Codage vidéo basé sur des modèles génériques 2D et 3D

Model-based video coding with generic 2D and 3D models

Gaël Sourimant Luce Morin

IRISA / INRIA Rennes
Campus Universitaire de Beaulieu,
Avenue du Général Leclerc, 35042 RENNES Cedex - France
{gael.sourimant,luce.morin}@irisa.fr

Résumé

L'extraction d'informations 3D à partir de vidéos fournit une représentation adaptée au codage bas débit et permet également de proposer des fonctionnalités avancées telles que la réalité augmentée ou la navigation interactive dans des environements photo-réalistes. Mais pour des mouvements de caméra dégénérés, tels que les rotations pures, les informations de profondeur ne peuvent pas être calculées. Nous proposons dans cette article une représentation originale basée sur un flux hybride de modèles 2D et 3D. L'idée de cette approche est de permettre une modelisation valide pour toutes les vidéos, y compris celles contenant des rotations pures. La séquence est divisée en plusieurs sous-parties, et pour chacune le mouvement de caméra est identifié. En fonction de ce mouvement de caméra, un modèle 3D, une mosaïque cylindrique ou une mosaïque sphérique est extrait. Les modèles générés pour chaque sous-partie sont contruits pour être visualisés de façon homogène. Des résultats sur des vidéos en environement réel et synthétique sont présentés.

Mots Clef

Codage basé modèles, reconstruction 3D, mouvements dégénérés.

Abstract

3D extraction from video gives a representation adapted to low bitrate coding and provides enhanced functionalities such as 3D cues for augmented reality and interactive navigation in photo-realistic environments. But for degenerated motions of camera, like pure rotation, 3D information can not be retrieved. In this article we propose an original representation based on a hybrid 2D/3D models stream. This approach provides a modelling for all sequences including those with rotations. The sequence is divided into portions and for each one the motion of the camera is identified. Depending on the type of motion a 3D model, a cylindrical mosaic or a spherical mosaic is extracted. They are constructed in order to be suitable for an homogeneous visualization process. Results are shown for synthetic and real video sequences.

Keywords

Model-based coding, 3D renconstruction, degenerated motion.

1 Introduction

1.1 Contexte

Le codage vidéo basé modèles 3D consiste en une représentation de la vidéo à l'aide d'un ou de plusieurs modèles 3D de la scène filmée. En reprojetant ces modèles on peut obtenir une séquence virtuelle similaire à l'originale mais en proposant en plus des fonctionnalités avancées telles que la réalité augmentée, la navigation virtuelle ou le changement d'illumination globale. De plus, ces représentations basées modèles sont beaucoup plus compactes que les représentations classiques basées images.

L'extraction de modèles 3D à partir d'images est basée sur une analyse du mouvement de la caméra et nécéssite donc plusieurs points de vue différents de la même scène. En particulier, une caméra décrivant un mouvement de rotation pure ne permet pas de remonter à une quelconque information de profondeur puisqu'il n'y a pas d'intersection entre les différentes lignes de vue (voir figure 1). Pour une reconstruction basée images classique on fait donc l'hypothèse que la caméra n'effectue pas un tel type de mouvement. D'un autre côté, les mosaïques sont tout à fait appropriées pour représenter une vidéo obtenue avec un mouvement rotationnel. Nous proposons donc une méthode hybride 2D/3D originale basée à la fois sur des modèles 3D et des mosaïques. Le but est de pouvoir traiter tout type de vidéo représentant une scène fixe, y compris celles acquises avec une caméra effectuant un mouvement de rotation pure.

1.2 Travaux antérieurs

Modélisation 3D. Retrouver les informations tridimensionelles à partir de vidéos est un sujet étudié depuis longtemps dans le domaine de la vision par ordinateur [13,

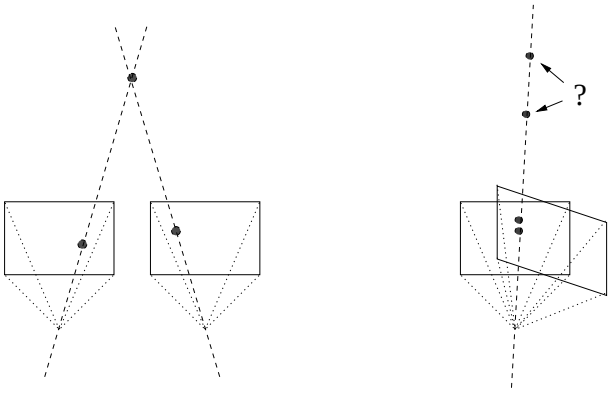


FIG. 1 – *Une translation de la caméra est nécéssaire pour retrouver les informations de profondeur à partir du mouvement : en cas de rotation pure, les lignes de vue se superposent*

7]. Cependant les méthodes proposées dans la littérature font généralement l'hypothèse que la vidéo est acquise sans mouvements dégénérés, ou alors qu'une intervention de l'utilisateur lors de la reconstruction ou la mise en correspondence est nécéssaire [1, 17].
En codage vidéo basé modèles 3D, les conditions d'acquisition ne sont pas contraintes mais des hypothèses sont faites sur le contenu de la scène. Un modèle 3D du contenu de la scène est connu à priori, dont la pose, les textures (et les possibles déformations non-rigides) sont estimées à partir de la vidéo elle-même. Cette approche est très efficace pour coder des vidéos au contenu spécifique, telles que des visioconférences [14].
Comme notre but est de proposer une représentation 3D pour toute vidéo, aucune hypothèse n'est faite sur les paramètres de caméra, le contenu de la scène ou la longueur de la vidéo. Dans ce contexte Galpin [3] a proposé une méthode basée sur un flux de modèles 3D plutôt que de chercher à estimer un unique modèle réaliste de la scène. Chaque modèle est valide pour une portion donnée de la séquence originale appelée GOP (*Group Of Pictures*). Ces GOPs sont délimités par des images-clefs sélectionnées automatiquement. Pour chaque GOP un modèle 3D est généré et la cohérence entre ces différents GOPs est assurée par un ajustement de faisceaux [5].

