ESTIMATION OF NON-PLANAR ROTATION FOR VIDEO CODING APPLICATIONS

M. García and H. Nicolas

IRISA/INRIA, Campus de Beaulieu
35042 Rennes Cedex, France
e-mail:mgarciav,hnicolas@irisa.fr

ABSTRACT

This paper presents a new method for the estimation of non-planar rotations, i.e. rotations around axis parallel to the image plane, in the context of video compression applications. This method is based on a non planar rotation model which assumes that the moving object as a planar surface. The proposed block-based motion estimation approach is performed between consecutive or non-consecutive images, which may contained large displacements, and aims at minimizing the motion compensation error. The efficiency of the method has been compared to the results obtained with the classical full search block matching approach. Experimental results have been done on real video sequences. These results show a significant gain in term of PSNR compared to the classical full search block matching approach, while the coding cost of the additional motion information is very low, which demonstrates the interest of the proposed rotation model.

1. INTRODUCTION

The motion estimation and compensation technique has proven to be effective to exploit the temporal redundancy of video sequences, and is therefore a central part of most of the video compression schemes, and, in particular, to the compression standards MPEG-2, MPEG-4 and the H.26x video compression algorithms [1]. These video compression standards are based on a block based hybrid coding concept, which was extended to support arbitrary shaped video objects within MPEG-4. However, these compression standards allow only the use of a translational motion model. As a consequence, the block-based motion estimation can theoretically compensate only 2-D translational displacements. In practice, it may be possible to compensate more complex motions, such as zoom or rotations, if the amplitude of the motion is low or if the texture is homogeneous in the considered block. Many papers have proposed and developed efficient algorithms for an efficient estimation of these translational parameters. Nevertheless, if the images contain large non-translational motion in textured areas, the motion compensation process may not be efficient.

In order to obtain an efficient motion compensation in areas containing non-translational displacements, different solutions have been investigated. A first approach consists in reducing the size of the blocks. This is for example the case in the under development H26L [2] compression scheme where the block size may be reduced up to 4x4 pixels. A second approach consists in the use of more complex motion models. Classically, the use of an affine model allows an efficient compensation of motion such as zoom or 2D rotations. Nevertheless, Non Planar Rotation (NPR) [5], i.e. rotations around an axis parallel to the image plane are not taken into account by such a model. Alternatively, other methods such as the control grid interpolation [3] or geometric transformation motion estimation [4] have also been developed. They are based on a warping process which allows the distortion of each block in order to warp it on the reference picture. If any kind of distortion of the blocks may be allowed, this model does not provide an explicit modeling of NPR.

In this paper, we propose a model of NPR which allows a better motion compensation efficiency when this kind of motion occurs. This model contains four motion parameters: two translational ones, one angle which defines the rotation axis, and the rotation angle.

2. NON PLANAR MOTION MODELING

In the context of block-based coding applications, the geometric model of the scene can be considered as a patchwork of rigid planar surfaces, one for each block, which can closely approximate 3D rigid bodies. Under this assumption, the 3D motion can be described by a rigid 3D motion model. This hypothesis is justified by the fact that the blocks are usually relatively small. Consequently, a representation of the 2D motion in the image plane can be easily derived by the projection of the 3D motion. Mathematically, for each point P of an object, its position at time t' can be expressed as:

$$\vec{P}' = R\vec{P} + \vec{D}, \qquad (1)$$

$\vec{D} = (D_x, D_y, D_z)^T$ is the translational displacement, and R denotes the rotation matrix which is defined as:

$$R = R_X R_Y R_Z = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos a & \sin a \\ 0 & -\sin a & \cos a \end{bmatrix}.$$
$$\begin{bmatrix} \cos b & 0 & -\sin b \\ 0 & 1 & 0 \\ \sin b & 0 & \cos b \end{bmatrix} \begin{bmatrix} \cos g & \sin g & 0 \\ -\sin g & \cos g & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

where R_X, R_Y, R_Z are the rotation matrix around X, Y, and Z axis, respectively, and α, β and γ the corresponding rotation angles. R_Z represents the 2D rotations, while R_X and R_Y are the non-planar rotations. In practice and for video compression purposes, it may be sufficient to consider only one matrix. If only the 2D rotation R_Z is considered, the model leads to the rotation parameter defines by the affine model. If we consider only one of the two non-planar rotation (R_X), and assuming that inside a block, the object is plane and a perspective projection for the camera:

$$x = f\frac{X}{Z} \text{ and } y = f\frac{Y}{Z}$$

where f denotes the focal length, and *(x,y)* the coordinates of point *P(X,Y,Z)* in the image plane, the projection of Eq. (1) in the image plane leads to:

$$x_2 - x_{g2} = \frac{(x_1 - x_{g1})\cos a}{1 - \frac{\sin a}{f}(x_1 - x_{g1})}$$
$$y_2 - y_{g2} = \frac{(y_1 - y_{g1})}{1 - \frac{\sin a}{f}(x_1 - x_{g1})}$$

where the coordinates (x_1, y_1, x_2, y_2) represents the position of each pixel at time t and t+1 respectively, and (x_g, y_g) denotes the gravity center of a block. It should be pointed out that the use of the gravity center as a reference point means that this gravity center is considered to be located on the rotation axis. If it is not the case, a translation needs to be added. Furthermore, the motion estimation is performed on small blocks (8x8 or 16x16 pixels), it is therefore possible to neglect the perspective term. Finally, a model combining a translation -defined by the two translational terms (t_x, t_y) - and a NPR can be obtained as follows:

$$x_2 - x_{g2} = t_x + (x_1 - x_{g1})\cos a$$
$$y_2 - y_{g2} = t_y + y_1 - y_{g1}$$

According these equations, the displacement can be computed if the two translational terms and the rotation angle a are estimated. Nevertheless, it is assumed that the rotation is done around the X axis. This non-planar rotation model can easily generalized to any axis Φ parallel to the image plane as follows:

$$\begin{pmatrix} x_2(f) \\ y_2(f) \end{pmatrix} = \begin{pmatrix} \cos f & -\sin f \\ \sin f & \cos f \end{pmatrix} \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$$

where $x_2(f)$ and $y_2(f)$ denotes the coordinates of pixel (x_2, y_2) in the coordinate system $(\Phi, \Phi^{\perp})$, and f is the angle between the X and Φ axis.

Finally, the four motion parameters (t_x, t_y, f, a) have to be estimated for each block.

3. MOTION ESTIMATION

In order to validate the non planar rotation model proposed in the previous section, a comparison of the efficiency of this model with the classical translational one has been done. For that purpose, the translational parameters have been estimated using a full search block matching algorithm (and the translational model) in order to obtain the better possible result, in term of minimization of the Mean Square Reconstruction Error (MSRE). For the non-planar motion estimation model, four parameters have to be estimated. It is therefore not reasonable, from a computational complexity point of view, to perform a full search on these four parameters. A sub-optimal approach has therefore to be defined. Furthermore, an efficient estimation of the rotation parameters f and a can be obtained only if the translational parameters have been previously obtained. This is due to the fact that the rotation is considered to rotate around the block gravity center. As a consequence, the estimation method proposed here performed into the three following stages:

1. **Rough estimation of the translational parameters** using a full search block matching algorithm. This first estimation is performed on a sub-sampled image (by a factor 2 in each direction) in order to get a rough and fast estimation of the translational parameters. The goal is to get a rough match between the block which should be predicted and the reference image in order to allow a correct estimation of the rotation parameters.
2. **Rough estimation of the two rotational parameters** using a full search method. The precision on the angles are fixed to an angle step of 5° in order to have a fast estimation.
3. **Refinement stage.** Once a first estimation for the four parameters have been obtained with the two first stages, a refinement stage is used to get a more precise estimation. A full search is performed on the four parameters with a maximal value of 4 for the translational parameters, and 5° for the rotation angles. The final precision is fixed to half-pixel for the translation parameters, and 1° for the angles.

3. EXPERIMENTAL RESULTS

Experimental test were performed in order to assess the performance of the presented method in sequences containing non planar rotations. Figure 1 shows some original frames from the tested sequences: "*Tai*", "*Foreman*" and "*Car*" sequences (CIF format). In *Tai* sequence the head has a non-planar rotation of around 180 degrees along the sequence. In the Foreman sequence, the camera has a panoramic displacement, which generates a non-planar rotation of the scene. Finally, the *Car* sequence shows a rigid non-planar rotation for the car composed with a moving camera (translational + zoom).

