

2D/3D HYBRID MODELING FOR VIDEO SEQUENCE

Eric Morillon^{1*}

Raphaèle Balter^{1et2}

Luce Morin¹

Stéphane Pateux^{1*}

¹ IRISA/INRIA-Rennes
Campus de Beaulieu avenue du Général Leclerc
35042 Rennes, France

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

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

ABSTRACT

3D extraction from video gives a representation adapted to low bit-rate coding and provides enhanced functionalities. But for pure rotation motion of camera, like rotation, 3D information can not be retrieved. In this article we propose an original representation based on a 3D model stream and on mosaics. The idea of this 2D/3D hybrid approach is to give a modeling 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 motion a 3D model or a mosaic is extracted. We also present an homogeneous visualization process for this representation.

Keywords

Video Modeling, 3D reconstruction, Mosaic.

1 INTRODUCTION

1.1 Context

3D model based video coding consists of the representation of a video with one or several 3D models of the captured scene. By re-projecting these 3D models we obtain a virtual sequence similar to the original but with enhanced functionalities such as augmented reality, free view point generation or lighting changes. Furthermore, 3D model based representations are more compact than image based-ones.

3D model extraction is based on structure-from-motion and thus requires different view points of the same scene. But in particular, a camera undergoing a pure rotation does not allow the recovery of 3D information as there is no intersection between different lines of view (see Figure 1).

That is why shape from motion requires an assumption of a non pure rotation motion of the camera.

On the other hand mosaics are very well suited to represent a video obtained with rotation motion.

So here we propose an original 2D/3D hybrid method based on both 3D models and mosaics. The aim is to deal with all types of video representing a fixed scene including those with camera pure rotation motion.

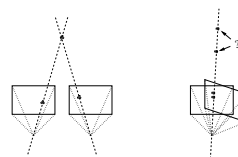


FIG. 1 – Camera translation is required for structure-from-motion, as for rotation, lines of view are the same

1.2 Previous works

1.2.1 3D Modeling

There are different ways to retrieve 3D information from videos [3]. But to meet coding requirements we do not want to make assumptions on camera parameters, scene contents or video length.

In this context Galpin proposed a method based on a 3D model stream [2] instead of aiming at a unique realistic model of the scene. Each model is valid for a portion of the original sequence called a GOF (Group of frames). These GOF are delimited with key images that are automatically extracted with three criteria based on mean motion between images, number of outgoing points and epipolar residual. For each GOF a 3D model is automatically estimated using the classic method and inter-GOF coherence is allowed by a sliding adjustment [2].

1.2.2 Mosaics

Types of mosaic can be obtained by homography, cylindrical or spherical projection [5].

Homographic mosaics are well adapted to reconstruct planar scenes and can be used in pure rotational cases. However, cylindrical and spherical mosaics are better adapted to pure rotational cases: they allow large rotations, and avoid the distortion of pictures that are far from the reference image.

2 PROPOSED METHOD

The proposed method is based on Galpin's scheme. The goal is to extract 3D structure or mosaic from usual monoscopic focal fixed video of a static scene. Our sequence is still divided into GOFs and, depending on the camera mo-

* now with France Telecom R&D

tion, a 3D model or a mosaic is estimated.

As there is no knowledge at first sight of the camera motion we have to characterize it from the video images. The result defines whether the current GOF is “2D GOF” or a “3D GOF”.

In order to obtain a homogeneous visualization process, mosaics are represented on a 3D cylinder and associated with virtual camera positions.

Figure 2 gives an overview of our algorithm. Grey rectangles correspond to a 3D model streams generation with Galpin’s algorithm. Other bricks are needed for the hybrid representation and will be detailed in the rest of the document.