Mosaïques. Les mosaïques peuvent être obtenues grâce à des projections homographiques, cylindriques ou sphériques [16, 15, 2].
Les mosaïques homographiques sont bien adaptées pour reconstruire des scènes planaires et peuvent également être utilisées en cas de rotations pures de faible amplitude. Cependant, les mosaïques cylindriques et sphériques sont mieux adaptées au cas des rotations pures, puisqu'elles permettent de gérer de larges rotations et évitent le problème de distortion dans les images éloignées de l'image de référence.

Sélection de modèle. La sélection de modèles dans le cas général a été étudiée par Kanatani [8, 9], qui présente une approche basée sur un critère combinant résidu et complexité pour chaque modèle. Ce principe a été utilisé en particulier pour la sélection de modèles de mouvement par Berger et al. [19] et Torr [18]. La sélection du modèle de mouvement basée sur la complexité n'est pas bien adaptée à notre approche, étant donné que nous voulons favoriser la modélisation 3D et utiliser la modélisation 2D uniquement en cas d'échec. Nous utilisons donc des critères uniquement basés sur des résidus, avec un algorithme favorisant la modélisation 3D.

2 Solution proposée

Nous allons présenter dans cette partie notre représentation hybride 2D/3D pour la vidéo, basée sur le schéma proposé par Galpin et al. [4] et Morillon et al. [11].

2.1 Principe général

Notre solution est basée sur un schéma d'analyse/synthèse. Lors de la phase d'analyse, la vidéo est partitionnée en *groups of pictures* (GOPs) (voir figure 2) et pour chaque GOP, un modèle 3D texturé (général, cylindrique ou sphérique) est calculé. Chaque modèle est associé à un ensemble de positions de caméra, une pour chaque image dans la vidéo. Le flux de modèles 3D est ensuite encodé puis transmis. Au récepteur, le flux est décodé en synthétisant la vidéo originale à partir des différents modèles 3D et des positions de caméra (voir figure 3), en faisant un rendu classique pour le GOP courant.

2.2 Notions importantes

- Les GOPs sont délimités par deux images particulières, appelées *images-clefs*.
- Deux GOPs consécutifs G_1 et G_2 partagent respectivement leur dernière et première image-clef (voir figure 2).
- Un maillage 3D est associé à chaque GOP. La texture appliquée sur ce maillage est soit sa première image-clef dans le cas général, soit une mosaïque regroupant toutes les images du GOP dans le cas de rotations pures.
- L'appelation "GOP 2D" se réfère dans la suite de cet article aux GOPs où la caméra décrit un mouvement rotationnel pur. Puisque dans ce cas aucune information 3D ne peut être estimée, ces parties de vidéo sont modélisées à l'aide de moaïques 2D, qui sont au final plaquées sur un maillage 3D cylindrique ou sphérique pour permettre une procédure de rendu homogène.

2.3 Hypothèses

Nous rappelons ici les hypothèses principales sur lesquelles se base notre méthode.

- La scène est supposée statique, ou au moins segmentée en mouvement.
- Cette scène est filmée par une caméra monoculaire en mouvement.