Experiments were carried out using 8x8 and 16x16 blocks. The maximal search range was set to ± 32 pixels for the translational motion and ± 90 degrees for the rotation angles. The quality gain, in term of PSNR, is about 0.8-1 dB for 8x8 blocks, and of 1.2-1.5 dB for 16x16 blocks, compared to the full search block matching algorithm. The gain is logically higher for larger blocks since the efficiency of the translational model decreases with the increase of the block size. Figure 2 shows for each image of each sequence the blocks in which a gain higher than 2dB is obtained. In a general way, a significant gain is obtained on rotating and textured for which the translation model are not efficient.

In term of coding cost, the overhead generated by the model is low, mainly if the model is used only when the gain is significant. This means that a flag must indicate to the decoder which model has been used. For 16x16 blocks, it represents less than 0.004 bit/pixels.

4. CONCLUSION

This paper proposes a four parameters model of non-planar rotation and its use in the context of motion compensation for video compression applications. It has been shown that a significant gain can be obtained, in term of PSNR, compared to the use of a translation model in sequences containing non planar moving objects, or having a rotating camera. The main perspectives of this work are related to the use of this model to improve the interpolation process existing in B images. Furthermore, an adaptive motion model representation, including translation, non-planar rotation and affine model, may be used to improve the motion compensation process in order to allow a block-based selection of the motion model.

5. REFERENCES

[1] http://www.cselt.it/mpeg.

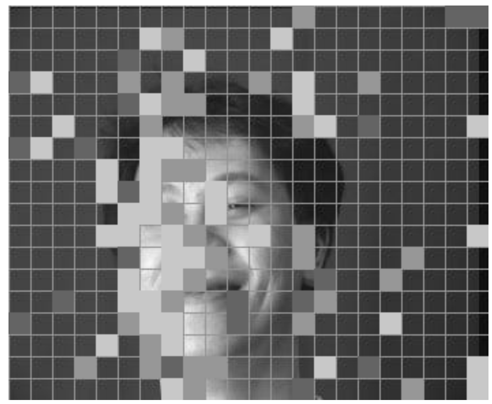
[2] www.ubvideo.com/public/h26l-white_paper.pdf

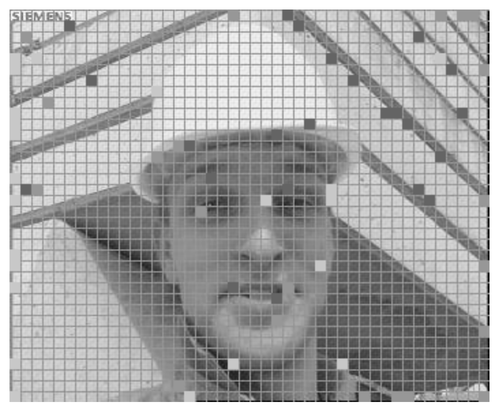
[3] G.J. Sullivan & R. L. Baker, “Motion compensation for video compression using video grid interpolation”,Proc. of ICASSP 91, pp.2713-2716,Toronto, May. 1991.

[4] S. Maciel de Faria, Very low bit rate video coding using geometric transform motion compensation, thesis, Univ. fo Essex 1996.

[5]M. Garcia, H. Nicolas. Video Object trajectory Analysis. Proceedings of ICIP 2002, Rochester,USA

Figure 1. Original frames from sequences Top to bottom: Tai(a), Foreman(b) and Car (c).





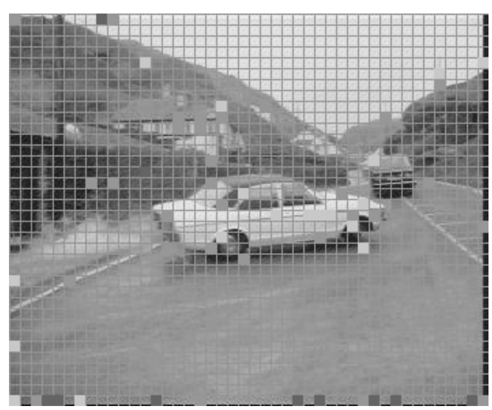


Figure 2.The marked blocks shows where the NPM model gives a gain of more than 2dB compared to the transnational model.