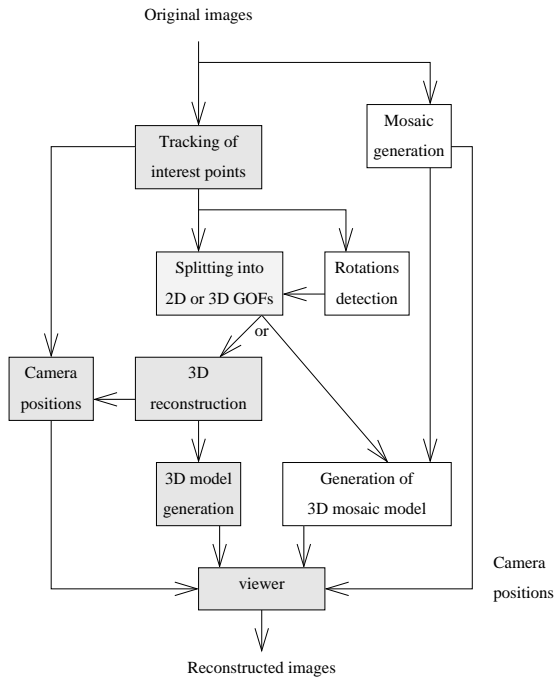


FIG. 2 – General block diagram

3 DETECTION OF ROTATIONS

This step consists of detecting 2D GOFs, which have to be represented by a mosaic, by determining whether or not the motion between the current frame and the first frame in the GOF is a pure rotation.

It uses interest points tracked during the sequence. We define a parametric motion model induced by pure camera rotation. We then estimate the model’s parameters from the tracked points. The residual indicates whether there exists a pure rotational motion of the camera compatible with the tracked points motion.

By computing the average residual between the model and the observed points, let x_i and x'_i be the tracked point in the first image and the current and $M(x_i)$ the model associated with rotation, the residual is defined as :

$$\frac{\sum_{i=1}^N \|x'_i - M(x_i)\|}{N}$$

We have tested three different models for the estimation of the transformation corresponding to a rotation : homography, cylindrical projection and spherical projection. We are going to introduce each method and compare them.

3.1 Estimation using homography

The transformation between two pictures captured with the same perspective camera undergoing pure rotation is an homography. The first method consists of estimating the optimal 2D homography between the two sets of interest points. Homographies have eight degrees of freedom and may be estimated with the help of four pairs of points. We estimate it from all given points by a linear minimization based on singular value decomposition.

3.2 Estimation using cylindrical projection

The principle is to project the picture in a space where the motion producing a pure rotation of the camera is a simple motion, whose estimation is easier.

The picture plane is thus projected on a cylinder with a vertical axis through the optical center of the camera. We then estimate the motion of the points at the cylinder’s surface.

The cylindrical projection assumes that the camera undergoes rotations around the vertical axis (perpendicular to the picture plane). In this case, the motion between the projected pictures is reduced to a horizontal translation, with only one degree of freedom. Its least square estimation is simply given by the average of the horizontal motion.

The computation of the geometrical residual is still done in the plane of the original picture, using retro projection of the points from the cylinder to the pictures, after estimating the translation.

3.3 Estimation using spherical projection

The principle of spherical projection is similar. It has the advantage of allowing any rotation axis.

By projecting the pictures onto a sphere, we obtain two 3D point sets linked by a rotation in 3D space with an axis going through the optical center of the camera.

The optimal rotation between the two clouds of points is estimated using unit quaternions representation, which provides a direct efficient solution via the resolution of a linear system [1].

3.4 Comparison of the different criteria

Tests on synthetic data show that the residual linearly increases with the amplitude of the translation and remains zero during pure rotation.

Here we present the obtained results on three real sequences. The *stairs* sequence corresponds to a translation along x-axis. The sequence called *Thabor* was captured with a tripod and is close to pure rotation.

During the *cloitre* sequence, the camera undergoes a translation, then a rough rotation, then a translation again.

