celles-ci, des méthodes de compression interimages exploitant la redondance temporelle de cette information de forme ont été développées, reprenant le principe de l'estimation / compensation des schémas traditionnels de codage, ou bien s'intéressant à la déformation de contours. Le travail présenté dans cet article se situe dans ce dernier contexte : la forme de l'objet est représentée par un maillage dont les arcs extérieurs décrivent au mieux le contour réel. Ce maillage, constitué à l'aide d'une segmentation initiale, est déformé avec un coût minimal au cours de la séquence vidéo, s'adaptant le plus précisément possible aux déformations réelles de l'objet vidéo. La texture associée au maillage initial peut ainsi être déformée via un modèle affine dans notre cas, capable de modéliser aussi bien des translations que des rotations ou zooms.

Traditionnellement, les *alpha*-plans sont obtenus par la technique du *chroma-key* permettant une segmentation précise et rapide mais au pris de conditions de prise de vue particulières (fond bleu). Pour des sources vidéo en mode trame, des procédés d'extraction automatique par segmentation spatio-temporelle ont été proposées, mais la difficulté d'extraire automatiquement des objets significatifs d'un point de vue sémantique est toujours d'actualité [10]. La technique de représentation d'objets par maillage 2D est une alternative semi-automatique à ce problème : partant d'une segmentation manuelle, pouvant être assistée par une segmentation spatiale, il est possible d'identifier des objets d'intérêt, le suivi automatique de ceux-ci permettra ainsi leur extraction jusqu'à leur disparition de la scène.

Au cours de cet article, nous allons décrire le modèle et la méthode d'optimisation du champ de mouvement basé maillage hiérarchique, situant nos travaux par rapport à ceux du domaine. Nous verrons ensuite plus spécifiquement l'originalité de notre procédé de génération du maillage ainsi que l'initialisation par une méthode globale robuste.

2 Estimation du mouvement et méthode de suivi

Le modèle de mouvement retenu est un modèle affine à six paramètres (1), permettant de représenter efficacement et continuement la majorité des déformations réelles présentes dans les séquences vidéo. Appliqué au cadre des maillages triangulaires, ce modèle est strictement équivalent à affecter un vecteur mouvement par nœud du maillage. On peut alors interpoler le mouvement intérieur $\vec{d}$ de composantes (x',y') au point $p(x,y)$ pour chaque maille e grâce aux coordonnées barycentriques $\psi_n^e(p)$ (2) associées aux nœuds i,j,k de e [7] (figure 1).

$$\begin{cases} x' = a.x + b.y + c \\ y' = d.x + e.y + f \end{cases} \quad (1)$$

$$\vec{d}(p) = \sum_{n=i,j,k} \psi_n^e(p).\vec{d}_n \quad (2)$$

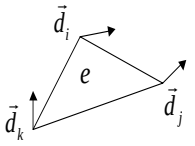

FIG. 1: *Estimation affine sur maille triangulaire*

Pour réussir à modéliser avec précision les mouvements, il est nécessaire d'utiliser des mailles de faible taille. Cependant un tel choix se fait au détriment de la robustesse de la méthode, et surtout de son incapacité à estimer de forts mouvements [2]. Une approche multirésolution a donc été appliquée, s'appuyant sur des travaux déjà publiés [7]. Elle consiste à définir une hiérarchie de maillages pour lequels le champ de mouvement sera successivement raffiné, partant d'un maillage grossier estimant les mouvements importants, puis propageant ce mouvement sur des niveaux de mailles plus fines permettant d'estimer le mouvement local. La méthode d'estimation est de type différentielle s'appuyant sur la méthode d'optimisation de *Levenberg-Marquardt*, extension de celle de *Gauss-Newton*.

Cette approche hiérarchique est aussi celle retenue dans [1], à la différence que *P. van Beek* différencie les méthodes d'estimation de mouvement pour le maillage grossier (basée gradient), et pour ses sous-divisions successives (recherche hexagonale généralisée, correspondant à une relaxation du vecteur mouvement autour de sa valeur estimée).

Il faut cependant remarquer que deux types de suivi peuvent être envisagés, suivant que l'on dispose du masque de segmentation tout au long de la séquence (obtenu par une acquisition *chroma-key*) ou non. Dans le premier cas, il est souhaitable que le déplacement des noeuds du contour de l'objet soient contraints à rester sur la frontière connue du VOP (*Video object plane)* [10]. Les mouvements des noeuds intérieurs à la forme sont alors soit déduit de ceux du contour, soit estimés indépendamment de la frontière de l'objet.

