

Modifications artificielles de trajectoires d'objets vidéo

Franck DENUAL¹, Henri NICOLAS²

¹MICROPROCESS Ingénierie
97 bis, rue de Colombes, BP 87, 92400 Courbevoie, France

²IRISA/INRIA
Campus de Beaulieu, 35042 Rennes Cedex, France
fdenual@microprocess.com, hnicolas@irisa.fr

Résumé – Avec l'apparition du nouveau standard pour la représentation de données multimédia MPEG-4, un grand nombre d'applications multimédia sont appelées à voir le jour. Une application que nous proposons dans ce papier utilise des outils proposés dans cette norme comme l'accès aux objets vidéo représentés par une suite de "Video Object Planes (VOP)", mais aussi les outils de codage d'images mosaïques utilisées pour représenter les fonds statiques de séquences vidéos. A partir de ces deux types de données, nous proposons la reconstruction ou la modification de séquences vidéo en agissant sur les positions et les trajectoires des objets la composant.

Abstract – With the emergence of the MPEG-4 standard, the standard for multimedia applications, a lot of multimedia applications should appear soon. An application proposed here deals with the utilisation of several MPEG-4 tools such as object-oriented coding using Video Object Planes (VOP) and mosaic images for static background (sprite) coding. We propose a method for video sequences reconstruction or modification based on video objects trajectories manipulations.

1 Introduction

Depuis quelques années, avec l'émergence de la nouvelle norme MPEG-4 [1], de nouvelles possibilités en matière de composition et de manipulation de séquences vidéo apparaissent. Certaines de ces techniques, (voir [5] ou [4]), permettent de réaliser des manipulations telles que la superposition de graphiques, de textes ou l'ajout de nouveaux objets vidéo sur une image quelconque et ce, directement dans le domaine compressé. Cependant, la plupart de ces manipulations représentent un lourd investissement en temps du fait de leur non automatisation. De plus ces techniques visent en général la superposition de plusieurs images avec des effets de transparence par exemple. La

principalement la modification du comportement dynamique des objets vidéo de la scène originale ou bien l'ajout d'objets vidéo provenant d'autres séquences. La modification de la séquence se fait de façon interactive, et éventuellement itérative, en se basant sur la définition d'images clés. Chaque image-clé nécessite de choisir un VOP (Video Object Plane) de référence appelé par la suite VOP-clé et d'en définir manuellement les paramètres de positionnement dans un référentiel fixe pour la scène désirée.

Le schéma bloc de la figure 1 décrit sommairement les principales étapes de notre méthode, chacune de ces étapes étant développée dans les paragraphes suivants.

2 Détection des VOPs-clés

La détection des VOPs-clés consiste à choisir parmi une séquence de VOPs l'indice temporel de quelques VOPs. Cette sélection peut être faite manuellement de façon arbitraire ou bien de façon plus pertinente pouvant conduire à une automatiser. Néanmoins, si l'on souhaite pouvoir automatiser le plus possible la méthode, il est nécessaire de déterminer si certains VOPs-clés peuvent être choisis automatiquement et sur quels critères.

2.1 Détection automatique

Il apparaît judicieux de séparer la séquence en intervalles de temps tels que le mouvement de l'objet considéré ne subisse aucun changement de vitesse ou d'accélération dans l'intervalle considéré. Pour cela, nous proposons ici un algorithme de détection automatique des instants clés basé sur le découpage de la fonction d'évolution des po-

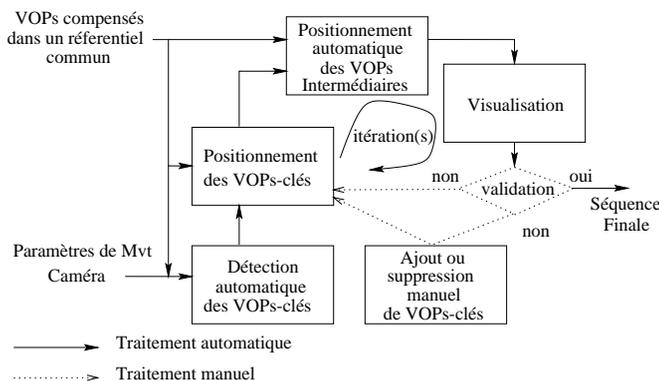


FIG. 1: Schéma-bloc de l'approche proposée

méthode que nous proposons ici vise à permettre la manipulation d'objets vidéo de façon semi-automatique et donc à permettre une interaction avec le contenu intrinsèque d'une séquence vidéo. Ces manipulations concernent prin-

sitions du centre de gravité de l’objet en intervalles de temps répondant à ces contraintes.

