MOTION ESTIMATION BASED ON AFFINE MOMENT INVARIANTS

G. Tzanetakis, M. Traka, and G. Tziritas
Institute of Computer Science - FORTH,
P.O. Box 1385, 711 10 Heraklion, Greece
and
Department of Computer Science, University of Crete
P.O. Box 1470, Heraklion, Greece
E-mail: {tziritas}@csd.uch.gr

ABSTRACT

A method is proposed for parametric motion estimation of an image region. It is assumed that the region considered undergoes an affine transformation, which means that the motion is composed of a translation and a pure affine function of pixel coordinates. The solution of the object correspondence problem is assumed to be known. The estimation of the six motion parameters is based on the moments of the corresponding image regions. Moments up to order three are needed. For the motion computation each region is transformed to a standard position which is defined using affine invariants. The result of motion estimation is checked in the construction of a mosaic image.

1 INTRODUCTION

Parametric model motion estimation can be used for object-based video coding [7], for content-based video manipulation or for video indexing and retrieval by content. An affine motion model is often adopted because it can describe many real motions. For example, a plane in general 3-D motion under orthographic projection, an object translating at a constant depth, etc. Two families of algorithms have been proposed in the past for estimating the affine model: parameter estimation based on a previously computed motion vector field and direct estimation using gradient techniques [7]. However the gradient methods need implicit point-to-point correspondences.

On the other hand, affine moment invariants have been proposed in the past and used for pattern recognition purposes [3] [4] [6]. Therefore objects in different views resulting from affine transformations can be recognized using these invariants. I. Rothe *et al* [5] derived affine moment invariants using the normalization method, in which a standard position is defined. Using this standard position the affine transformation, and therefore an affine motion model, between two views of the same object can be determined. We use here this property for measuring an affine parametric motion model of an object, which is tracked in an image sequence. There is no need for calculation of point correspondences, since only object correspondence is required. The issue of object correspondence is not addressed here, although our method can be used for solving the matching problem, as in [1], knowing that the number of objects, or of interesting objects, in the image sequence is small.

2 AFFINE INVARIANTS FOR MOTION ESTIMATION

The aim of the proposed method is to estimate an affine parametric motion model between two corresponding image regions, without point-to-point correspondences. Let $p = (x, y)$ and $p' = (x', y')$ be two corresponding points belonging to the corresponding image regions. The affine model is given by

$$\begin{aligned} x' &= a_{10} + a_{11}x + a_{12}y \\ y' &= a_{20} + a_{21}x + a_{22}y \end{aligned}$$

or in matrix form

$$p' = c + Ap \qquad (1)$$

Therefore the vector describing the motion of all the region points is the following

$$v = c + (A - I)p$$

In such a motion straight lines remain straight lines, polygonal regions are maintained, and the convexity property is preserved.

The estimation of motion parameters will be based on the region moments; thus point correspondences are not needed

$$m_{pq} = \iint_{\mathcal{R}} x^p y^q f(x, y) dx dy$$

The moment evaluation might be limited only on the boundary points of the corresponding regions. In what follows only the region shape is taken into acoount. Thus $f(x, y) = 1$, if the point (x, y) belongs to the object, otherwise $f(x, y) = 0$. This means that no photometric information is considered in the moment computation, and therefore only geometrical characteristics determine the motion features.

Without loss of generality we assume that the origin is placed at the center of region (set of points) $\mathcal{R}$ in the initial view before the transformation. This means that: $m_{10} = m_{01} = 0$. Therefore, the translation vector c is given by

$$a_{10} = \frac{m'_{10}}{m'_{00}} \quad \text{and} \quad a_{01} = \frac{m'_{01}}{m'_{00}} \tag{2}$$

where m' means the moments of the object in the second view. Now the transformation matrix A could be estimated using the central moments, which are translation invariant. In addition, the moments are also normalized by the object surface

$$m_{00} = \iint_{\mathcal{R}} f(x,y)dxdy$$

Therefore from here all the moments are central and normalized.