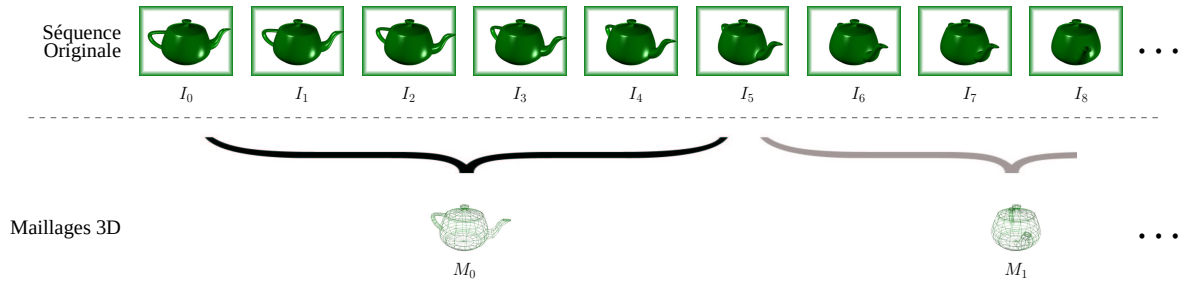


FIG. 2 – *Principe général d'une représentation par flux de modèles 3D.*

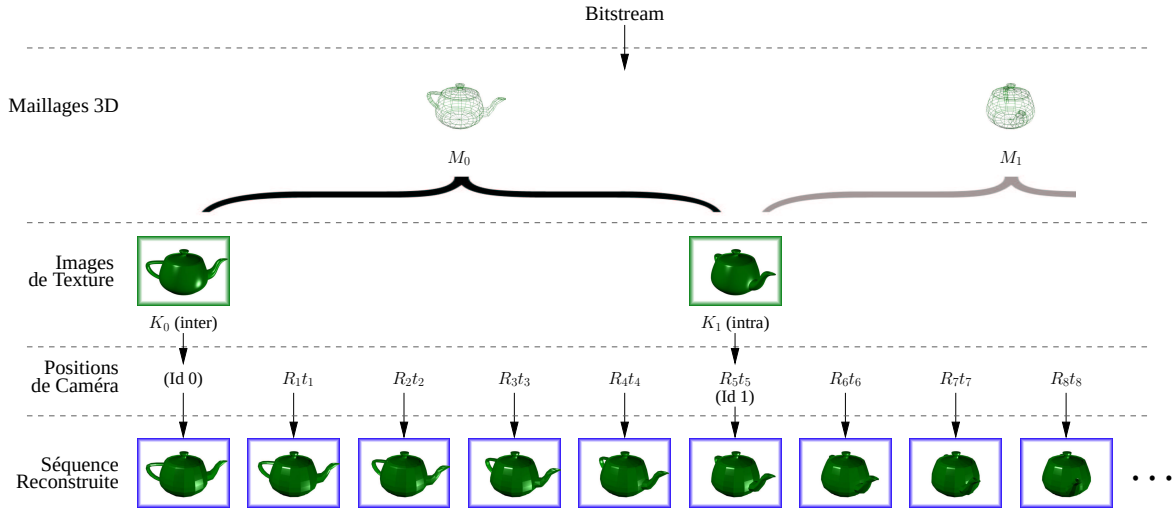


FIG. 3 – *Phase de synthèse : les maillages 3D, les images de textures et les positions de caméra sont utilisées pour reconstruire les images originales.*

- Le mouvement de la caméra n'est pas contraint.
- Les paramètres intrinsèques et extrinsèques de la caméra sont inconnus.
- La focale de la caméra est fixe (pas de zoom).
- Aucune hypothèse n'est faite sur le contenu de la scène, à l'exception du fait qu'elle ne contient pas ou peu de surfaces spéculaires.

2.4 Etapes de l'algorithme

Nous décrivons ici chaque étape de notre algorithme, illustré sur la figure 4.

Estimation de mouvement : La première étape consiste à estimer le champ de mouvement tout au long du GOP, c'est-à-dire le vecteur déplacement pour chaque pixel entre la première et la dernière image-clef. Cette estimation est faite en utilisant un maillage 2D déformable, dont on peut trouver une présentation dans [10, 12, 3]. L'estimation du mouvement entre deux images I_t et I_{t+1} est faite en maillant régulièrement I_t avec un maillage 2D. Puis, pour chaque nœud de ce maillage, on recherche le mouvement qui minimise l'erreur quadratique moyenne entre I_t et I_{t+1} compensée en mouvement. Le mouvement $\vec{u}$ de chaque pixel à l'intérieur d'un triangle est donné par la somme pondérée du mouvement de chaque sommet de ce triangle. Le mouvement estimé pour un GOP donné est alors calculé en accumulant les différents champs de mouvement entre les différentes images.

Extraction et suivi de points d'intérêt : L'extraction de points d'intérêts se fait sur la première image-clef du GOP courant, en utilisant un détecteur de Harris [6] sur l'ensemble des nœuds du maillage 2D déformable à sa résolution la plus fine. Le suivi de ces points est déduit des différentes déformations de ce maillage, calculées lors de la phase d'estimation dense de mouvement. L'utilisation d'une décimation des points du maillage ayant une réponse suffisante à un détecteur de Harris se justifie d'une part pour garantir une répartition homogène des sommets, et d'autre part pour ne pas retenir les points situés dans les zones trop uniformes du point de vue photométrique, c'est-à-dire là où notre confiance en les vecteurs mouvement estimés est la moins forte.

Estimation de la pose : L'estimation du mouvement de la caméra est effectuée en se basant sur le suivi des points

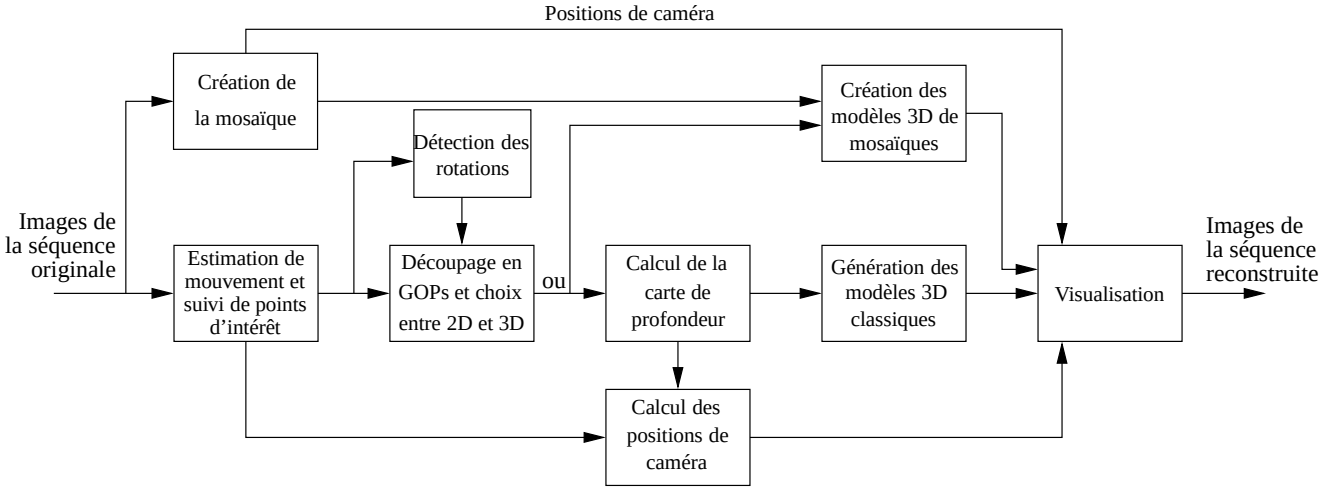


FIG. 4 – *Chaîne algorithmique*

d'intérêts en utilisant des valeurs approximatives pour les paramètres intrinsèques.

