

VIDEO OBJET TRAJECTORY ANALYSIS

M.García and H. Nicolas

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

ABSTRACT

This article proposes a new method which allows the detection of variations of orientation of video objet due to rotations around axis parallel to the image plane. For that purpose, the video objet sequence is decomposed into temporal segments separated by key instants. The motion between these key instants are therefore estimated using a planar motion model. The validity of this model is tested by comparing their texture and shape. It is assumed that the considered video object is rigid and has been previously segmented for each key instant with a high quality level. If a variation of orientation has been detected, the direction of the rotation axis is estimated. Experimental results shows that this technique can classify different key views according to modification of camera/object orientation, and have an accurate estimation of the rotation axis with large motion.

1. INTRODUCTION

Effective motion analysis is an essential part of digital video processing. So it is not surprise that over the last twenty years the determination and analysis of the block or region based 2D motion has spurred considerable research activity [1]. Now, with the development of MPEG-4 standard [2], this research topic present a new point of view: 2D motion object-based. MPEG-4 enables contents based functionalities by introducing the concept of video object planes (VOP's). This new concept introduces the development of applications related to the compression, the manipulation, the edition and the composition of video object sequences [3]. For video compression, one of the key problems is to find the optimal motion representation which allows the compensation of the differences between the image which has to be coded and a reference one (for example the previous one). 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 [1]. 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 [4]. Classically, motion estimation algorithms are able to correctly estimated displacement 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 detection and the estimation of rotation around axis parallel to the image plane is not easy. This is due to the fact that motion model such as the affine model does not represent this kind of motion.

In this paper, we propose a new technique which allows the detection of variations of orientations generated by non-planar rotations and the estimation of the rotation axis. In practice, small rotations displacements may be assimilated to translations. In order to avoid this problem, it is therefore necessary to have larger rotations angles. This can be obtained by segmented the video object sequences into temporal segment separated by key instant (in this paper, we consider that these key instants are regularly sampled). The detection of the variations of orientations of the video object relatively to the camera is therefore performed only between two successive key instants. This allows to detect only significant rotations. This is done by testing the validity of a planar motion model. Each temporal segment is therefore classified as planar or non-planar motion using criterion based on the relative shape and texture of the two considered key VOP. Finally the direction of the rotation axes is automatically estimated in order to characterize the displacement in the non-planar temporal segments. Figure 1 illustrates the decomposition of a video object sequence. The method proposed here assume that the video object is rigid and has been previously segmented using a semi-automatic method.

2. 2D MOTION ESTIMATION

The 2D apparent displacement of video objects in an original video sequence is generated by the relative 3D object/camera displacement [5]. The two following main categories of motion can be distinguished according to a variation of orientation criterion.

Planar Motion. Planar motions represent the relative camera/object displacements which do not modify the

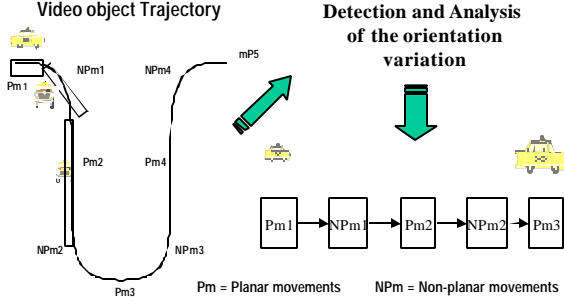


Fig.1. Decomposition of a video object trajectory

relative orientation object/camera. They corresponds to a combination of translation (low amplitude compared to the distance camera/object), divergence or rotation around an axis parallel to the image plane. Under reasonable assumptions such kinds of motion can be represented using affine or homographic motion models.

Non-planar Motion. Non-planar motions correspond to rotations around axis parallel to the image plane, to large amplitude (compared to the object-camera distance) translation, and eventually combined with a planar motion. With the exception of particular configurations (when the object can be regarded as plane or when the optical center of the camera is fixed relatively to the scene), this type of motions cannot be represented using a simple model such as the affine or homographic ones, often use in classical motion estimation algorithms.

