Modélisation Hybride 2D/3D de Séquences Vidéos

E. Morillon[1*] R. Balter[1et2] L. Morin[1] S. Pateux[1*]

[1] IRISA/INRIA-Rennes
Campus de Beaulieu avenue du Général Leclerc
35042 Rennes

[2] France Telecom R&D
4 rue du Clos Courtel
35512 Cesson-Sévigné

{raphaele.balter,luce.morin}@irisa.fr
{eric.morillon,stephane.pateux}@rd.francetelecom.com

Résumé

L'extraction de l'information 3D à partir d'une séquence vidéo permet d'obtenir une représentation très avantageuse pour le codage bas débit tout en ajoutant des fonctionnalités avancées à la vidéo. Mais dans le cas de mouvements de caméra dégénérés, comme les rotations pures, cette information 3D ne peut pas être extraite. Dans cet article on propose une représentation originale de vidéos, basée sur un flux de modèles 3D et des mosaïques. L'idée de cette approche hybride 2D/3D est de pouvoir modéliser de manière automatique une séquence vidéo y compris celles comportant des parties correspondant à une rotation pure. La séquence est divisée en portions et pour chacune d'entre elles on identifie le type de mouvement de la caméra. Dans le cas d'une rotation pure on reconstruit une mosaïque et pour les autre mouvements de caméra on extrait un modèle 3D. On présente également la visualisation de la séquence reconstruite par cette représentation.

Mots clefs

Modélisation de vidéos, Reconstruction 3D, Mosaïques.

1 Introduction

1.1 Contexte

Le codage de séquences basé modèles 3D consiste à représenter la séquence par un ou plusieurs modèles 3D de la scène filmée et par des positions de caméra associées. La séquence est restituée en visualisant les modèles à partir des positions caméra spécifiées.

Cette représentation ouvre de nombreuses perspectives d'applications, notamment en compression vidéo. En effet, ce type de représentation demande beaucoup moins d'information à transmettre (les modèles 3D et les positions caméra) que les représentations basées images (tous les pixels).De plus, la représentation par modèles 3D permet des fonctionnalités comme la possibilité de naviguer virtuellement dans la scène, la réalité augmentée ou le rendu en relief par visualisation stéréoscopique [1].

*Maintenant chez France Telecom

L'extraction des modèles 3D à partir d'une caméra monoculaire en mouvement n'est cependant pas toujours possible. Certains mouvements de caméra dits dégénérés ne permettent pas de reconstituer d'information 3D. C'est le cas des rotations pures. La reconstruction 3D, basée sur la stéréovision, nécessite deux points de vue distincts d'un même point 3D. Or lors d'une rotation pure le centre optique de la caméra est immobile comme l'illustre la figure 1, on ne possède donc que des images correspondant au même point de vue.

Les techniques développées pour la modélisation 3D à partir d'images font donc l'hypothèse d'un mouvement de caméra non dégénéré.

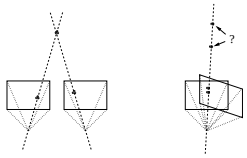

Figure 1 – *Problème introduit par les mouvements dégénérés pour la reconstruction de l'information 3D*

Les mosaïques d'images quant à elles sont des représentations 2D des images bien adaptées au cas des rotations pures de caméra. Elles sont largement utilisées pour la représentation de scènes fixes car elles permettent un codage efficace à bas débit ainsi que la fonctionnalité de navigation interactive.

On propose donc ici une méthode hybride 2D/3D pour permettre de modéliser n'importe quelle séquence vidéo de scènes fixes, y compris celles comprenant des mouvements de caméra dégénérés. Dans les portions de séquences où le mouvement de caméra n'est pas dégénéré, on utilisera une modélisation 3D ; dans celles où une rotation pure de la caméra est détectée, on utilisera une mosaïque.

1.2 Travaux antérieurs

L'extraction de l'information 3D. Il existe différentes méthodes de reconstruction 3D. Dans le cadre de la compression, il faut respecter les contraintes du codage, c'est à dire de travailler sans connaissance *a priori* ni hypothèses

sur les paramètres de caméra, le contenu de la scène ou la longueur de la séquence. Dans ce contexte, F. Galpin a proposé une représentation des vidéos par un flux de modèles 3D plutôt que par un modèle unique et réaliste de la scène [1]. La séquence est découpée en GOFs (Group of Frames), délimités par des images clé communes à deux GOFs successifs et détectées de manière automatique. Pour chaque GOF, le modèle 3D est automatiquement extrait à l'aide d'algorithmes classiques [2] suivis d'un ajustement de faisceaux. La sélection automatique de ces images clé est basée sur :

– le mouvement moyen entre les images,
– le pourcentage de points communs entre les images,
– le résidu épipolaire.