Création des mosaïques : Lors de l'analyse d'un GOP, le type de modèle qui correspondra le mieux aux images en entrée est inconnu. Une mosaïque est donc construite en parallèle de l'estimation de mouvement. Le recalage entre les différentes images à intégrer dans la mosaïque est calculé grâce à l'estimateur de mouvement par maillage déformable. Cette mosaïque sera abandonnée plus tard si un maillage 3D classique peut être généré.

Estimation de la carte de profondeur : La carte de profondeur est déduite de la carte de disparité (c'est-à-dire le champ de mouvement) par une triangulation classique, en utilisant les paramètres de caméra estimés.

Reconstruction 3D : Quand une image-clef est sélectionnée pour clôturer un GOP le maillage correspondant est alors généré. Si un GOP 3D peut être reconstruit, on applique un maillage triangulaire régulier sur l'image de profondeur et l'on déplace les nœuds à la profondeur correspondante. Dans le cas contraire un maillage cylindrique ou sphérique associé à une mosaïque est construit.

Visualisation de la séquence reconstruite : La séquence reconstruite finale peut-être visualisée avec un logiciel dédié, qui prend en entrée les différents maillages texturés associés aux positions successives estimées de la caméra. En plus de la visualisation, certains traitements spécifiques peuvent également être appliqués à la séquence, comme un changement simple d'illumination globale, une modification du chemin emprunté initialement par la caméra, ou la génération d'une séquence stéréo en vue d'une visualisation immersive de la vidéo traitée.

3 Sélection automatique des images-clefs

La taille des GOPs n'est pas fixée, mais déterminée par les données vidéo. Un algorithme de sélection des images-clefs de la séquence a donc été proposé. Cette sélection se fait de façon automatique, grâce à différents critères basés sur une analyse du suivi des points d'intérêt. Ce choix est d'autant plus important qu'il va au final déterminer la viabilité des modèles reconstruits. Nous voulons également que cet algorithme réponde aux contraintes suivantes :

- Déterminer le type du GOP courant parmi 3D, panoramique et sphérique.
- Favoriser les GOPs de type 3D et ne passer aux GOPs de type 2D que si la reconstruction 3D est impossible.
- Maximiser la taille des GOPs pour éviter la redondance, et assurer que la reconstruction soit faite avec une ligne de base suffisament importante.

3.1 Critères de sélection

Les critères de sélection utilisés sont inspirés par ceux définis par Galpin [3] et Morillon et al. [11] dans le cas des GOPs 3D et panoramiques. Ces critères ont été adaptés pour prendre en compte le cas sphérique, et l'algorithme de sélection a été complété pour qu'il puisse traiter n'importe quelle séquence comportant des mouvements 3D, panoramiques ou sphériques. Ces critères sont estimés pour chaque image courante I_{t+p} et déterminent si oui ou non celle-ci est choisie comme la prochaine image-clef K_{t+1}.

Déplacement apparent $\mathbf{C_d}$. $D_{t,t+p}$ représente le mouvement moyen apparent des points entre l'image courante I_{t+p} et l'image-clef I_t qui la précède dans la séquence. Le critère testant $D_{t,t+p}$ est défini comme :

$$C_d \Leftrightarrow D_{t,t+p} > S_d$$
$$\text{où} \quad D_{t,t+p} = \frac{1}{N_{t+p}} \sum_{i=1}^{N_{t+p}} \|\vec{u}(m^i_{t,t+p})\| \tag{1}$$

avec N_{t+p} le nombre de points suivis depuis la dernière image-clef I_t jusqu'à l'image courante I_{t+p}, $\vec{u}$ le vecteur mouvement du point m^i entre I_t et I_{t+p}, et S_d un seuil sur le mouvement apparent moyen des pixels qui a été fixé expérimentalement à 10 pixels. Ce critère assure que l'estimation de la profondeur sera faite avec une ligne de base significative.

Points restants $\mathbf{C_p}$. C_p supervise le pourcentage de points communs entre I_t et I_{t+p}, et est exprimé comme :

$$C_p \Leftrightarrow \frac{N_{t+p}}{N_t} > S_p \tag{2}$$

avec S_p le seuil sur le pourcentage de points restants. Ce critère assure que deux images-clefs partagent suffisament d'information pour construire un modèle valide de la portion de séquence qu'elles délimitent. Ce seuil a été fixé à 70%.

Résidu épipolaire $\mathbf{C_e}$. C_e est défini comme :

$$\begin{gathered} C_e \Leftrightarrow \tfrac{1}{N_{t+p}} \sum_{i=1}^{N_{t+p}} (\mathfrak{D}^i_{t,t+p} + \mathfrak{D}^i_{t+p,t}) < S_e \\ \text{où} \quad \mathfrak{D}^i_{t,t+p} = d^2(\tilde{m}^i_t, F_{t,t+p}.\tilde{m}^i_{t+p}) \\ \text{et} \quad F_{t+p,t} = F^{\mathbf{t}}_{t,t+p} \end{gathered} \tag{3}$$

avec $\tilde{m}_t$ un point d'intérêt dans l'image I_t exprimé en coordonnées homogènes, $\tilde{m}_{t+p}$ son correspondant dans I_{t+p}, $F_{t,t+p}$ la matrice fondamentale estimée et S_e un seuil sur la précision de la mise en correspondance, fixé à 0,5 pixels. Ce critère permet de tester le résidu épipolaire, calculé à partir de la matrice fondamentale et des points mis en correspondance entre la dernière image-clef et l'image courante. Il assure ainsi que le modèle 3D se reprojette sur la deuxième image-clef avec une erreur sub-pixelique, et il garantit que le mouvement de la caméra ainsi que la matrice fondamentale sont cohérents avec le champ de mouvement calculé par l'estimation dense par maillage.

