ROBUST ESTIMATION OF MULTI-COMPONENT MOTION IN IMAGE SEQUENCES USING THE EPIPOLAR CONSTRAINT

Eckehard Steinbach [1] * *Subhasis Chaudhuri* [2] * *Bernd Girod* [1]

[1] Telecommunications Institute, University of Erlangen-Nuremberg
Cauerstrasse 7, 91058 Erlangen, Germany
{steinb,sc,girod}@nt.e-technik.uni-erlangen.de

[2] Department of Electrical Engineering, Indian Institute of Technology, Bombay
Powai, Mumbai - 400 076, India

ABSTRACT

Given two frames of a dynamic scene with several rigid body objects undergoing different motions in the three-dimensional space, we robustly estimate the motion and structure of each object. The *least median of squares* (LMedS) estimator is integrated into a robust 3D motion parameter estimation and scene structure recovery framework to deal with the multi-motion problem. Experimental results underline the capability of the approach to deal successfully with multi-component motion. We apply the approach presented in this paper to the problem of automatic insertion of artificial objects in real image sequences.

1. INTRODUCTION

The estimation of motion parameters for objects in an image sequence and simultaneous recovery of the scene structure is a very important problem in various computer vision and image communication applications. The primary focus in the literature has been on motion of a single object in the scene. However, in most practical situations the motion field is not homogeneous as there may be several objects undergoing different motions. Most of the existing motion analysis methods would fail to perform under these circumstances. Recently, there had been some attempts to identify and estimate the various motion components present in the scene [1, 2, 3, 4].

In this paper we address the following problem: *Given two frames of a dynamic scene with several rigid objects (and/or the camera) undergoing different motions in three-dimensional space, how can we robustly estimate the motion and structure of each object?* The problem simultaneously imposes the task of inherent segmentation of the scene into regions that contribute to different motion fields.

A recently published [5] robust approach for motion recovery is extended for multiple objects in this paper. The concept of *epipolar line constraint* [5] is used to recover the motion parameters and to simultaneously compute the motion disparity used to reconstruct the depth map. The *least median of squares* (LMedS) estimator [6] is used to differentiate or classify the motion field into various constituent groups.

This paper is organized as follows. First, we briefly discuss how we use the epipolar constraint to analyze a scene with a single moving object. Then we extend the approach to deal with the multi-component motion in the field of view of the camera using the proposed technique. In the first section of our experimental results we demonstrate the ability of the algorithm to separately estimate multi-component motion. Finally we use the algorithm for automatic insertion of artificial 3D objects into a real image sequence to demonstrate its application.

*This work is supported by the 'Graduiertenkolleg 3-D Bildanalyse und -synthese' at the Univ. of Erlangen-Nuremberg.

2. SINGLE COMPONENT MOTION ANALYSIS

Traditionally, structure-from-motion algorithms utilize a two-stage approach. First, feature points are extracted from the current image and their correspondence to features in the previous image is established. In a second step, rigid body motion parameters and depth values are computed from these feature point correspondences. There is no feedback from the computation of motion parameters and depth to the feature matching process, and, typically, the results are very sensitive to errors in feature correspondences [7, 8, 9, 10].

A new structure-from-motion algorithm was presented in [5] that does not separate feature matching and the 3-D motion recovery computation and thus overcomes the inherent limitations of the conventional two-stage approach. The algorithm is based on the observation that feature correspondences for a given 3-D motion are constrained to lie on a straight line (called the epipolar line) in the image [7]. The position along the epipolar line corresponds to different depth values.