Les mosaïques. Les mosaïques [3, 4] permettent de construire une image unique à partir de plusieurs images.
On distingue les mosaïques obtenues par homographie, les mosaïques cylindriques et les mosaïques sphériques.
Nous avons choisi d'utiliser un algorithme de génération de mosaïques à partir de séquences vidéos qui repose sur une estimation dense et locale du mouvement basé sur des maillages 2D déformables [5].

2 Méthode proposée

La méthode que nous proposons se base sur le schéma de représentation par flux de modèles 3D développé par F. Galpin. On conserve la représentation basée sur un découpage de la séquence vidéo en GOFs. L'idée est alors de développer un schéma hybride utilisant des modèles 3D si le mouvement de la caméra n'est pas dégénéré ou des mosaïques sinon. Dans le but de rester le plus homogène possible, les mosaïques sont représentées sur un cylindre 3D associé à des positions de caméra permettant la restitution des images d'origine. Ainsi, elles seront visualisées exactement comme les modèles 3D.
Le schéma global 2 présente la structure générale de l'application. Les rectangles gris correspondent à la méthode de F. Galpin (flux de modèles 3D). Les autres briques sont ajoutées pour permettre la représentation hybride 2D/3D et sont détaillées dans la suite de l'article dans l'ordre de leur intervention dans l'algorithme.

3 Détection des rotations pures

Cette étape sert à détecter les portions de séquence qui correspondent à des rotations pures de la caméra afin de choisir le type de représentation adaptée au GOF.
Le principe est de comparer le mouvement réel entre les images avec un modèle de mouvement de rotations pures.
Les points d'intérêts suivis dans la séquence permettent de calculer les paramètres du modèle de mouvement.
Pour les N points en correspondance m_1^i , $i = 1..N$ de la première image du GOF et les points m_j^i , $i = 1..N$ de l'image courante j, on calcule les paramètres du modèle de rotation M_{1j}.
Nous estimons la validité du modèle obtenu par le résidu $\frac{1}{N}\sum_{i=1}^{N}$ distance$(M_{1j}(m_1^i), m_j^i)$. Ce résidu correspond à la distance moyenne entre les points prédits par le modèle et les points suivis.
Nous détaillons maintenant trois modèles de mouvement de rotation pure, basés sur une homographie, une projection cylindrique et une projection sphérique.

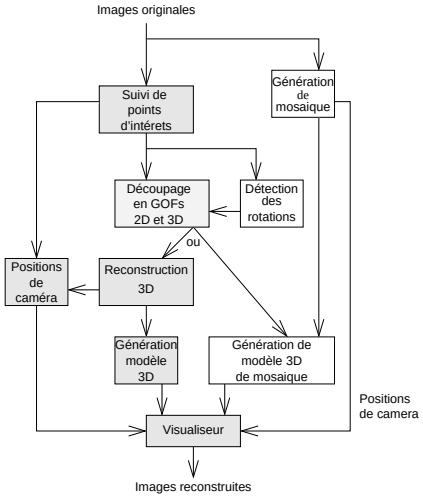

Figure 2 – *Schéma global de la méthode proposée*

3.1 Estimation par homographie

La transformation entre deux images issues d'une caméra effectuant une rotation pure est une homographie.
La première méthode consiste donc à estimer l'homographie 2D optimale entre les deux ensembles de points d'intérêts. L'homographie comporte huit degrés de liberté et peut être estimée à l'aide de quatre couples de points. Nous l'estimons à partir de tous les points disponibles par une minimisation linéaire basée sur une décomposition en valeurs singulières.

3.2 Estimation par projection cylindrique

Le principe est de projeter l'image dans un espace où le mouvement produit par la rotation pure de la caméra est un mouvement simple, dont l'estimation est facilitée.
La projection cylindrique suppose que la caméra effectue des rotations autour de l'axe vertical (parallèle au colonnes du plan image).
Elle consiste à projeter le plan image sur un cylindre d'axe vertical passant par le centre optique de la caméra. On estime alors le mouvement des points à la surface du cylindre. Dans ce cas, le mouvement entre les images projetées se réduit à une translation horizontale, avec un seul degré de liberté. Son estimation aux moindres carrés est simplement donnée par la moyenne des déplacements horizontaux. Le calcul du résidu est toujours effectué dans le plan image d'origine, par rétro-projection des images depuis le cylindre vers les images, après estimation de la translation.