Résidu de rotation $\mathbf{C_r}$. Le critère C_r permet de détecter les mouvements rotationnels purs. Il est défini comme :

$$C_r \Leftrightarrow \frac{E_r}{D_{t,t+p}} < S_r$$

On estime ici la transformation image induite par une rotation pure (homographie planaire) qui correpond le mieux au mouvement estimé des points d'intérêt. Le résidu moyen est calculé de la façon suivante :

$$E_r = \frac{\sum_{i=1}^{N} \|H.m^i_2 - m^i_1\|}{N} \tag{4}$$

où m^i_1 et m^i_2 sont des points d'intérêt en correspondance, respectivement dans le première et la dernière image du GOP courant, et où H est la matrice d'homographie estimée. C_r évalue le résidu de rotation par rapport au déplacement moyen dans l'image. Il considère donc le contribution

```
3DFaisable = faux;
TypeGOP = inconnu;

Si C_d Alors
  Si C_p Alors
    Si (C_e ∧ ¬C_r) Alors
      3DFaisable = vrai; ContinueGOP;
    Sinon
      Si 3DFaisable Alors
        Finalise3D;
      Sinon ContinueGOP;
  Sinon
    Si (C_e ∧ ¬C_r) Alors
      Si TypeGOP=2D Alors Finalise2D;
      Sinon Finalise3D;
    Sinon
      Si 3DFaisable Alors Finalise3D;
      Si (TypeGOP==2D ∧¬C_r) Alors Finalise2D;
      TypeGOP=2D; ContinueGOP;
Sinon ContinueGOP;
```

FIG. 5 – *Algorithme de sélection des images-clefs*

relative des composantes rotationnelles et translationnelles dans le déplacement image observé. Le seuil S_r est fixé expérimentalement à 0.05 pixels. On peut également noter que quand C_r est *vrai* (i.e. un mouvement rotationnel a été détecté), le critère épipolaire C_e n'est alors plus signifiant.

3.2 Algorithme de sélection

Nous présentons ici l'algorithme de sélection des images-clefs utilisant les critères définis précédemment. Il est exécuté pour chaque image de la séquence et indique, si une image-clef est sélectionnée, si elle délimite un GOP de type 2D ou 3D. Le GOP en cours d'analyse est classifié par défaut comme *inconnu*. Si à l'instant t la géométrie 3D de la scène est suffisament bien estimée alors le GOP sera classifié comme étant de type 3D. Il sera de type 2D uniquement s'il n'y a pas assez de points suivis dans la séquence et si le critère épipolaire n'a jamais été vrai. La fin d'un GOP 2D a lieu dès que C_r devient faux. La figure 5 illustre plus précisément cet algorithme.

4 Génération des modèles 3D

Une fois que le type du GOP est déterminé le modèle correspondant est construit. Pour les GOPs 2D, des modèles 2D planaires seraient suffisants. Cependant, pour garder le processus homogène au décodage, nous voulons aussi représenter les GOPs 2D comme des maillages 3D texturés associés à un ensemble de positions de caméra. Nous construisons pour cela un cylindre en cas de rotations d'axe vertical, et une géosphère en cas de rotations plus générales. Les deux ont une mosaïque associée en guise de texture.

4.1 Modèles 3D

Dans un contexte de mouvement générique, un modèle 3D est reconstruit une fois que l'image-clef correspondante est sélectionnée. Dans un premier temps, la carte de disparité

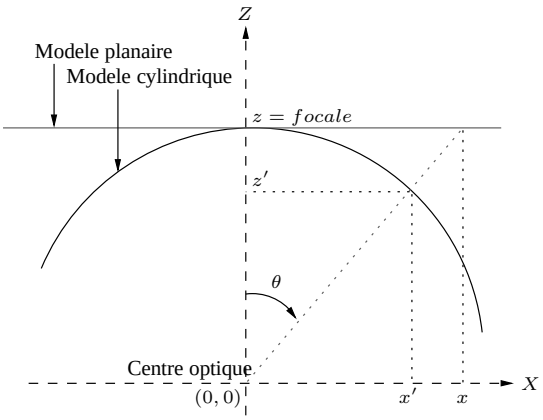
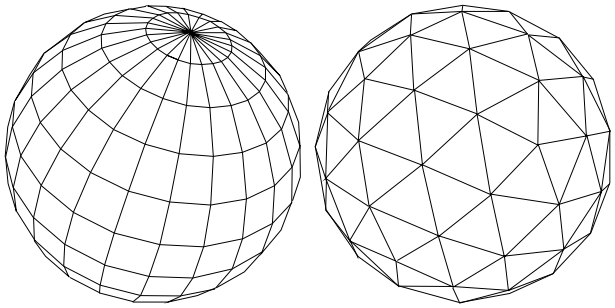
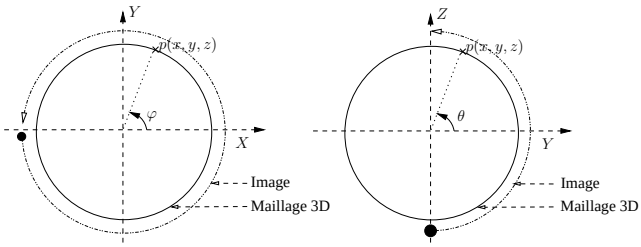


FIG. 6 – *D'un modèle planaire à un modèle cylindrique*



(a) Sphère (b) Géosphère

FIG. 7 – *Comparaison des modèles 3D de type sphérique*



(a) Mapping horizontal (b) Mapping vertical

FIG. 8 – *Mapping d'une image sur une sphère*

est utilisée dans une phase de reprojection pour calculer la carte de profondeur en chaque pixel. Notons que nous souhaitons que cette reprojection sur la première image du GOP soit parfaite. Dans un deuxième temps, la carte de profondeur est maillée uniformément, et chaque nœud se voit alors assigné la profondeur correspondante sur la carte. L'utilisation d'un maillage uniforme plutôt qu'adaptatif est justifiée par le fait que dans le cas uniforme nous avons juste à transmettre les profondeurs de chaque nœud dans un ordre prédéfini, alors que dans le cas adapatatif nous aurions à transmettre la topologie entière du modèle.

