Analyzing video object motion focusing on non-planar rotation for two video applications

Abstract

In this paper an object-based non-planar rotation estimation for video analysis and video coding is presented. This method is based on a non-planar rotation model which assumes that the moving object as a planar surface. It is also assumed the considered video objects are rigid and have been previously segmented for each key instant considered. The model can be easily adjusted by choosing a suitable region of support (block or arbitrary rigid region). The presented results indicate that the proposed technique is a suitable approach for motion estimation, which can be used for applications such as MPEG-4 block based hybrid coding or MPEG-4 object based video manipulations. Experimental results have been performed on real video sequences.

1. Introduction

The MPEG-4 [1] video standard introduces the notion of Video Object Plane (VOP) in the framework of normalization. This new concept introduces the development of applications related to the compression, the manipulation, the edition and the composition of video object sequences. The motion estimation is a key technique for all these applications, primarily because the temporal correlation of intensities (and colors) in an image sequence is very high in the direction of motion. In video compression, the efficiency of the compression depends on the coding cost of the motion parameters, and the gain obtained on the reduction of the motion compensation error [2]. In the context of video post-production applications, an efficient motion or trajectory representation of video object sequences would allow to extract meaningful information about the behavior of the video objects [3]. Classically, motion estimation algorithms are able to correctly estimate displacements such as translation, divergence and rotations around an axis parallel to the optical axis of the camera, at least when the amplitude of the displacement is not too large. At the opposite, the estimation of rotation around an axis parallel to the image plane is not easy. This is due to the fact that motion model such as the affine is not able to correctly represent this kind of motion.

The aim of this paper is to show an efficient motion estimation technique, which can be used for video object manipulation applications and video object compression. To this effect, in the next sections we present how this technique can be implemented for each different application.

2. Motion analysis based on video object

For applications related to 2D natural video compositing, it is very interesting to get information related to the camera object orientation [4]. For that purpose, in our earlier work[5] we have developed –within the framework of object-based– a classification method which is based on the detection of non-planar motion existing between each couple of key video objects. In this context, we can distinguish two categories of motion according to a variation of orientation criterion. First, the planar motions that represent the relative camera/object displacements which do not modify the relative orientation object/camera (translation, rotation, divergence). At the opposite, the non-planar motions which correspond mainly to rotations around axis parallel to the image plane (eventually combined with a planar motion) even if large amplitude (compared to the object-camera distance) translation can also generate significant variations of orientation.

In order to characterize the non-planar rotation of the video objects as well as to allow an efficient semi-automatic manipulation of video object in the framework introduced above, we estimate the direction of the rotation axis and the rotation angle using geometric constraints. It is important to note that it is difficult to correctly detect non-planar rotations with small rotations angles. This is the reason why the temporal distance between the key VOP may be high, with the risk to have more noise or illumination variations which make the motion compensation less efficient.
2.1 Rotation axis estimation

The objective of this analysis phase consists in the estimation of the direction of the rotation axis \( \Phi \) of the non-planar motion. This is done assuming the following hypothesis: 1) The image projection is in perspective, 2) the physical object is a plane with constant depth and 3) the VOP turns in depth around an axis that crosses by its gravity center. Let \( A_1 \) and \( A_2 \) (\( A_1' \) and \( A_2' \), respectively) be the two areas of the VOP separated by \( \Phi \) at time \( t_1 \) \( (t_2, \) respectively). These areas are modified when the VOP is rotating, according to the following ratio:

\[
\frac{A_1}{A_1'} = \frac{A_2}{A_2'} \tag{1}
\]

If \( \Phi_\perp \) represents the axis perpendicular to \( \Phi \), then we have also:

\[
\frac{B_{1\perp}}{B_{2\perp}} = \frac{B_{1\perp}}{B_{2\perp}},
\]

where \( B_{1\perp} \) and \( B_{1\perp} \) (\( B_{2\perp} \) and \( B_{2\perp} \), respectively) represents the two areas separated by \( \Phi_\perp \) at time \( t_1 \) \( (t_2, \) respectively). The direction of \( \Phi \) is therefore computed by minimizing the following expression:

\[
\alpha = \arg \min_{\alpha=-90,90} \left| \frac{p_{1\perp} - p_{1\perp}^-}{p_{2\perp} - p_{2\perp}^-} \right| \tag{2}
\]

where \( \alpha \) represents the angle between the horizontal axis and \( \Phi \). Furthermore, when successive non-planar temporal segments are detected, the estimated axis are temporally smoothed to have a more robust estimation. Figure 1 shows that the direction of the non-planar rotation has been correctly estimated for various rotation axis of the sequences "tai" and "car".

