Robust pan, tilt and zoom estimation

I. Grinias and G. Tziritas
Department of Computer Science, University of Crete,
P.O. Box 2208, Heraklion, Greece
E-mails: {*grinias,tziritas*}@*csd.uoc.gr*

ABSTRACT

In this paper we propose a method for estimating the motion parameters for a rotating camera with a possibly changing focal length. This method is suitable for analysing sport event videos. The estimation process consists of three stages: (1) robust estimation of the 2-D translational components, (2) 2-D block-based motion estimation and (3) robust estimation of a parametric motion model. For more reliability confidence measures for 2-D motion vectors are introduced. The reliability of the method is illustrated on some difficult real image sequences.

1 Introduction

Video content can be characterized using a camera motion description [5]. For example, in TV sports coverage camera panning means that a panoramic view is shown or that an athlete moves in a specific direction. When the camera zooms on a given player this means that the intention is to fixate on this particular player. The camera's activity is related to the most interesting action in such a video.

In many real image sequences the camera undergoes only rotational motion with the focal length possibly changing. This happens for example in videos of sport events, where the camera may rotate around either the horizontal, or vertical axis, attempting to fixate on the moving athlete. In addition, the focal length may change for zooming in on the fixated person. As seen in Section 2 this situation leads to a 2-D parametric motion model.

The most widely used method with a parametric motion model is the robust differential method [3] [6]. J.-L. Dugelay and H. Sanson [4] present a detailed analysis of the differential techniques for parametric motion estimation. Various parametric models are considered, including affine and quadratic models.

In the case of large motion and less textured images we have experimented some problems with the differential approaches. In this paper we propose a new method which is less sensitive to the amount of motion and the image texture. A block diagram of our method is shown in Figure 1. In a first stage only the global translational motion is estimated using a least median of absolute deviations method. As we do not require a very accurate measure for this initial estimation, we limit the search to a predefined scheme with 1/4 pixel precision. This initial estimation constrains the second stage which is applied at a block level. Then a sparse field of motion vectors is obtained with the desired accuracy, either by a matching technique or a differential approach. However, all blocks are not equally reliable for motion estimation. As our objective is to obtain a reliable 3-D motion model, we have to measure the confidence of the estimated vectors. We use two measures of confidence, one *a priori*, *i.e.* before the motion vector estimation, and an *a posteriori* based on the displaced frame difference. We present in detail the confidence measures in Section 4. Finally we use a robust estimation method for obtaining the model parameters.

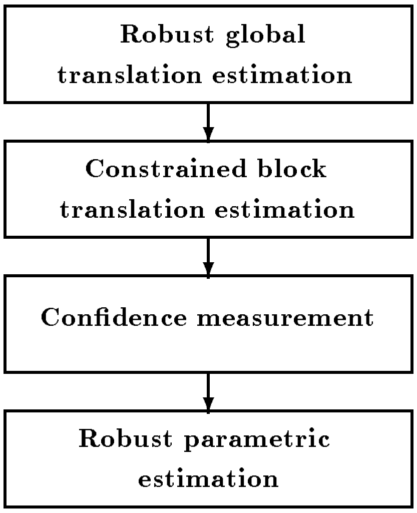

Robust global translation estimation
Constrained block translation estimation
Confidence measurement
Robust parametric estimation

Figure 1: The steps of the parameter estimation method.

2 Motion model

Since we assume only rotational motion, the 2-D motion field is not dependent on depth. Therefore, a

2-D parametric model can fully describe the 2-D motion field. Let (X, Y, Z) be a 3-D point and a perspective projection to (x, y),

$$x = f\frac{X}{Z} \quad \text{and} \quad y = f\frac{Y}{Z} \tag{1}$$

where f is the focal length. We assume that the focal length may change with a rate

$$\beta = \frac{df}{fdt}, \tag{2}$$

while the camera may undergo rotational motion around the vertical and/or horizontal axes. The 3-D velocity vector of the 3-D point will then be

$$\frac{dX}{dt} = -Z\Omega_Y, \frac{dY}{dt} = Z\Omega_X, \frac{dZ}{dt} = -Y\Omega_X + X\Omega_Y \tag{3}$$

Using Eq. (2) and (3) in the projection relations of Eq. (1) we obtain the 2-D motion field

$$u = -f\Omega_Y + \beta x + \frac{\Omega_X}{f}xy - \frac{\Omega_Y}{f}x^2 \tag{4}$$

$$v = f\Omega_X + \beta y + \frac{\Omega_X}{f}y^2 - \frac{\Omega_Y}{f}xy \tag{5}$$