4.2 Modèles cylindriques

Dans le cas d'un mouvement panoramique, le modèle 3D généré est un cylindre centré sur le centre optique de la caméra, et dont le rayon est égal à la focale estimée. La mosaïque correspondante est plaquée sur le cylindre de telle sorte que les reprojections du modèle sur de plan image de la caméra virtuelle génère les images originales. Cette procédure est très simple et peut être vue comme une transformation de coordonnées cartésiennes en coordonnées cylindriques (voir figure 6):

$$\begin{cases} x' = f sin\theta \\ z' = \sqrt{f^2 - x'^2} \\ \theta = tan^{-1}(\frac{x}{f}) \end{cases} \quad (5)$$

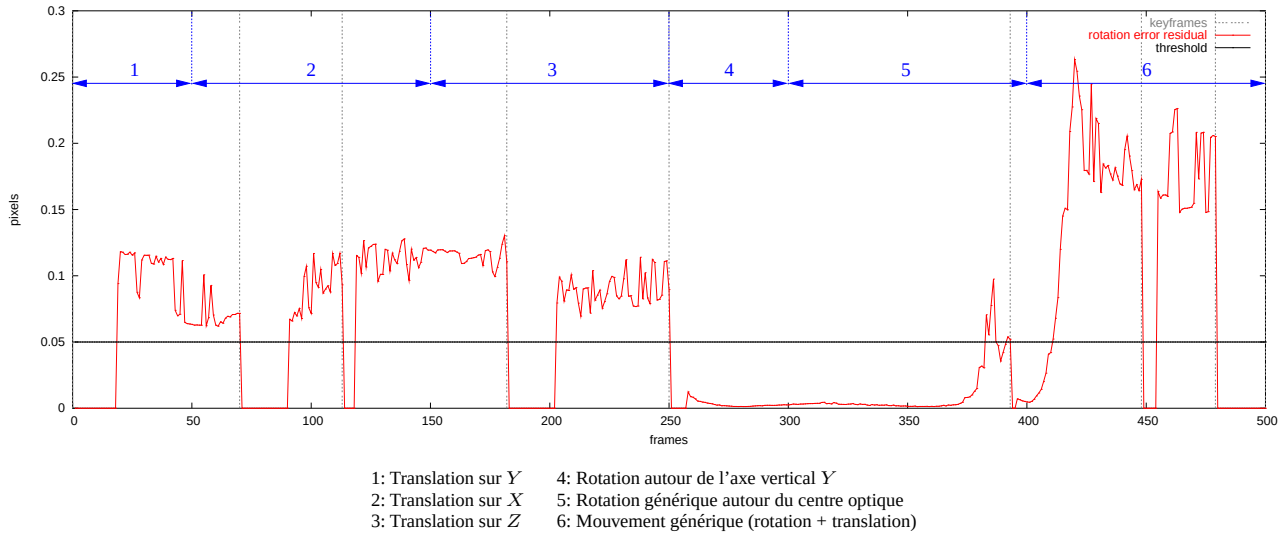
4.3 Modèles sphériques

Dans le cas d'un mouvement de rotation plus général, la scène est reconstruite comme une sphère centrée sur le centre optique de la caméra. Comme dans le cas panoramique, la texture correspondante est une mosaïque générée lors de l'anayse du GOP courant. Pour pouvoir proposer des modèles qui peuvent être codés efficacement, le maillage généré est de type géosphérique en opposition au modèle sphérique classique. Ceci se justifie par le fait qu'une sphère classique est construite sur une base longitude/lattitude, alors qu'une géosphère est construite en raffinant un polyèdre régulier (ici un icosaèdre). En conséquence, les points d'une géosphère sont distribués de façon beaucoup plus homogène le long de la surface du modèle (en particulier au niveau des pôles, voir figure 7), et la quantité d'information est moins importante pour une qualité visuelle équivalente. Comme pour les modèles cylindriques, les coodonnées de texture du modèle sont calculées de telle sorte que le rendu avec la caméra virtuelle génère des images identiques aux images originales. Une transformation des coordonnées cartésiennes en coordonnées sphériques est donc effectuée pour ces coordonnées de texture. Les transformations horizontales et verticales sont illustrées respectivement sur la figure 8, où

$$\begin{cases} \varphi = tan^{-1}(\frac{y}{x}) \\ \theta = cos^{-1}(\frac{z}{f}) \end{cases} \quad (6)$$

5 Résultats

L'approche présentée a été implémentée et testée sur plusieurs vidéos réelles ou de synthèse pour être validée. Nous montrons les résultats obtenus sur deux séquences spécifiques. La première, *archi*, est une vue synthétique d'une voiture garée près un bâtiment, dont le principal avantage est de décrire tous les différents mouvements de caméra à tester (voir figure 11). La deuxième séquence, *petite-rotation*, est une séquence réelle en extérieur, caractérisée par un



1: Translation sur Y
2: Translation sur X
3: Translation sur Z
4: Rotation autour de l'axe vertical Y
5: Rotation générique autour du centre optique
6: Mouvement générique (rotation + translation)