2.2 Rotation angle estimation

The objective of this analysis phase consists in the estimation of the rotation angle about Y axis of the non-planar motion. This is done assuming the following hypothesis: 1) The image projection is in perspective, 2) the physical object is a plane with constant depth and 3) the VOP turns in depth around an axis that crosses by its gravity center. According to this hypothesis, a change in the projected horizontal position of the pixel is determined by the product of a perspective factor and a cosine factor \([6]\). The vertical position, on the other hand, is affected only by a perspective factor. The perspective factor is the same for both the horizontal and vertical coordinates of the all projected pixels of the video object. This gives the following results:

\[
\text{horizontal change} = \text{perspective factor \times \cosine},
\]

\[
\text{vertical change} = \text{cosine}, \tag{3}
\]

where a change is defined as the ratio of the present horizontal or vertical coordinate of the projected pixel to a previous coordinate. Applying the mentioned ideas, the angle of the rotation can be obtained minimizing the objective function that it is explained in the next section.

3. Motion estimation for compression based on video object

In the context of block-based coding applications, many papers have proposed and developed efficient algorithms for an efficient estimation of 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 \([7]\) 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, 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 \([8]\) or geometric transformation motion estimation \([9]\) have also been developed. They are based on a warping process that 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 non-planar rotation.

We propose a model of non-planar rotation which allows a better motion compensation efficiency when this type of motion occurs. This model contains four motion parameters: two translational, one angle which defines the rotation axis, and the rotation angle.

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:

\[
\vec{P'} = \vec{R} \vec{P} + \vec{D},
\]

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

\[
R = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & \cos \alpha & -\sin \alpha & 0 \\
0 & \sin \alpha & \cos \alpha & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
\]

where \( R_x, R_y, R_z \) are the rotation matrix around X, Y, and Z axis, respectively, and \( \alpha, \beta \) and \( \gamma \) 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 defined by the affine model. If we consider only one of the two non-planar rotation \( (R_x) \), and assuming that inside there is 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. (3) in the image plane leads to:

\[
x_2 - x_{22} = \frac{(y_1 - y_{22}) \cos \alpha}{1 - \sin \alpha (y_1 - y_{22})} \quad y_2 - y_{22} = \frac{(y_1 - y_{22})}{1 - \sin \alpha (y_1 - y_{22})}
\]

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 non-planar motion can be obtained as follows:

\[
x_2 - x_{22} = t_x + (x_1 - x_{22}) \cos \alpha \\
y_2 - y_{22} = t_y + (y_1 - y_{22})
\]

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

\[
\begin{bmatrix}
x_2(\phi) \\
y_2(\phi)
\end{bmatrix} =
\begin{bmatrix}
\cos \phi & -\sin \phi \\
\sin \phi & \cos \phi
\end{bmatrix}
\begin{bmatrix}
x_2 \\
y_2
\end{bmatrix}
\]

where \( x_2(\phi) \) and \( y_2(\phi) \) denotes the coordinates of pixel \((x_2,y_2)\) in the coordinate system \((\Phi, \Phi^\perp)\), and \( \phi \) is the angle between the X and \( \Phi \) axis. Finally, the four motion parameters \((t_x, t_y, \alpha, \Phi)\) have to be estimated for each block. In order to validate the non planar rotation model, 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 terms 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 \( \phi \) and \( \alpha \) 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 is according to 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. 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 models 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

In this paper an object-based non-planar rotation estimation for video analysis and video coding is presented. This method is based on a non-planar rotation model which assumes the moving object as a planar surface. The model can be easily adjusted by choosing a suitable region of support (block or arbitrary rigid region). The main perspectives of this work are 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