The proposed classification method is based on the detection of non-planar motion existing between each couple of key instants. For that purpose, the validity of the planar motion model is successively tested between each key instant. If this model is not able to represent correctly the motion between these two VOP, it can reasonably be assumed that the motion between these two time instants contains a variation of orientation. The two main assumptions done to perform this test are relative to the object which is considered as rigid, and to the segmentation which is considered as perfect. The planar motion model is defined as:

$$\begin{aligned} d_x &= t_x + k(x_1 - x_g) - \mathbf{q}(y_1 - y_g) \\ d_y &= t_y + k(y_1 - y_g) + \mathbf{q}(x_1 - x_g) \end{aligned} \quad (1)$$

Where d_x, d_y are the displacement of pixel $p(x, y)$. The planar motion parameters are estimated as follows:

1-Translation estimation. If there is no variation of orientation, the centers of gravity of the successive VOP correspond to the same physical 3D point [6]. The translation can therefore be simply estimated as the difference between the coordinates of the gravity centers

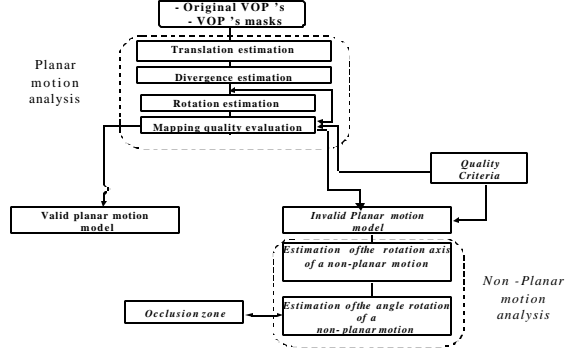


Fig.2. General outline of the proposed approach

(x_{g1}, y_{g1}) and (x_{g2}, y_{g2}) corresponding to the two considered key instants t_1 and t_2 :

$$\begin{aligned} t_x &= x_{g1} - x_{g2} \\ t_y &= y_{g1} - y_{g2} \end{aligned} \quad (2)$$

2 - Divergence estimation. The scaling factor k is estimated by comparison of the VOP surfaces at times t_1 and t_2 as follows:

$$k = \sqrt{\frac{S(VOP_2)}{S(VOP_1)}} - 1 \quad (3)$$

where $S(VOP)$ represents the surface of the VOP, and where $S_2 > S_1$ (this is only introduced to facilitate the notation).

3 - Rotation estimation. The estimation of the 2D rotation angle is done using the Hough transform [7]. The main principle of this method is the determination of the general orientation of the object between times t_1 and t_2 . It is done by comparing, for the two considered time instants, the orientation of each VOP. This orientation is defined by a line L crossing the VOP gravity center, and for which the distance between the two contours points (l and l') included in it is maximal. The orientation of the line L , defined by \mathbf{q}_L , can be represented as follows:

$$\mathbf{q}_L = \arg \max Dist[l(\mathbf{q}), l'(\mathbf{q})]$$

where function $Dist$ represents the Euclidian distance. The estimated rotation angle \mathbf{q} is then the difference between the angles \mathbf{q}_{L_1} and \mathbf{q}_{L_2} . Finally, we have

$$\mathbf{q} = \mathbf{q}_{L_1} - \mathbf{q}_{L_2} \quad (4)$$

3. CLASSIFICATION METHOD

The classification method is based on the comparison of the two considered VOP after motion compensation. The general principle is illustrated in Figure 2. Such a comparison requires the definition of a similarity criterion.

Classically the Mean Square Error (MSE) is used to compare two images, as it is for example done for video compression applications. Nevertheless, the quality of the similarity measure provided by this criterion is not sufficiently robust. Effectively, for a given pixel, it is important to know only if it is well compensated or not. The amplitude of the compensation error is not significant by itself. The MSE allots an important weight to the pixels having a very large compensation error. A weak minority of badly compensated pixels is then susceptible to give a high MSE, and consequently, to invalidate the model, even if these badly compensated pixel comes from small motion estimation errors. To avoid this problem, a solution consists of giving a binary label to each pixel specifying if it is correctly compensated or not. If the attribution of this label is carried out by comparison of the error to a given quality threshold S_L , the expression of the evaluation criterion is:

$$C_1 = 100 \frac{1}{N} \text{Card}\{p / \text{MSE}(p) < S_L\} \quad (5)$$

Where $\text{MSE}(p) = [\text{VOP}_1(p) - \overline{\text{VOP}_2}(p)]^2$, $\overline{\text{VOP}_2}$ is the compensated VOP_2 . C_1 represents the percentage of the well compensated pixels and the N the common area between the two VOP . Moreover, pixels being near to a contour or in a textured area are likely to have a high MSE due to small motion estimation errors. To reduce this effect, it is possible to perform a local motion estimation. The quality criterion can then be written as:

$$C_2 = 100 \frac{1}{N} \text{Card}\{p / \text{MSEV}(p) > S_1\}$$

With $\text{MSEV}(p) = \min_{q \in V(p)} |\text{VOP}_1(p) - \text{VOP}_2(p+q)|$
 $V = (x, y) / x \in [-x_0, x_0], y \in [-y_0, y_0]$ is the search window ($\text{MSEV}(p) \leq \text{MSE}(p)$). Since the non-covered areas between the two VOP are relevant to estimate the degree of similarity between them, the similarity criterion can be modified as follows:

$$C_3 = 100 \frac{1}{N_T} [\text{Card}\{p / \text{MSEV}(p) > S_1\} + \text{Card}\{p / p \in \text{NRZ}\}]$$

N_T is the total number of pixels (covered and uncovered areas) and NRZ the non recovering area. Finally, the geographical distribution of the non-covered points can be taken into account as follows:

$$C_4 = 100 \frac{1}{N_T} [\text{Card}\{p / \text{MSEV}(p) > S_1\} + \sum d(p)]$$

Where function $d(p)$ for $p \in \text{NRZ}$ denote the distance between p and the closest point in the recovering area. This function allows to give more weight to large regions

of pixel in the non recovering regions, which are more significant than isolated points.

The proposed method has been validated on video test sequences. The test sequences “cube”, “tai”, “car”, and “face” shows a rigid object with planar (zoom, translation) and non planar (rotation) motion with large rotation angles. Figure 4, shows the evolution of criterion C_4 versus the key VOP. It can be seen that the classification process has correctly assigned each temporal segment, the segment is classified as planar (resp. Non-planar), with $S = 20$. Nevertheless, the main limitations of the method are related to the fact 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.

4. NON-PLANAR MOTION ANALYSIS

The objective of this analysis phase consists in the estimation of the direction of the rotation axis P 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 A1 and A2 (A1' and A2', respectively) be the two areas of the VOP separated by P at time t_1 (t_2 , respectively). These areas are modified when the VOP is rotating, according to the following ratio:

$$\frac{A1}{A1'} = \frac{A2}{A2'} \quad (6)$$

If P_{\perp} represents the axis perpendicular to P, then we have also:

$$\frac{B1_+}{B2_+} = \frac{B1_-}{B2_-}$$

where $B1_+$ and $B1_-$ ($B2_+$ and $B2_-$, respectively) represents the two areas separated by P_{\perp} at time t_1 (t_2 , respectively).

The direction of P is therefore computed by minimizing the following expression:

$$\mathbf{a} = \arg \min_{\mathbf{a}=-90,+90} \left| \frac{p1_+}{p2_+} - \frac{p1_-}{p2_-} \right| \quad (7)$$

where \mathbf{a} represents the angle between the horizontal axis and P. Furthermore, when successive non-planar temporal segments are detected, the estimated axis are temporally smoothed to have a more robust estimation. Figure 3 shows that the direction of the non-planar rotation has