3.3 Estimation par projection sphérique

Le principe de la projection sphérique est similaire. Il a l'avantage d'autoriser tous les axes de rotation. En proje-

tant les images sur une sphère, on obtient deux nuages de points 3D liés par une rotation de l'espace 3D dont l'axe passe par le centre optique de la caméra. La rotation optimale entre les deux nuages de points est estimée par la méthode des quaternions, qui fournit une solution directe par résolution d'un système linéaire [2].

3.4 Discussion sur les critères

Nous avons validé ces trois méthodes par des tests sur des séquences synthétiques qui montrent que le résidu pour les différentes méthodes augmente linéairement avec l'amplitude des translations et reste nul pendant les rotations pures. Nous avons également testé l'efficacité de nos critères sur des séquences réelles. La séquence *thabor* correspond à une scène filmée par une caméra en rotation sur un trépied. La séquence *escalier* quant à elle correspond à une translation latérale.
Les figures 3 et 4 montrent l'évolution des critères au cours des séquences *thabor* et *escalier*, pour 2 GOFs successifs. Elles ont les mêmes allures d'une séquence à l'autre, et ce quel que soit le critère observé. Pourtant, la première séquence correspond à une rotation et la seconde à une translation. Les résidus de la séquence en rotation peuvent être dus à une légère translation et à du bruit. Cependant, ils restent faibles devant l'amplitude du mouvement effectué par la caméra. À l'opposé, les résidus augmentent vite par rapport au mouvement de la caméra dans la séquence en translation.
On propose donc comme nouveaux critères $\frac{\text{résidu de rotation}}{\text{déplacement total}}$ et obtenons les courbes 5 et 6 pour les mêmes séquences.

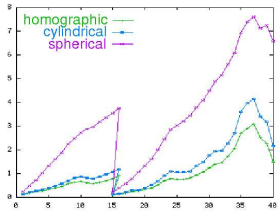

Figure 3 – *Résidu pour la séquence thabor*

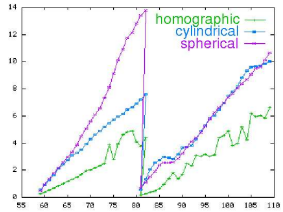

Figure 4 – *Résidu pour la séquence escalier*

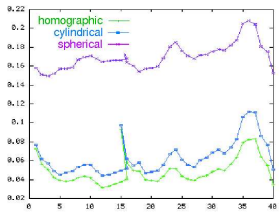

Figure 5 – *Résidu de la rotation/déplacement total pour la séquence thabor*

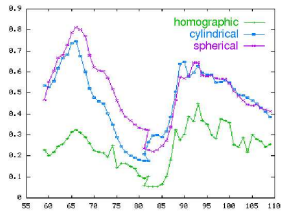

Figure 6 – *Résidu de la rotation/déplacement total pour la séquence escalier*

Le seuil de 0.2 pour le critère basé sur la projection sphérique permet nettement de distinguer la séquence en rotation (*thabor*) de la séquence en translation (*escalier*).

Ce critère a donc été intégré dans la sélection des images clé délimitant les GOFs, pour déterminer si le GOF courant est un GOF-3D ou un GOF-2D.

Maintenant que nous savons détecter les parties de séquences correspondant à des rotations, on va s'intéresser à la production des mosaïques pour représenter ces groupes d'images.

4 Génération des mosaïques

4.1 Pré-traitement par projection

Nous avons choisi d'utiliser les mosaïques produites par la méthode proposée dans [5] car elle ne fait pas explicitement l'hypothèse d'un mouvement particulier de la caméra. Elle permet donc de traiter les séquences correspondant à des rotations pures ou couplées à un faible mouvement de translation, souvent présent dans les séquences acquises à la main.

Cependant, cette méthode de génération de mosaïques fait l'hypothèse d'un mouvement affine, or le mouvement entre nos images est homographique. Pour cela, nous avons adaptés nos données d'entrées en projetant chaque image sur un morceau de cylindre avant de les fournir au mosaicker. Cette méthode à l'avantage d'être simple et ne fait pas d'hypothèses sur l'ampleur des rotations considérées, par contre elle nécessite la connaissance de la focale et elle est limitée aux rotations d'axe vertical.

4.2 Résultats

Nous a testé la génération des mosaïques sur la séquence en rotation *thabor* dont la première et dernière images ainsi qu'une image intermédiaire sont présentées figure 7.