The moments of the object in the second view m'_{pq} are related to the moments in the original view by the following equation

$$m'_{pq} = \sum_{i=0}^{p}\sum_{j=0}^{q} a_{11}^{i} a_{12}^{p-i} a_{21}^{j} a_{22}^{q-j} m_{i+j,p-i+q-j} \tag{3}$$

For estimating the transformation matrix A we use affine invariants similar to those described by I. Rothe *et al* [5] based on the normalization method. We use here their results for estimating the parameters of the affine model. Any object is known in two different views; the object correspondence problem is assumed as been solved. The normalization transformations are calculated separately for the two views. As we are concerned with two views of the same object, the standard position is the same. Let s denote the standard position and let T_1, T_2 be the two affine transformations, from the object position to the standard position. The following relations hold true

$$p' = Ap, p = T_1 s \quad \text{and} \quad p' = T_2 s$$

Consequently

$$A = T_2 T_1^{-1} \tag{4}$$

In Fig. 1 are illustrated two views of an object and the standard position for both. As the two views result each other by an affine transformation, the standard position is the same.

Each transformation matrix T_i $(i = 1, 2)$ is decomposed in x-shear, anisotropic scaling and rotation,

$$T = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix} \begin{bmatrix} \alpha & 0 \\ 0 & \delta \end{bmatrix} \begin{bmatrix} 1 & \beta \\ 0 & 1 \end{bmatrix} \tag{5}$$

For simplicity the index i is suppressed in the above equation.

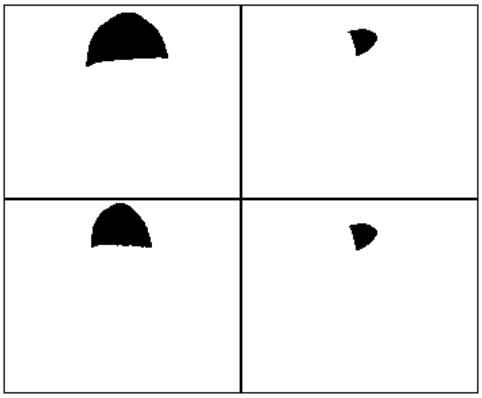

Figure 1: A pair of affine views (left) and their standard positions (right).

By using this decomposition we calculate the decomposition parameters ϕ,α,δ,β for T_1 and T_2 and use them to calculate the decomposition parameters for A (the affine transformation between the two views). The decomposition parameters are calculated by successively obtaining invariants for each of the decomposition steps (m^{sh} are x-shear invariants and $m^{\text{sc,sh}}$ are x-shear and scale invariants) :

x-shearing invariance The requirement of x-shearing invariance provides the value of parameter β for the standard position

$$\beta = -\frac{m_{11}}{m_{02}} \tag{6}$$

anisotropic scaling invariance Normalizing to scale 1, and after transforming moments according to the x-shearing invariance, we obtain the values of parameters α and δ,

$$\alpha = \frac{1}{\sqrt{m_{20}^{\text{sh}}}} \quad \text{and} \quad \delta = \frac{1}{\sqrt{m_{02}^{\text{sh}}}} \tag{7}$$

rotation invariance After transforming moments according to the above relations, the normalization for parameter ϕ gives

$$\phi = \arctan \frac{m_{30}^{\text{sc,sh}} + m_{12}^{\text{sc,sh}} + m_{21}^{\text{sc,sh}} + m_{03}^{\text{sc,sh}}}{m_{30}^{\text{sc,sh}} + m_{12}^{\text{sc,sh}} - m_{21}^{\text{sc,sh}} - m_{03}^{\text{sc,sh}}} \tag{8}$$

To strengthen the robustness of the solution we also have introduced an hierarchical method for parameter estimation. Thus at the same time the nature of the transformation is identified, and only the needed number of parameters is calculated. For selecting the appropriate motion model a distance measure is defined between the second view and the view computed using the estimated parameters. As only sets of points are considered, the Hamming distance is adopted, and therefore the total error is equal to the number of points in the two views with different value.