Figures 3 and 4 show the evolution on the real sequences of the tree criteria : the one based on homography, and those based on cylindrical and spherical projection. A new GOF

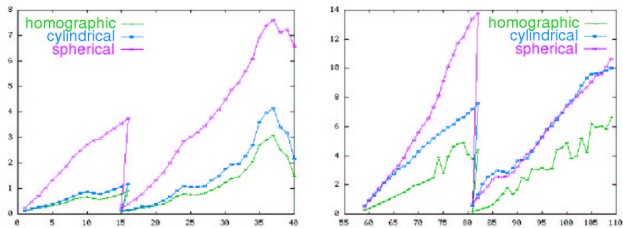


FIG. 3 – *Thabor* sequence residual

FIG. 4 – *stairs* sequence residual

starts at image 15 and 80 respectively. For both sequences we observe a regular increase of the criteria for two successive GOFs.

This is expected in the case of translation. In the case of the rotation, it comes from the fact that rotation is not really pure. The little part due to translation is detected by the criterion and is growing during the sequence.

Even if the growing speed is not of the same order, these criteria can not be directly exploited because there is no threshold distinguishing the two types of motion.

We thus have to consider the residual relative to global image displacement. A quasi pure rotation produces a large image displacement and a small residual showing the small translation contribution. A small pure translation produces a small residual, but also a small image displacement. Thus it is the comparison between the residual and the image displacement which characterizes the 3D motion type. We thus propose to use the relative residual : $\frac{\text{rotation residual}}{\text{total displacement}}$ as our new criteria.

Figures 5 and 6 show the evolution of these criteria on the same sequences. In the case of the criterion based on the spherical projection the threshold of 0.2 allows us to distinguish between the *Thabor* rotation sequence and the *stairs* translation sequence.

So the proposed criterion has been integrated in the selection of the key pictures delimiting the GOFs, in order to determine whether the current GOF is a 3D GOF or a 2D GOF. Now that we have seen how to detect the rotation parts of the sequence, we are going to focus on the production of the mosaics to represent them.

4 MOSAICS GENERATION

4.1 Pre-treatment using projection

We use the mosaics reconstruction method proposed by Pateux [4]. It does not explicitly make the hypothesis of a particular camera movement, so it allows the treatment of the sequences corresponding to pure rotations or coupled to a small translation movement.

However, this method makes the hypothesis of a locally affine movement between the pictures, whereas the real movement is homographic. The estimation of the local affine movement allows the absorption of artifacts. To adapt our incoming data we project each picture onto a cylinder which will then be "unrolled" before being sent to the mosaicker.

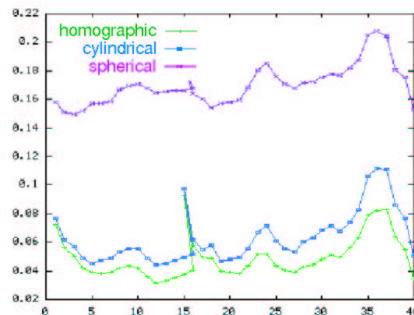


FIG. 5 – *Thabor* rotation/total displacement residual

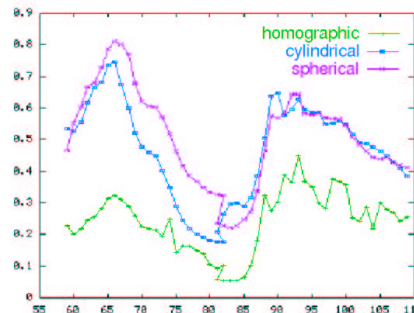


FIG. 6 – *stairs* rotation/total displacement residue

The movement between the projected pictures is a translation. This method is simple and does not make hypothesis on the size of the considered rotations, but it requires knowledge of the focal length and is limited to vertical axis rotations.

4.2 Results

We have tested the mosaic generation on the *Thabor* rotation sequence (see Figure 7).

Without pre-projection (see Figure 8) we can see an widening distortion when going away from the first image located on the right side of the mosaics. This deformation is not any more present on the resulting mosaic after projection (see Figure 9).