FIG. 9 – *Evolution du résidu de rotation au long de la séquence* archi

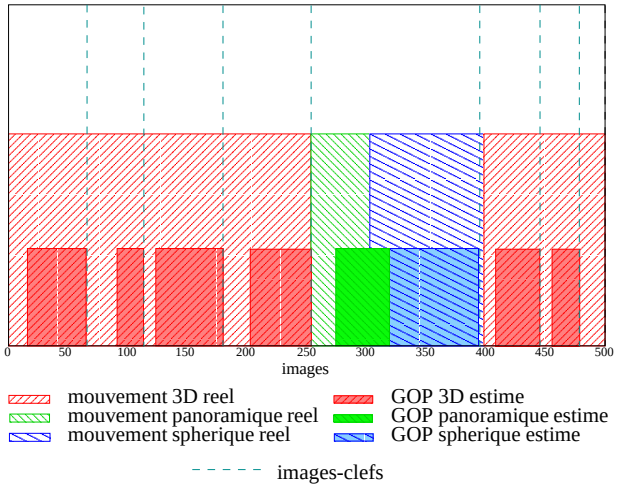


FIG. 10 – *Séquence* archi*: comparaison entre les GOPs estimés et le mouvement réel*

mouvement de type panoramique (voir figure 13(a)).

La carte de profondeur donne une information réaliste sur la géométrie de la scène. Elle permet de reconstruire les images dans le GOP avec une bonne qualité. Dans la figure 12, nous voyons la première image-clef d'un GOP 3D généré sur la séquence *archi* et sa carte de profondeur associée, calculée depuis la première et la dernière image-clef du GOP. Un rendu du modèle généré depuis un point de vue virtuel est également montré. Il est assez satisfaisant, malgré de petits artefacts comme des élongations de textures sur certaines parties du modèle.

Dans cette séquence *archi*, la caméra décrit un mouvement rotationnel entre les images 250 et 400. Notre algorithme de sélection des images-clefs a détecté un GOP sphérique entre les images 250 et 398 (voir figure 10). Nous pouvons également voir sur la figure 9 la valeur du résidu de rotation pour chaque image de la séquence. Noter que celui-ci n'est calculé que pour les images où le mouvement apparent est suffisament important, ce qui explique les brusques retours à 0 du résidu en début de GOP, ainsi que les "trous" dans l'estimation du type de GOPs sur la figure 10. Sur la figure 14 nous présentons le maillage géosphérique correspondant avec sa mosaïque associée. La différence sur la figure 14(c) entre l'image originale et l'image reconstruite (visible en particulier près des lignes à fort contraste) est due en grande partie à un effet d'anti-aliasing introduit par OpenGl dans l'image reconstruite.

Dans la séquence *petite-rotation*, puisque la caméra est fixée sur un trépied, elle décrit un mouvement de rotation presque pur. Comme on pouvait s'y attendre, l'algorithme de sélection des images-clefs n'a détecté qu'un seul et unique GOP panoramique, illustré par la mosaïque correspondante sur la figure 13(b) et par le maillage panoramique généré (figure 13(c)).

6 Conclusion et travaux futurs

Nous avons présenté ici une méthode permettant de représenter des vidéos au contenu fixe mais quelconque sous forme de modèles 2D ou 3D en fonction du mouvement de caméra décrit. Nous avons donc étendu un schéma existant au cas des rotations pures quelconques et un algorithme de sélection d'images-clefs a été proposé. De plus, les test effectués sur des séquences réelles ou de synthèse ont validé cette approche. Cette méthode pourrait être améliorée de différentes façons. L'estimateur de mouvement par maillage pourrait être adapté pour calculer des transformations homographiques entre les images plutôt que des transformations affines, pour produire des mosaïques sphé-

riques correctes. Ceci est lié au fait qu'un mouvement projeté sur une sphère n'est pas globalement affine. Nous souhaitons également étudier une représentation où il n'y aurait pas une distinction claire entre GOPs 3D, panoramique et sphérique, mais plutôt une dégradation plus élégante des modèles 3D en modèles sphériques en cas de rotation pure.

Références

[1] P.-L. Bazin, J.-M. Vézien, and A. Gagalowicz. Shape and motion estimation from geometric primitives and parametric modelling. In *Proceedings of the Machine Vision Applications workshop*, Tokyo, 2000.

[2] S. Coorg and S. Teller. Spherical mosaics with quaternions and dense correlation. *International Journal of Computer Vision*, 37(3):259–273, 2000.

[3] F. Galpin. *Représentation 3D de séquences vidéo; Schéma d'extraction automatique d'un flux de modèles 3D, applications à la compression et à la réalité virtuelle*. PhD thesis, Thèse de doctorat en Informatique, Université de Rennes 1, France, 2002.

[4] F. Galpin, R. Balter, L. Morin, and S. Pateux. Efficient and scalable video compression by automatic 3d model building using computer vision. In *Picture Coding Symposium, PCS'2004, San Francisco, USA*, 2004.

[5] F. Galpin and L. Morin. Sliding adjustment for 3d video representation. *EURASIP Journal on applied Signal Processing - Special issue on 3D Imaging and Virtual Reality, volume 2002, No. 10*, pages 1088–1101, 2002.

[6] C. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings of the 4th Alvey Vision Conference*, pages 147–151, 1988.

[7] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, ISBN: 0521623049, 2000.

[8] K. Kanatani. Geometric information criterion for model selection. *Int. J. Comput. Vision*, 26(3):171–189, 1998.

[9] K. Kanatani. Model selection for geometric inference. In *Proceedings of the 5th Asian Conference on Computer Vision (ACCV 2002)*, pages xxi–xxxii, Melbourne, Australia, January 2002.