Comme précisé en introduction, le sujet de ce travail est de réaliser un suivi avec une unique segmentation initiale. Afin de conserver l'acuité du découpage, il est nécessaire d'avoir un suivi efficace. Certains travaux se basent sur des techniques de contours actifs [4], ou s'appuyent sur une segmentation spatiale. Dans notre cas, le suivi est réalisé sur la seule estimation du mouvement, exploitant la structure hiérarchique du maillage et simplifiant le processus de raffinement des vecteurs de mouvement nodaux.

3 Création de la hiérarchie de maillages

La réalisation du maillage est une étape particulièrement délicate. En effet, de celle-ci dépendra pour une large part l'efficacité de l'estimation de mouvement. D'une manière générale, ce maillage est obtenu grâce à une triangulation contrainte, imposant le contour des objets de la scène. Très classiquement, la triangulation utilisée est la triangulation de *Delaunay* contrainte, celle-ci maximisant les rapports d'aspect (ou compacité) des triangles, menant ainsi à de bonnes propriétés de conditionnement. Le contour est quant à lui obtenu par approximation polygonale de la frontière réelle du *VOP* [6].

Deux approches coexistent pour générer la hiérarchie de maillage : celle grossier vers fin, ou au contraire fin vers grossier. La première stratégie consiste à diviser les mailles selon un critère prédéfini, la seconde sélectionne les noeuds à conserver à un niveau de maillage et réalise une triangulation avec ces noeuds pour générer le maillage décimé. *Celasun & al.* [3] sélectionnent les nœuds intérieurs grâce à un critère "d'importance" fonction du gradient local de l'intensité de l'image et de l'efficacité de la compensation de mouvement caractérisée par la DFD.

Notre approche se différencie de celles exposées en ce sens qu'aucune approximation de contour polygonal n'est réalisée. Il s'agit d'une méthode fin vers grossier où le maillage le plus fin est de structure régulière. Pratiquement, un maillage fin et global à toute l'image est appliqué, les mailles n'appartenant pas à des objets sont tour à tour supprimées et les arcs formant les contours des *VOPs* sont identifiés. Un sous-échantillonnage du maillage est alors réalisé pour obtenir la hiérarchie telle que celle présentée en figure (4).

Il est à noter l'approximation très grossière des contours réels des objets pour la figure (4-a), pouvant sembler préjudiciable à une estimation de mouvement dans le sens où une partie non négligeable de la surface définie par éléments triangulaires n'appartient pas aux objets eux-même. Notre implémentation contourne ce problème en utilisant le masque binaire du *VOP* associé au niveau de maillage le plus fin (4-c) lors de toutes les optimisations de mouvement. Seuls les pixels appartenant à masque sont considérés lors de l'estimation du mouvement.

4 Initialisation par modèle polynomial global

Bien que l'estimation de mouvement hiérarchique apporte un avantage certain à la robustesse de l'estimation de mouvement ([1], [7]), une étape préalable nous a apporté des résultats encore plus prometteurs. Celle-ci consiste à modéliser le mouvement de l'objet entier défini par son masque de segmentation par un unique modèle affine à 6 paramètres. Nous avons donc utilisé les travaux développés dans [9] s'appliquant à l'analyse du mouvement de régions quelconques par modèles polynomiaux. S'appuyant sur le même procédé d'optimisation différentielle que notre méthode basée maillage [7], ce premier traitement permet l'obtention de vecteurs mouvement nodaux initiaux homogènes, appliqués au maillage grossier. Les figures 2 et 3 montrent les deux étapes du procédé de suivi : initialisation par le modèle affine global, puis estimation hiérachique sur les niveaux successifs de maillage.

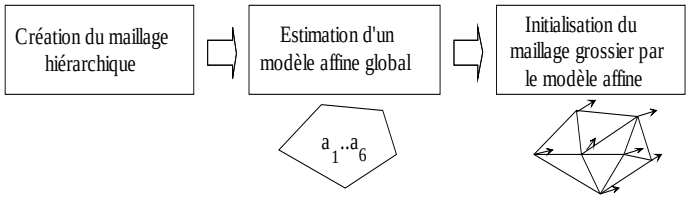

FIG. 2: *Initialisation du modèle grossier*

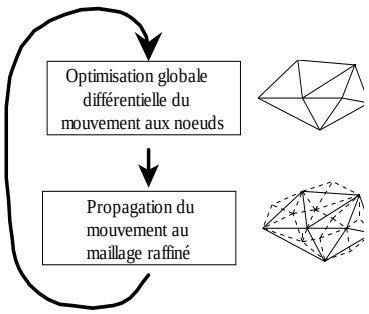

FIG. 3: *Procédé hiérarchique d'estimation*

5 Résultats

Les résultats présentés en figures (5) et (6) montrent le suivi réalisé sur les séquences *foreman* et *mobile and calendar*. Les 6 premières images, correspondant à une rotation du visage vers sa gauche, font apparaitre deux caractéristiques apportées par la méthode :

- le contour externe de l'objet est correctement suivi, à la condition que les mailles soient suffisamment petites,
- le mouvement intérieur à l'objet est également suivi : une concentration de nœuds se forme vers la gauche du visage, due à la rotation.