Les instants de la séquence originale correspondant à des changements de vitesse d’un objet vidéo peuvent être détectés en considérant les gradients des positions successives du centre de gravité de cet objet:

$$\Delta dx_g(t) = dx_g(t+1) - dx_g(t) \quad (1)$$

$$\Delta dy_g(t) = dy_g(t+1) - dy_g(t) \quad (2)$$

où dx_g, dy_g représentent les coordonnées horizontales et verticales du centre de gravité de l’objet traité dans un référentiel fixe et commun à toutes les images de la séquence (celui d’une image panoramique dans notre cas). Les instants correspondant à des changements d’accélération sont localisés sur les extrema et sur les valeurs nulles des fonctions $\Delta dx_g, \Delta dy_g$, qui correspondent alors aux VOPs-clés. Cependant, afin d’éliminer les nombreuses détections parasites dues aux erreurs d’estimation du mouvement caméra et à la non rigidité de certains objets vidéo, un seuillage des fonctions Δdx_g et Δdy_g doit être effectué. Les VOPs-clés détectés sont donc les suivants:

- Le premier et le dernier VOPs de la séquence.
- Les zéro des fonctions Δdx_g et Δdy_g seuillées (uniquement le premier et le dernier en cas de zéro consécutifs).
- Les extrema significatifs des fonctions Δdx_g et Δdy_g seuillées (un extremum est jugé significatif si la différence entre son amplitude et celle des extrema qui l’entourent dépasse un seuil prédéfini)

2.2 Affinement manuel

La détection automatique ne s’avère pas toujours suffisante pour repérer toutes les modifications de la trajectoire originale d’un objet. De plus, la trajectoire voulue par l’opérateur peut contenir des changements de sens, de vitesse voire d’accélération à des instants différents des originaux. C’est donc l’opérateur qui, au vu de la qualité de la séquence générée, décide de l’ajout, de la suppression ou de la modification de VOPs-clés. Les VOPs ayant été sélectionnés sont ensuite positionnés manuellement dans un repère fixe pour toute la séquence: la position du centre de gravité de l’objet dans ce référentiel est spécifiée ainsi qu’un facteur d’échelle sur l’objet qui traduit la sensation de profondeur. Une rotation de l’objet dans le plan image peut également être introduite. Par la suite, les VOPs-clés associés aux paramètres de positionnement ci-dessus définissent des points de passage obligatoires de l’objet. En d’autres termes, l’étape de positionnement décrite ci-dessus revient à échantillonner la trajectoire de l’objet vidéo. En pratique, une première séquence est générée à partir des VOPs-clés obtenus automatiquement puis affinée manuellement jusqu’à validation.

3 Positionnement automatique des VOPs intermédiaires

Les positions des VOPs intermédiaires dans la scène manipulée sont obtenues par interpolation linéaire des para-

mètres de positionnement définis pour le VOP-clé précédent en (t_1) et le VOP-clé suivant en (t_2) . Selon [2], les termes de positionnement, zoom et rotation peuvent être injectés dans un modèle affine représenté par les coefficients a_i (i pouvant aller jusqu’à 6 paramètres). L’interpolation peut donc s’écrire:

$$a_i(t) = \frac{(t_2 - t) * a_i(t_1) + (t - t_1) * a_i(t_2)}{(t_2 - t_1)} \quad (3)$$

Les VOPs intermédiaires sont alors automatiquement positionnés et mis à l’échelle dans le repère de la scène finale en fonction de leurs paramètres interpolés.

4 Résultats expérimentaux

Les résultats présentés sur les figures 7 à 10 proviennent de la séquence test “Stefan” limitée aux images 220 à 290 qui correspondent au déplacement le plus important. Dans un souci de lisibilité, seules les positions sous échantillonnées d’un facteur 7 sont montrées sur l’image panoramique obtenue par les techniques décrites dans [3], tout comme les paramètres de mouvement de la caméra (translation plus zoom dans le cas présent). Le plaquage des VOPs sur la mosaïque n’est bien sûr qu’une étape intermédiaire avant la régénération des images de la séquence résultat, comme par exemple les images 3 à 5 à comparer avec l’image 2.