3 EXPERIMENTAL RESULTS

At first, in Fig. 2 we give a result on synthetic data. The synthetic motion was

$$A = \begin{bmatrix} 1.159 & 0.233 \\ -0.259 & 0.869 \end{bmatrix} \quad \text{and} \quad c = \begin{bmatrix} -5 \\ 40 \end{bmatrix}$$

The estimated motion was

$$A = \begin{bmatrix} 1.157 & 0.234 \\ -0.260 & 0.870 \end{bmatrix} \quad \text{and} \quad c = \begin{bmatrix} -5 \\ 40 \end{bmatrix}$$

The difference between the image resulting from the synthetic motion and the image resulting from the estimated motion is given in the right bottom part of Fig. 2. For all the experimental results the x (resp. y)-coordinate corresponds to the vertical (resp. horizontal) coordinate.

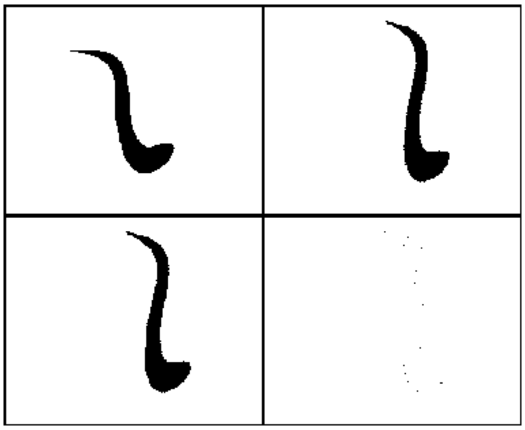

Figure 2: Top left: the original object, Top right: the affine transformed object, Bottom left: the estimated object, Bootom right: the difference after motion compensation

We have applied our method for estimating the global camera motion in the *Stefan* sequence using an affine parametric model, which could be considered as an approximation of the real camera motion. Camera operates several different movements over the whole time. For testing the method we used the estimated motion parameters for the alignment of the image frames in order to construct a mosaic or panoramic view of the scene. For each couple of frames a couple of corresponding contours is provided. In Fig. 3 is given a pair of frames in an interval of 1/6 sec, and in a part of the sequence where a rotational movement is also present. In the bottom of the same figure are given the two corresponding polygons extracted manually. For the needs of the mosaic image construction, an adjustement should be done for obtaining the global movement reported to the image center. Finally, the estimated camera motion was

$$A = \begin{bmatrix} 0.924 & 0.021 \\ 0.000 & 0.906 \end{bmatrix} \quad \text{and} \quad c = \begin{bmatrix} 1.5 \\ 108.5 \end{bmatrix}$$

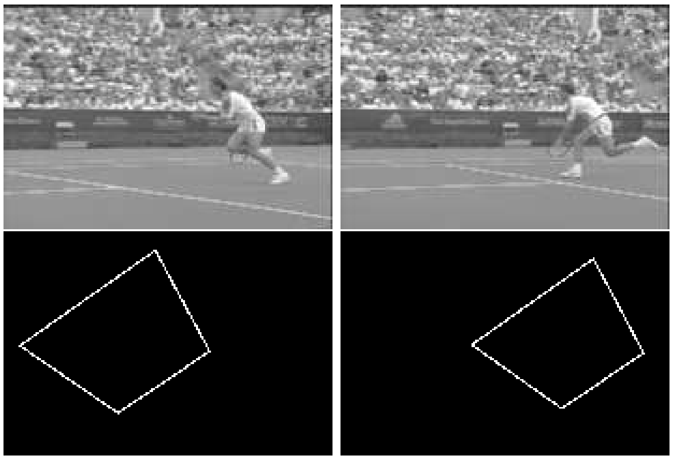

Figure 3: Top: Two frames of the *Stefan* sequence. Bottom: The two corresponding polygons.