On observe cependant des déformations importantes de certaines mailles. Ces déformations pourraient être évitées en ajoutant un terme de conservation des formes comme celui utilisé par *Nakaya* [8]. Ce serait cependant au prix d'un positionnement moins précis des nœuds dans les régions d'occultation, pouvant dégénérer le contour externe des objets. Nous avons donc préféré nous appuyer sur l'approche hiérarchique pour limiter ces déformations, en bloquant au cours du procédé itératif d'estimation les retournements éventuels de triangles.

La figure (6) illustre l'aptitude de notre méthode à suivre des objets en mouvement dans une séquence. Ici, le ballon a été initialement segmenté puis suivi. En (6a) nous avons montré l'effet de la seule initialisation du mouvement par le modèle affine global, et en (6b) le raffinement de cette estimation selon notre méthode basée nœuds. La première séquence s'appuyant sur un modèle robuste à peu de paramètres permet de suivre l'objet pendant toute sa période visible (les 180 premières images de la séquence), lui assignant correctement ses mouvements de rotation et de translation. Cependant les déformations internes à l'objet ne sont içi pas prises en compte.

C'est justement l'optimisation du maillage suivante qui a pour rôle de compenser les déplacements internes à l'objet, ce qui est réalisé en (6b). On constate cependant que le maillage "s'accroche" aux motifs du ballon qui disparaissent au cours de sa rotation, déformant celui-ci et nuisant au suivi global de la forme. Un compromis doit donc être réalisé entre le suivi global d'un objet et l'estimation précise de son contenu.

6 Conclusion

Au cours de cet article, nous avons présenté un procédé global de création et de suivi d'objets vidéo par maillage. L'intérêt d'une telle démarche a été mise en évidence par le standard MPEG-4, lui attribuant de nombreuses fonctionnalités, aussi bien d'un point de vu services interactifs, que progressivité de codage ou indexation par le contenu. Notre approche a pour but d'estimer efficacement et de manière robuste le mouvement des objets en question, et ce quel que soit le type de leur déplacement et déformation. Nous avons donc choisi une approche visant à affiner très progressivement la précision de notre estimation, partant de très peu de paramètres (donc d'une estimation plus robuste) puis au fur et à mesure en donnant au modèle davantage de degrés de liberté, pour une optimisation précise. Les résultats en cours confirment la validité de l'approche mais mettent en évidence le compromis à réaliser entre précision de l'estimation et persistance du suivi.

Références

[1] Peter van Beek, A.Murat Tekalp, Ning Zhuang, Isil Celasun, and Minghui Xia. Hierarchical 2d mesh representation, tracking and compression for object-based video. *IEEE Trans. Circ. and Syst. for Video Tech.*, 9(2):353–369, Sept. 1999. (special issue).

[2] I. Celasun, E. Ilgaz, A.M. Tekalp, P. van Beek, and N. Zhuang. Optimal hierarchical design of 2d dynamic meshes for video. pages 899–903, Chicago, USA, Oct. 4-7 1998.

[3] I. Celasun, M. Xia, P.J.L. van Beek, and A.M. Tekalp. Hierarchical 2d mesh design and compression for video. In *Proc. of Picture Conding Symp.*, Berlin, Germany, Sept 1997.

[4] Marc Gelgon. *Segmentation Spatio-Temporelle et Suivi dans une Séquence d'Images : Application à la Structuration et à l'Indexation de Vidéo*. PhD thesis, Université de Rennes I, 1998.

[5] ISO/IEC JTC1/SC29/WG11. N2459. http://www.cselt.stet.it/mpeg/standards, October 1998.

[6] A.K. Katsaggelos, L.P. Kondi, F.W. Meier, J. Ostermann, and G.M. Schuster. Mpeg-4 and rate-distorsion-based shape-coding techniques. *Proceedings of the IEEE*, 86(6):1126–1154, June 1998.

[7] P. Lechat, M. Ropert, and H. Sanson. "Hierarchical Mesh-based Motion Estimation Using a Differential Approach and Application to Video Coding". In *Proc. of EUSIPCO'98*, Rhodes, GRECE, Sept 8 - 11 1998.

[8] Y. Nakaya and H. Harashima. Motion compensation based on spatial transformations. *IEEE Transactions on Circuits and Systems for Video Technology*, 4(3):339–56, 1994.

[9] H. Sanson. Vers une méthodologie pour l'identification paramétrique robuste du mouvement de régions par optimisation non-linéaire. In *Proc. GRETSI*, pages 817–820, Juan-Les-Pins, 18-21 Sept 1995.