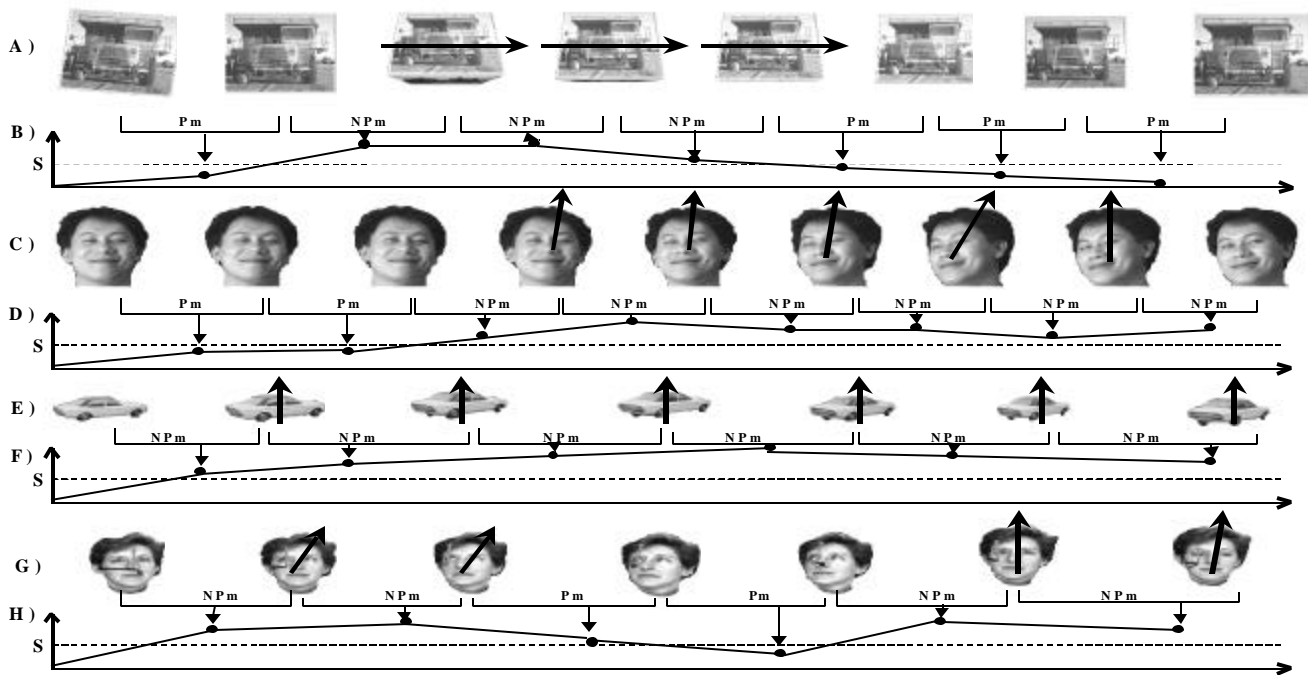


Fig3. Classification of object motion according to a variation of orientation criterion. **B,D,F,H.** Evolution of criterion C_4 versus the key VOP. **S:** Classification threshold, **Pm:** planar segment, **NPm:** non-planar segment.

been correctly estimated for various rotation axis of the sequences “cube,” “tai,” “car” and “face”.

5. CONCLUSION

This paper proposed a new method which allows to decompose the trajectory of rigid video object into temporal segments classified into two categories depending if the object displacement in each segment contains or not variations of orientations of the object relatively to the camera. These variations of orientations generally comes from rotations of the object around axis parallel to the image plane. This analysis algorithm is fully automatic, but it is assumed that the object contours have been previously segmented with a high level of quality, which is generally possible only with a semi-automatic segmentation approach. Experimental results show that good object trajectory decomposition can be obtained even for large rotation angles. Furthermore, the direction of the rotation axis can be correctly estimated. The main applications of this work are related to video object manipulation for video post-production applications for which it is interesting to know the different points of view available for each manipulated object. The main perspectives of this work is related to the estimation of the occlusion areas and of the rotation angles.

6. REFERENCES

- [1] Eric Dubois, T. Huang. Motion Estimation. IEEE Signal Proc. Magazine. March 1998.
- [2] <http://www.cselt.it/mpeg>
- [3] MPEG Requirements, MPEG-4 Applications, Doc.ISO/MPEG N2724, Seoul MPEG M., March 1999
- [4] J. Mosch and H. Nicolas. Object-based motion classification. In Proceedings of ICIP, Kobe, October 1999
- [5] G. Adiv. Determining three-dimensional motion and structure from optical flow generated by several moving objects. IEEE Trans. On Pattern Analysis and Machine Intelligence, Vol. 7, pp. 384-401, July 1985.
- [6] L. Shapiro, G. Stockman. Computer Vision Prentice Hall 2001.
- [7] C. Huang, Y. Lin Region-based Video Coding Using a Geometric Motion Compensation. Journal of Visual Communication and Image Representation 11, 279-301 (2000).