Four parameters, $(\Omega_X, \Omega_Y, \beta, f)$, can be used for describing the whole 2-D motion field. However, in order to obtain linear equations, we extend the parameters to the coefficients of the above polynomials of the image coordinates. Finally, we obtain the following parametric model:

$$u = \alpha_1 + \beta x + \gamma x^2 + \delta xy \tag{6}$$

$$v = \alpha_2 + \beta y + \gamma xy + \delta y^2 \tag{7}$$

with $\frac{\gamma}{\alpha_1} = \frac{\delta}{\alpha_2} > 0$.

3 Global translation estimation

As presented in the introduction, the estimation process begins with a global translation estimation; then motion vectors are estimated at the block level; and finally, the motion parameters are computed. It is very important to have a reliable and robust initial global estimation. We are interested primarily in scenes with a single moving person or object. Therefore, independent moving objects should be rejected for estimating the camera motion. We use a robust least median of absolute deviations method for rejecting outliers [9]. However, the full implementation of a least median method results in an excessive computational cost. As the accuracy requirement is not extremely high, we limit the search to a finite set of possible displacement vectors. We thus obtain the best vector and the minimum deviation as follows:

$$D = \min\ \text{med}\ \{|I(x, y, t) - I(x - \alpha_1, y - \alpha_2, t - 1)|\}, \tag{8}$$

where the median value is determined using the histogram method.

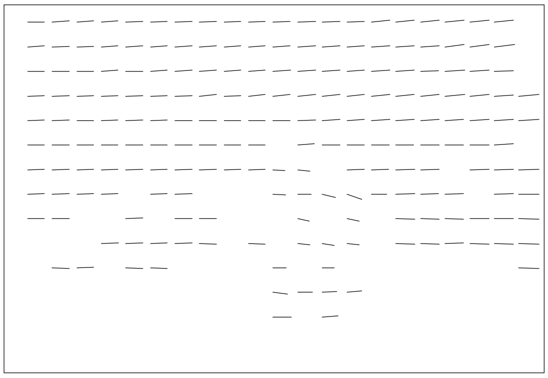


Figure 2: The estimated displacement vectors on a frame of the *Stefan* sequence.

The displacement vector estimation at the block level is initialized by this initial estimate. Since this estimate is sufficiently accurate, the search is limited to a narrow window around the initial estimate. These estimates are quite accurate for blocks taken from the background, while those from the moving part of the foreground are less accurate, but are generally rejected as described in the following. Estimated displacement fields are shown graphically in Figures 2 and 3.

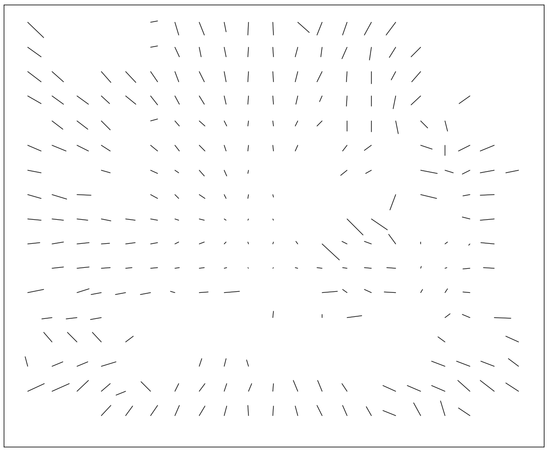


Figure 3: The estimated displacement vectors on a frame of the *Tennis table* sequence

4 Confidence measure

In the past work has been done on the reliability of the estimated displacement vector using matching techniques. Anandan [1] uses the curvature of the matching criterion, the sum of squared differences, for measuring the confidence of the correspondence. Patras [8] expresses the confidence measure in terms of the *a posteriori* probability of the motion vectors, the whole estimation framework being probabilistic.

In our work we use two confidence measures for weighting the estimated motion vectors. The first is related to the block texture and is independent of the estimation method and result. The second is based on the matching criterion.