[10] G. Marquant, S. Pateux, and C. Labit. Mesh-based scalable image coding with rate-distortion optimization. In *Image and Video Communications and Processing 2000*, volume 3974, pages 101–110, San Jose, USA, 2000.

[11] E. Morillon, R. Balter, L. Morin, and S. Pateux. 2d/3d hybrid modeling for video sequence. In *Wiamis 2004,International Workshop on Image Analysis for Multimedia Interavtive Services,april 2004*, 2004.

[12] S. Pateux. Estimation de mouvement par maillages actifs - application au codage vidéo. rapport technique projet cohrainte. Technical Report MSR-TR-97-23, IRISA, 2001.

[13] M. Pollefeys, M. Vergauwen, K. Cornelis, J. Tops, F. Verbiest, and L. Van Gool. Structure and motion from image sequences. In *Proc. Conference on Optical 3-D Measurement Techniques*, pages 251–258, Vienna, October 2001.

[14] F. Prêteux and M. Malciu. Model-based head tracking and 3d pose estimation. In *Visual Conference on Image Processing*, pages 94–110, San Jose, California, 1998.

(a) Image-clef (b) Carte de profondeur correspondante

(c) Point de vue virtuel

FIG. 12 – *Un exemple de modèle 3D généré*

[15] H. Shum and R. Szeliski. Panoramic image mosaics. Technical Report MSR-TR-97-23, Microsoft Research, 1997.

[16] R. Szeliski. Image mosaicing for tele-reality applications. In *Proceedings of the Second IEEE Workshop on Applications of Computer Vision*, pages 44–53, 1994.

[17] C. J. Taylor, P. E. Debevec, and J. Malik. Reconstructing polyhedral models of architectural scenes from photographs. In *Proceedings of the 4th European Conference on Computer Vision-Volume II*, pages 659–668. Springer-Verlag, 1996.

[18] P.H.S. Torr. An assessment of information criteria for motion model selection. In *Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*, page 47. IEEE Computer Society, 1997.

[19] J.-F. Vigueras-Gomez, M.-O. Berger, and G. Simon. Calibration multiplanaire d'une caméra : augmenter la stabilité en utilisant la sélection de modèles. In *Journées Francophones des Jeunes Chercheurs en Vision par Ordinateur - ORASIS'2003, Gérardmer, France*, pages 147–156, May 2003.

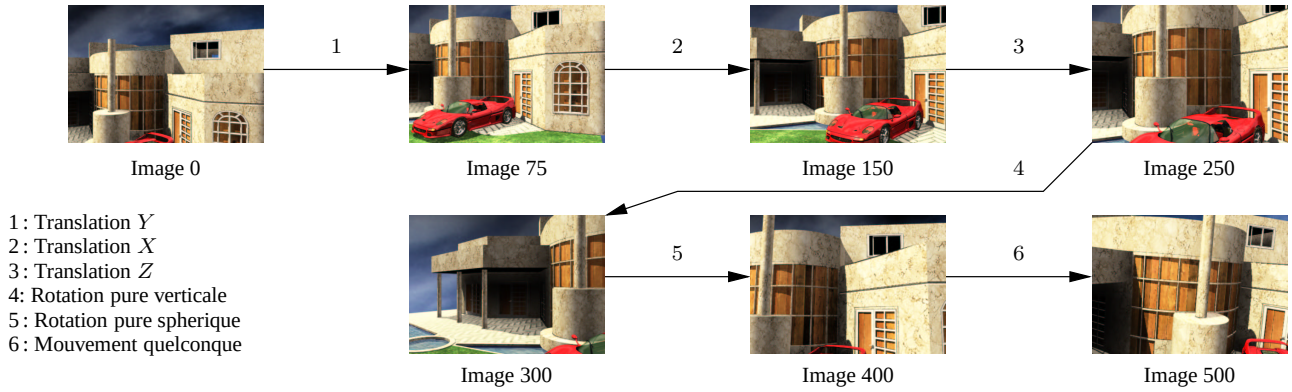
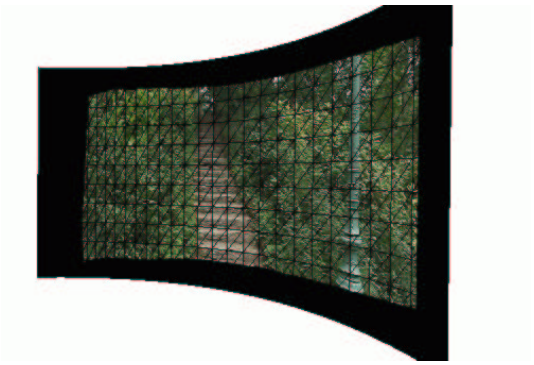


FIG. 11 – *Décompostition du mouvement de caméra sur la séquence* archi

(a) Quelques images de la séquence

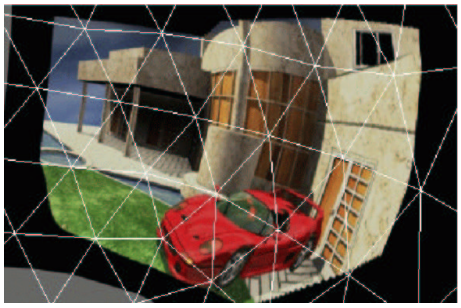
(b) La mosaïque générée, couvrant la séquence entière



(c) Modèle 3D panoramique correspondant

FIG. 13 – *Séquence* petite-rotation

(a) Mosaïque associée à un GOP sphérique



(b) Maillage 3D géosphérique correspondant

(c) Comparaison entre des images intra-GOP originale et reconstruite

FIG. 14 – *Exemple de modèle de GOP sphérique*