[10] A.M. Tekalp, P. van Beek, C. Toklu, and B. Günsel. Two-dimensional mesh-based visual-object representation for interactive synthetic/natural digital video. *Proceedings of the IEEE*, 86(6):1029–1051, June 1998.

(a) (b) (c)

FIG. 4: *Hiérarchie de maillages*

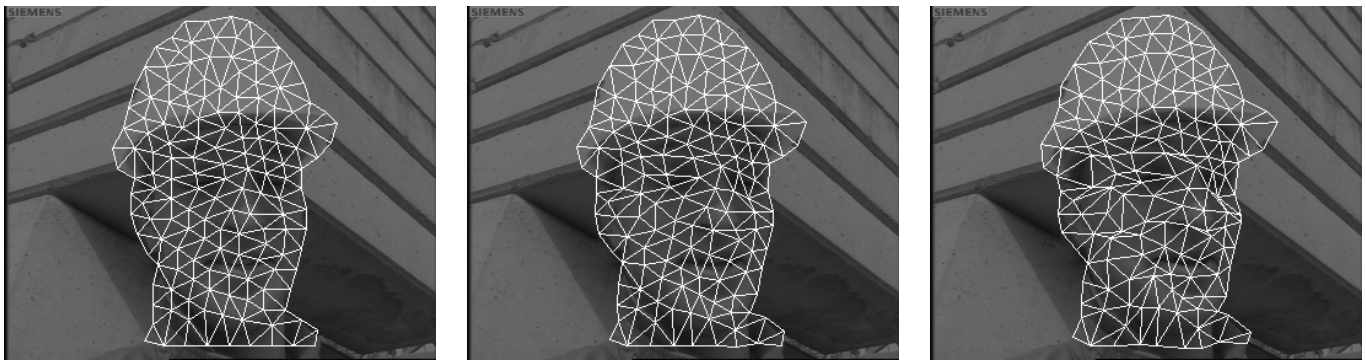

Image 1 Image 2 Image 3

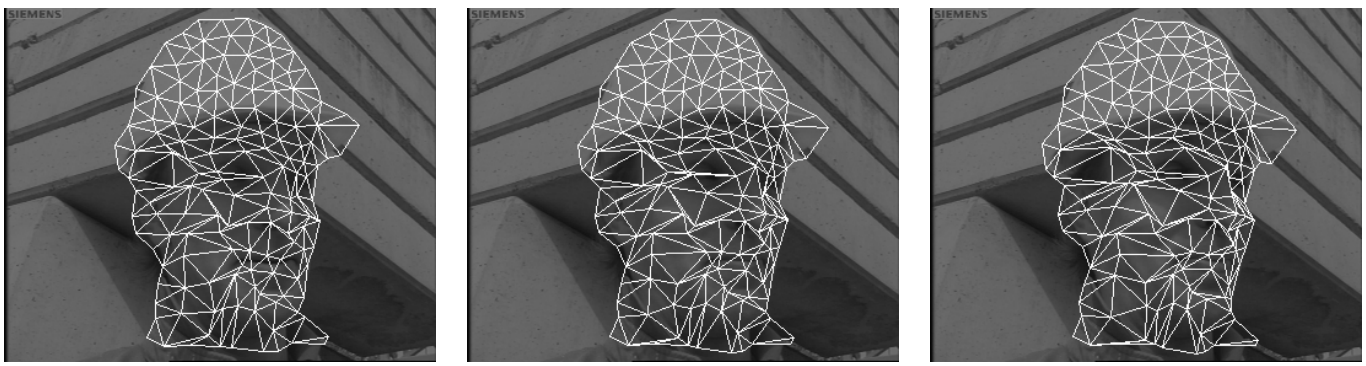

Image 4 Image 5 Image 6

FIG. 5: *Suivi séquence Foreman*

Maillage initial construit sur l'image 1

Image 15 Image 30 Image 45 Image 60

(a) Utilisation du modèle affine global uniquement

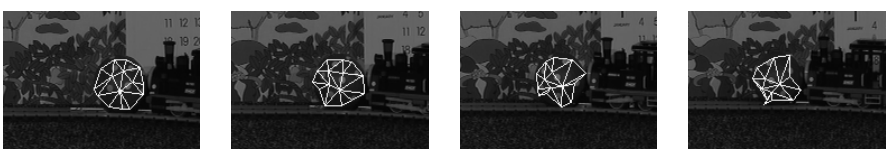

Image 15 Image 30 Image 45 Image 60

(b) Utilisation du modèle affine global et optimisation basée nœuds du maillage

FIG. 6: *Suivi d'objet sur la séquence mobile and calendar*