L'image de mosaïque 8 est obtenue sans pré-traitement des images d'entrée. On observe un élargissement de l'image quand on s'écarte de la première image. Cette déformation n'apparaît plus dans la mosaïque 9 obtenue après projection.

Figure 7 – *Images d'entrées originales*

Figure 8 – *Mosaïque obtenue sans pré-traitement*

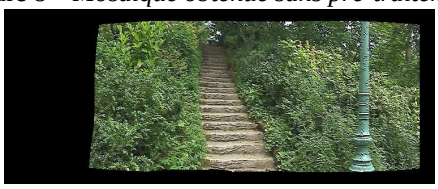

Figure 9 – *Mosaïque obtenue avec projection cylindrique*

Figure 10 – *Images originale et reconstruite.*

5 Visualisation

Afin de visualiser de manière homogène les GOFs 2D et 3D, nous proposons de construire, à partir des mosaïques, un modèle 3D associé à des positions de caméra qui sera re-projeté à la manière des modèles des GOF 3D pour reconstruire les images d'origines. Ce traitement homogène est intéressant en particulier pour les applications interactives où une visualisation en temps réel est nécessaire. On choisit d'associer à la mosaïque un modèle 3D cylindrique texturé par la mosaïque ayant pour rayon la focale utilisée lors du pré-traitement de génération de mosaïque. Les positions de caméra sont paramétrées par une translation horizontale par rapport à la position de départ, orientée vers la première image de la mosaïque. Lors de la génération de la mosaïque, on estime la translation horizontale δX de chaque image avec la première. L'angle cherché est $\frac{\delta X}{\text{focale}}$.

La figure 10 présente une image originale ainsi que l'image reconstruite en projetant la mosaïque cylindrique. Les figures 11 et 13 montrent un GOF 2D et un GOF 3D dans le visualiseur pour la séquence *cloître* contenant des mouvements de translation ainsi que des mouvements de rotation pure de la caméra. Le visualiseur comprend différentes vues du modèle reconstruit ainsi que la position de caméra courante. On présente également des images reconstruites extraites de ces deux GOFs dans les figures 12 et 14.

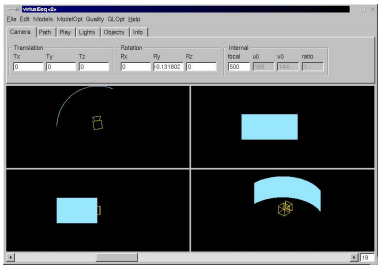

Figure 11 – *Visualiseur pour un GOF 2D de la séquence cloître*

Figure 12 – *Image du GOF 2D reconstruite*

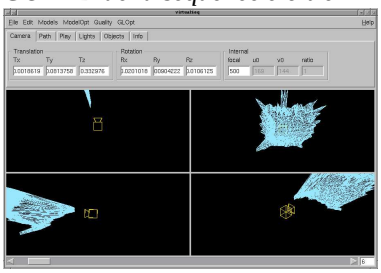

Figure 13 – *Visualiseur pour un GOF 3D de la séquence cloître*

Figure 14 – *Image du GOF 3D reconstruite*

6 Conclusion

Nous avons présenté dans cet article une représentation hybride 2D/3D pour les séquences vidéo de scènes statiques. Elle est basée sur un flux de modèles 3D et de mosaïques. On conserve ainsi les avantages des représentations basées modèles 3D, c'est à dire le fait d'être adaptée au codage bas débit et d'ajouter des fonctionnalités à la vidéo. De plus on peut traiter de manière homogène les cas ou le mouvement de caméra ne permet pas de reconstruction 3D. La génération des mosaïques reste cependant limitée aux rotations d'axe vertical. Il serait intéressant d'étendre cette méthode à l'ensemble des trois rotations en utilisant une projection sphérique plutôt que cylindrique.

Références

[1] Franck Galpin et Luce Morin. Sliding adjustment for 3d video representation. *Eurasip Journal ASP, special issue on Signal Processing for 3D Imaging and Virtual reality*.

[2] Olivier Monga Radu Horaud. *Vision par ordinateur : outils fondamentaux*. Editions Hermès, 1993.

[3] R. Szeliski. Image mosaicing for tele-reality applications. Dans *WACV94*, pages 44–53, 1994.

[4] H. Shum et R. Szeliski. Panoramic image mosaics. Rapport technique MSR-TR-97-23, Microsoft Research, 1997.

[5] G. Marquant S. Pateux et D. Chavira-Martinez. Object mosaicking via meshes and crack-lines technique. application to low bit-rate video coding. Dans *Picture Coding Symposium 2001*.