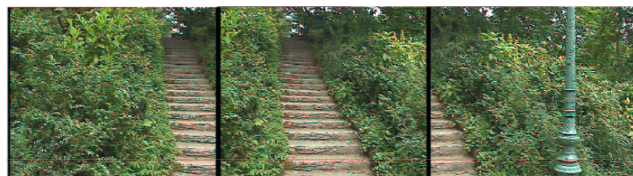


FIG. 7 – Original pictures

5 VISUALIZATION

Instead of using a classical mosaic viewer, we propose the creation of a 3D model from each mosaic, so that 3D GOFs and 2D GOFs will be visualized with the same process, i.e. by re-projection of a 3D model onto the associated camera viewpoints.



FIG. 8 – Mosaic without pre-projection



FIG. 9 – Resulting mosaic with a cylindrical projection



FIG. 10 – Original and reconstructed images.

The 3D model generated from the mosaic is a cylinder with radius equal to the estimated focal length, centered on the optical center of the first camera in the GOF, and textured with the mosaic image. For each frame in the GOF, a virtual camera position is specified so that it provides the frame reconstruction by re-projection of the 3D model onto this camera viewpoint. As the mosaic is cylindrical, the virtual camera motion is a set of horizontal rotations defined with the following parameters: image center position in the mosaic for the first image and relative horizontal rotation angle for other images in the GOF. These parameters are estimated during the mosaic generation.

PSNR is unfortunately unadapted to evaluate geometric distortion in images. We thus rather use visual quality to evaluate the quality of the reconstruction. Figure 10 shows the original and reconstructed images using projection of the cylindrical mosaic. Figures 11 and 13 show the viewer interface for a 3D GOF and for a 2D GOF from the *cloitre* sequence. The different windows provide different viewpoints on the 3D model and current camera position. The corresponding reconstructed images are shown on figures 12 and 14.

6 CONCLUSION

In this paper, we have presented a hybrid 2D/3D representation for low bit-rate coding of video sequences. This repre-

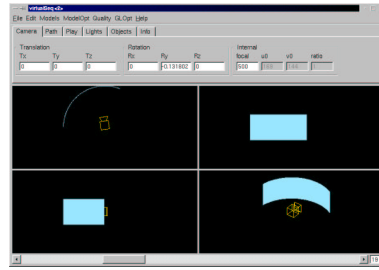


FIG. 11 – Visualization interface for a 2D GOF, for sequence *cloitre*

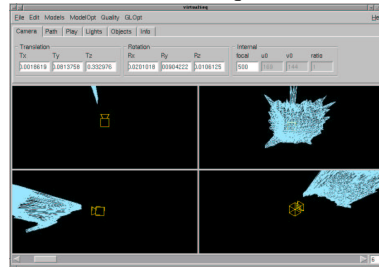


FIG. 13 – Visualization interface for a 3D GOF, for sequence *cloitre*



FIG. 12 – Reconstructed image for 2D GOF



FIG. 14 – Reconstructed image for 3D GOF

sentation is based on a stream of 3D models and mosaics and it benefits from 3D modeling functionalities when 3D information is available, and can also process and model sequence parts where 3D information can not be recovered. However the proposed representation is limited to rotations along a vertical axis. It could be interesting to extend this method in the context of a camera undergoing general rotations by the use of a spherical model.

REFERENCES

- [1] Olivier Faugeras. *Three Dimensional Computer Vision, a geometric viewpoint*. The MIT Press, Cambridge, 1993.
- [2] Franck Galpin and Luce Morin. Sliding adjustment for 3d video representation. *Eurasip Journal ASP, special issue on Signal Processing for 3D Imaging and Virtual reality*, 2002.
- [3] M. Pollefeys, M. Vergauwen, F. Verbiest, K. Cornelis, and L. Van Gool. From image sequences to 3d models. In *Third International Workshop on Automatic Extraction of Man-made Objects from Aerial and Space Images*, 2001.
- [4] G. Marquant S. Pateux and D. Chavira-Martinez. Object mosaicking via meshes and crack-lines technique. application to low bit-rate video coding. In *Picture Coding Symposium 2001*, 2001.
- [5] H. Shum and R. Szeliski. Panoramic image mosaics. Technical Report MSR-TR-97-23, Microsoft Research, 1997.