In [11] it is proven that a unique displacement vector does not exist if the intensity function depends on only one orientation. More precisely, let $I(x, y)$ be the intensity function. It will be not possible to obtain a unique motion vector if $I(x, y) = f(c_0 + c_1 x + c_2 y)$. In this singular case all the gradient vectors are parallel. This totally ambiguous situation happens when only one gradient orientation exists. For defining a confidence measure let us consider the following gradient matrix:

$$G = \left[\begin{array}{cc} \sum I_x^2(x,y) & \sum I_x(x,y) I_y(x,y) \\ \sum I_x(x,y) I_y(x,y) & \sum I_y^2(x,y) \end{array} \right] \tag{9}$$

The smaller of the two eigenvalue of matrix G can be used as the index of the block capacity for estimating the displacement. If the smaller eigenvalue is 0, the measure confidence is taken to be null. This measure is also suggested in [2]. We define the range [0,1] as being the interval of possible values for the confidence measure. On the other hand, the confidence measure should only define the relative confidence of the different blocks. Therefore, it is not necessary to have a unique mapping of the eigenvalues to the confidence measures. For this reason we assign the value 0.5 to the eigenvalue ranked at 1/3 of the total number of blocks, say λ_m. Finally, we propose a sigmoid function for measuring the confidence,

$$C_0 = \frac{1}{1 + e^{\mu\left(1 - \frac{\lambda}{\lambda_m}\right)}} \tag{10}$$

After the estimation the quality can be measured using the resulting displaced frame difference. Let δ be the average displaced frame difference for a given block. A sigmoid function is also used for measuring the confidence,

$$C_1 = \frac{1}{1 + e^{\delta - \delta_M}}, \tag{11}$$

where δ_M is a reference level for the displaced frame difference. It could be possible to compute an adaptive reference level based on the observation of the whole result on a given frame. However, it will be faster to give a generic rule for determining the reference level.

The reference level for the absolute value of the displaced frame difference will be fixed as being the maximal expected. This means that even in the case of maximum expected deviation the confidence remains at 0.5. Thus blocks well compensated are taken into consideration with confidence measure very close to 1. Let us now consider the expected value of the absolute displaced frame difference,

$$E\{|d|\} = E\{|I(x,y,t) - I(x - \hat{u}, y - \hat{v}, t-1)|\} =$$

$$E\{|I(x,y,t) - I(x-u, y-v, t-1)\}$$
$$+(u - \hat{u})I_x + (v - \hat{v})I_y|\}$$

$$\leq E\{|I(x,y,t) - I(x-u, y-v, t-1)|\}$$
$$+E\{|u - \hat{u}|\}|I_x| + E\{|v - \hat{v}|\}|I_y|$$

where (u, v) is the real motion vector and $(\hat{u}, \hat{v})$ is the estimated motion vector. The estimation error on the motion vector is assumed to be proportional to the real motion vector, which is unknown and for this reason replaced by the estimated one. We propose to use the following inequalities

$$E\{|I(x,y,t) - I(x-u, y-v, t-1)|\} \leq 3D,$$

where D is taken from Equation (8). Finally, we have

$$E\{|d|\} \leq 3D + r(u|I_x| + v|I_y|).$$

Concerning a block, we use the averages of the two absolute gradient components on the whole block. We obtain then

$$\delta_M = 3D + r(\hat{u}\overline{|I_x|} + \hat{v}\overline{|I_y|}) \tag{12}$$

The final confidence measure results from the product of the two above defined measures. The equations for parameter motion estimation are weighted by

$$W = C_0 C_1 \tag{13}$$

In Figure 4 the results of confidence measure on frames of the *Stefan* and *Tennis table* sequences are shown. The darker a block is, the stronger the confidence value is for this block. The more transparent a block is, the lower the confidence measure is for this block. It is seen that homogeneous blocks, or blocks with colinear gradient vectors, or independently moving blocks are not taken into consideration for the motion parameters estimation.

5 Motion parameter estimation

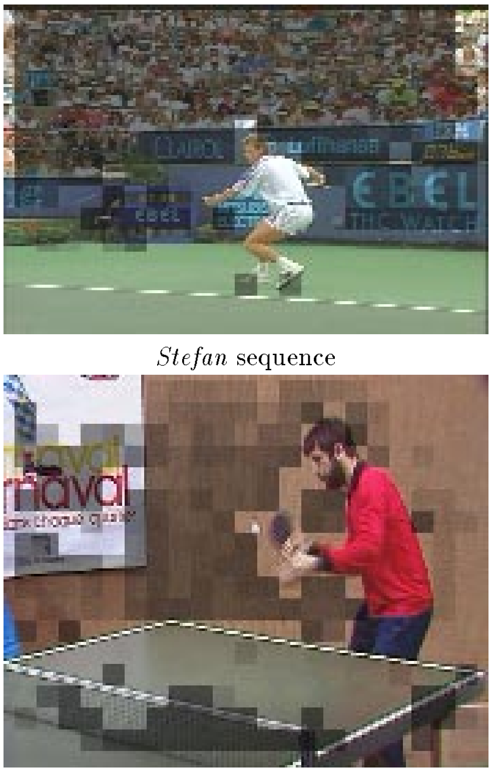
We now propose to use a robust M-estimator [7] for estimating the model parameters. The M-estimation criterion which should be minimised is