This result shows that important movements could be estimated. The mosaic image is given in Fig. 4, where all the frames are aligned with the coordinate system of the first frame.

4 CONCLUSION

The main advantage of our method is that only region or boundary correspondence is needed, and not point-to-point. Of course, the shape of the corresponding regions should be sufficiently precise, for having results of an acceptable accuracy. The proposed method is also very fast as it proceeds to the direct computation of the parameters, based on the moment calculation. The hierarchical application of the method allows the identification of the motion model, and gets more robust results. The use of affine invariants provides also a criterion for checking the correctness of the correspondence of the two views or of the motion model.

Acknowledgement: This work is funded in part under the NEMESIS ESPRIT project.

References

[1] D. Bhattacharya and S. Sinha, "Invariance of stereo images via the theory of complex moments", *Pattern Recognition*, Vol. 30, No. 9, pp. 1373-1386, 1997.

[2] J. Flusser and T. Suk, "Pattern recognition by affine moment invariants", *Pattern Recognition*, Vol. 26, pp. 167-174, 1993.

[3] M. K. Hu, "Visual pattern recognition by moment invariants", *IRE Trans. on Information Theory*, Vol. IT-8, pp. 179-187, 1962.

[4] Y. Li, "Reforming the theory of invariant moments for pattern recognition", *Pattern Recognition*, Vol. 25, No. 7, pp. 723-730, 1992.

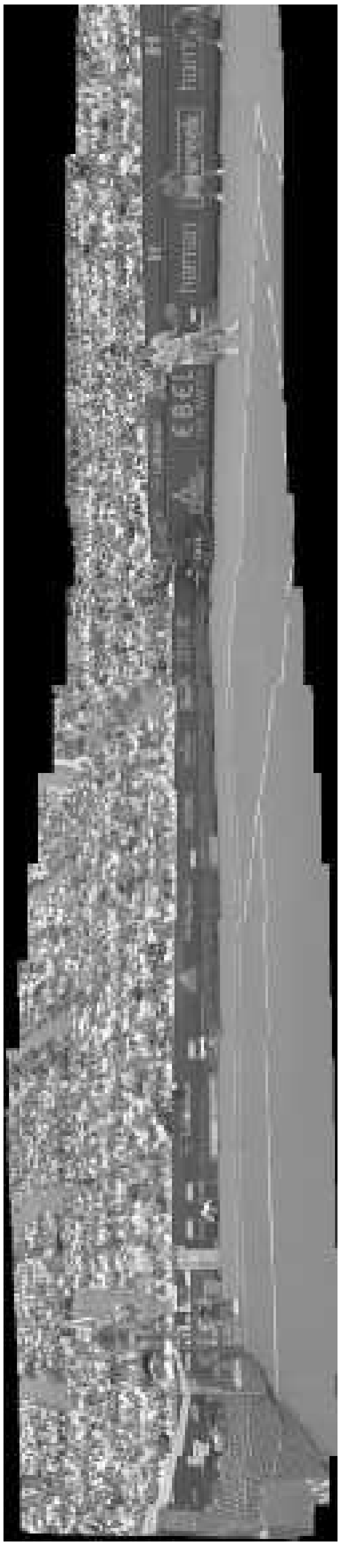

Figure 4: The mosaic image

[5] I. Rothe, H. Süsse, and K. Voss, "The method of normalization to determine invariants", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 18, No. 4, pp. 366-376, April 1996.

[6] C.-H. Teh and R. Chin, "On image analysis by the method of moments", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 10, pp. 496-513, July 1988.

[7] G. Tziritas and C. Labit, *Motion Analysis and Image Sequence Coding*, Elsevier, 1994.