La figure 7 correspond à la trajectoire originale de l’objet et sert de référence. Les figures 8, 9 et 10 montrent des trajectoires différentes obtenues en se basant sur des VOPs-clés détectés automatiquement avec un seuil sur les extrema des gradients de 5 (cf. figure 6). Par affinement successif des paramètres de positionnement des VOPs-clés, il a été possible de générer des trajectoires soit en arrière-plan (cf. figure 8), soit au premier plan (cf. figure 9). Enfin, la figure 10 montre une trajectoire inversée de l’objet, gérée par un modèle de mouvement affine à 6 paramètres qui a nécessité l’ajout manuel de deux VOPs-clés afin d’obtenir une séquence visuellement satisfaisante. En effet, les résultats produits ne peuvent être validés automatiquement, et seul l’utilisateur au vu de la séquence générée décide ou non de sa validité.

5 Conclusion

Les résultats proposés montrent clairement que la modification du comportement dynamique des objets d’une séquence vidéo est réalisable assez simplement même lorsque les objets manipulés sont déformables. Cette approche purement 2D représente un avantage au niveau de la gestion des données, des modèles de représentation et des temps de calcul, mais se trouve directement limitée dans le cas de transformations autres que dans le plan de l’image. Cette difficulté peut néanmoins être contournée en utilisant les prises de vue disponibles dans la séquence originale de VOPs classifiés comme proposé dans [6]. A terme, cette application doit s’intégrer sur une plate-forme logicielle de développement d’applications multimédia gérée par l’exécutif temps-réel objet *SoftKernel* de Microprocess.



FIG. 2: Image 230 originale de la séquence "Stefan"



FIG. 4: Extraction de l'image 230 à partir d'une trajectoire modifiée du VOP "Stefan"



FIG. 3: Image 230 reconstruite par plaquage du VOP "Stefan" sur la mosaïque



FIG. 5: Extraction avec un effet de perspective de l'image 230 résultant du plaquage du VOP "Stefan" sur la mosaïque

Références

- [1] R. Koenen. *Overview of the MPEG-4 Standard*. ISO/IEC/JTC1/SC29/WG11 N2459, Octobre 1998
- [2] E. François. *Interprétation qualitative du mouvement à partir d'une séquence d'images*. Thèse Université de Rennes 1, 1991
- [3] H. Nicolas. *Mosaic Representation and Video Object Manipulations for post-production applications*. ICIP, Vol II, pp. 451-455, Chicago Septembre 1998
- [4] B. Shen, V. Bhaskaran, I. K. Sethi. *Digital video blue screen editing in compressed domain*. VCIP, pp. 971-982, San Jose Février 1997
- [5] S.F. Chang. *Compositing and manipulations of video signals for multimedia network video services*. U.C. Berkeley, Aout 1993
- [6] J. Motsch, H. Nicolas. *Classification de mouvement des objets*. Actes du GRETSI99, Vannes, Septembre 1999