$$\sum W_k \rho\left((\hat{u}_k - u_k(\theta))^2 + (\hat{v}_k - v_k(\theta))^2\right),$$

where $\rho(\cdot)$ is defined as

$$\rho(x) = \begin{cases} 1, & |x| \leq T \\ 0, & |x| > T \end{cases}$$

The threshold T is set proportional to the motion parameters. The summation is taken over all estimated motion vectors, while θ contains all the unknown parameters. An iterative least squares method is employed for solving the above minimisation problem, leading to iterations of linear systems of equations. In Figure 5 the resulting horizontal (or panning) global displacement for the *Stefan* sequence is shown. Only 3 parameters $(\alpha_1, \alpha_2, \beta)$ were considered for obtaining the global alignement. A comparison is also shown (dashed line) with a global displacement resulting from an estimated affine parametric model [10].



Stefan sequence

Tennis table sequence

Figure 4: The confidence measures on blocks.

6 Conclusions

In this paper we have proposed a new technique for estimating the motion parameters in the case where the camera undergoes panning and tilting, while the focal length may change. As the camera does not translate, a pure parametric model must be estimated, not depending on depth. We have developed and tested a fast and robust method based on block matching, confidence measure computation and M-estimation. We have obtained very good results on difficult sport image sequences.

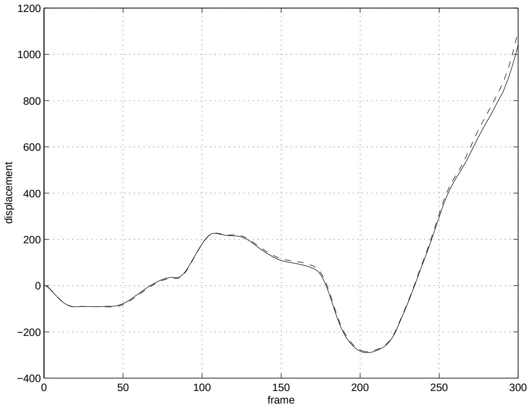


Figure 5: The global horizontal diasplacement for the *Stefan* sequence.

Acknowledgment: This work has been funded in part by the European IST “Personalised Immersive Sports TV Experience” project.

References

[1] P. Anandan. Computing dense displacement fields with confidence measurement of visual motion. In *Proc. in SPIE Conference on Intelligent Robots and Computer Vision*, pp. 184–194, 1984.

[2] J.L. Barron, D.J. Fleet, and S.S. Beauchemin. Performance of optical flow techniques. *Intern. J. of Computer Vision*, 12:43–77, 1994.

[3] M. Black and P. Anandan. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer Vision and Image Understanding*, 63(1):75–104, Jan. 1996.

[4] J.-L. Dugelay and H. Sanson. Differential methods for identification of 2-d and 3-d motion models in image sequences. *Signal Processing: Image Communication*, 7:105–127, 1995.

[5] S. Jeanin *et al.* Motion descriptors for content-based video representation. *Signal Processing: Image Communication*, 16:59–85, 2000.

[6] M. Hoetter. Differential estimation of the global motion parameters zoom and pan. *Signal Processing*, 16:249–265, 1989.

[7] P. Huber. *Robust statistics*. Wiley, 1981.

[8] I. Patras. *Object-based video segmentation with region labeling*. Ph.D. thesis, Technische Universiteit Delft, 2001.

[9] P.J. Rousseeuw and A.M. Leroy. *Robust Regression and Outlier Detection*. Wiley-Interscience, New York, 1987.

[10] M. Traka and G. Tziritas. Panoramic view construction. *Signal Processing: Image Communication*, 2002 (submitted).

[11] G. Tziritas and C. Labit. *Motion analysis for image sequence coding*. Elsevier, 1994.