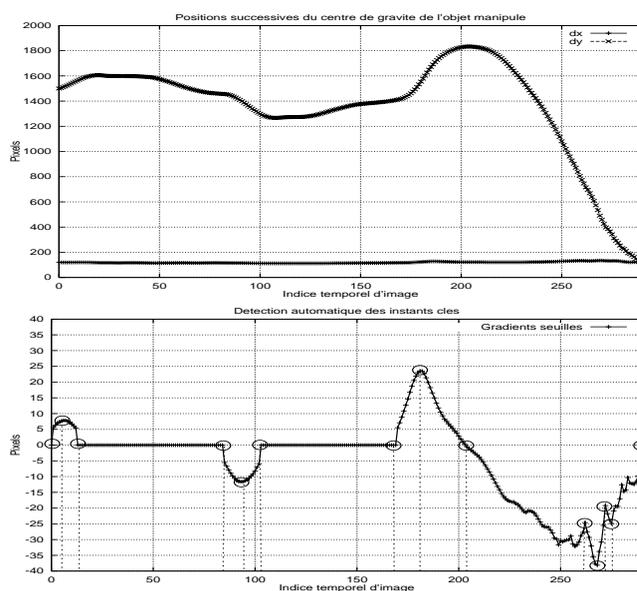


FIG. 6: Détection des instants-clés (séquence "Stefan"), seuil égal à 5

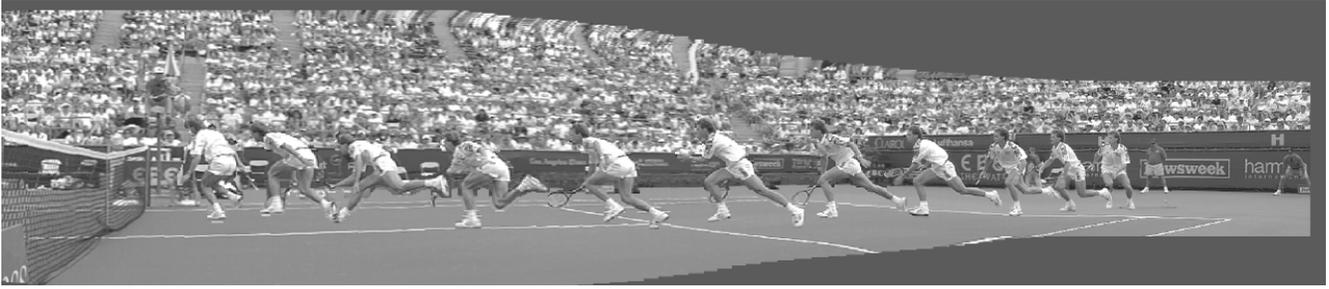


FIG. 7: Scène originale

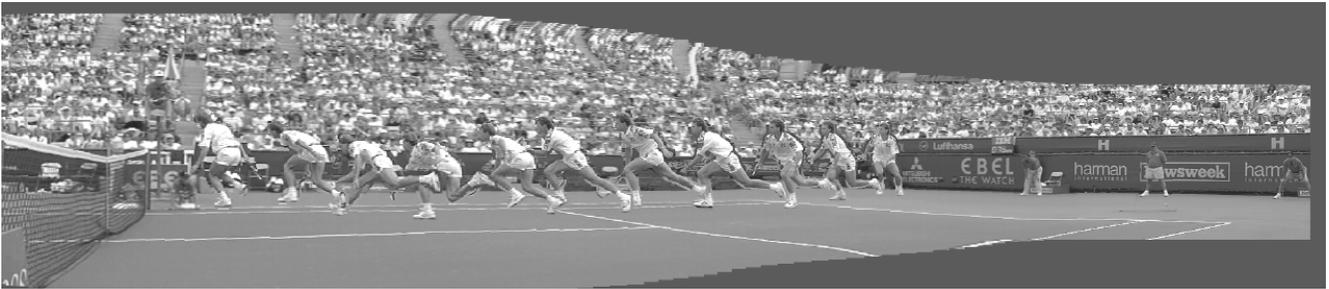


FIG. 8: Trajectoire en second plan

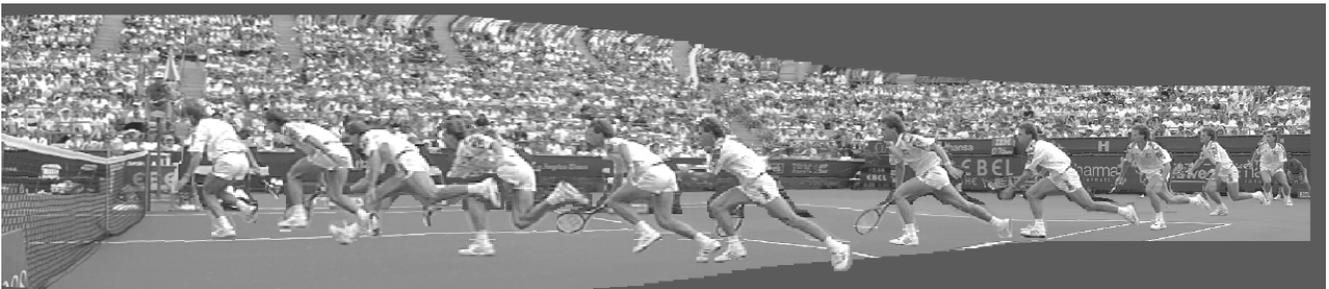


FIG. 9: Trajectoire au premier plan



FIG. 10: Trajectoire inversée, les rectangles indiquent les VOPS-clés ajoutés à la main