For a given set of 3-D rigid body motion parameters, we can readily compute the parameters of the epipolar line for the ith feature point in the image from the well known equation [7]

$$\left[X_i' \; Y_i' \; 1\right] \; [\mathbf{E}] \; [X_i \; Y_i \; 1]^T \; = \; 0, \qquad (1)$$

with (X_i, Y_i), (X_i', Y_i') being corresponding image plane locations in two successive frames and $[\mathbf{E}]$ depending on the rotation and translation. For a given point (X_i, Y_i), the corresponding point (X_i', Y_i') must lie on the epipolar line $L_i(E)$ given by $aX_i' + bY_i' + c = 0$, where $[a \; b \; c]^T = [\mathbf{E}][X_i \;\; Y_i \;\; 1]^T$. We first divide the second image into rectangular measurement windows of fixed size (typically 7×7 or 15×15 pixels) and mean squared *displaced frame difference* (DFD) surfaces are computed with respect to the first

image. The DFD surface at a point (X_i, Y_i) is given by

$$DFD(X_i,Y_i;d_X,d_Y) = \sum_{r=-\frac{N-1}{2}}^{\frac{N-1}{2}} \sum_{s=-\frac{M-1}{2}}^{\frac{M-1}{2}} (I_1(X_i+d_X+r,Y_i+d_Y+s) - I_2(X_i+r,Y_i+s))^2 \quad (2)$$

where I_1 and I_2 are the intensity images, (d_X, d_Y) the displacement, and N, M the size of the measurement windows in horizontal and vertical directions, respectively. Please note that the computation of the DFD surfaces is very similar to the computation for full-search *block matching*. The search for correspondence is now restricted along the epipolar line over the DFD surface. The point along this line where the DFD surface has the least value is taken as the match.

Since the search for correspondence on the epipolar line requires a knowledge of the 3-D motion parameters the algorithm searches the 5-dimensional motion parameter space. The search terminates when a local minimum of the accumulated mean squared DFD values is encountered. The cost function to be minimized for F feature points (centers of measurement windows) selected in the image is given by

$$\min_E \sum_{i=1}^{F} \min_{(X_i+d_X,Y_i+d_Y)\in L_i(E)} DFD(X_i,Y_i;d_X,d_Y) \quad (3)$$

A simple conjugate direction search method is used to arrive at the solution. Now, given the motion parameters, a dense map of depth (scaled by a factor) can be recovered for all points in the scene by searching for their displacement vectors (d_X, d_Y) along the corresponding epipolar line.

3. MULTI COMPONENT MOTION ANALYSIS

Let us now consider the case where there are two objects undergoing different rigid motions, say E_1 and E_2. Obviously, the least squares minimization of the cost function as given in equation (3) does not yield good results as the epipolar lines corresponding to these motions for a particular point could be very different. While estimating the dominant motion E_1, the second motion field creates outliers that must be identified and rejected in order to obtain a good estimate of the dominant motion parameters. We propose the use of the LMedS estimator which can efficiently detect such outliers, enabling one to obtain a robust estimate of the motion parameters.

The LMedS estimator has been used in various computer vision applications, including motion analysis [11]. An extremely important property is that the LMedS estimator can tolerate up to 50% data contamination by outliers [6]. Here, one replaces the mean of the squared residuals by their median to achieve the robustness. Hence, we modify our cost function to estimate the motion parameters E_1 by

$$\min_{E_1} \left(med_{\forall i} \left\{ \min_{(X_i+d_X,Y_i+d_Y)\in L_i(E_1)} DFD(X_i,Y_i;d_X,d_Y) \right\} \right) \quad (4)$$

There is an inherent assumption in using equation (4) to estimate the motion parameter E_1. It is assumed that *at least* half of the feature points used in the estimation process belong to the object moving with motion E_1. In other words, the cardinality of the segmented region belonging to the first object when normalized with respect to the total size of the image must be greater than 0.5. Having estimated the first component of the motion, a dense map of displacement vectors for all pixels is obtained. The residual error for each feature point given by the minimum of the DFD surface along the corresponding epipolar line is sorted by their magnitude. A search is now initiated in the bottom half of the residuals to locate the break point when there is a sudden large increase in magnitude. The points above the break point are outliers and should belong to the second object undergoing a different motion E_2. Hence, we achieve an automatic segmentation of the scene based on motion parameters. The parameters E_2 are now estimated using equation (3), but the computation is restricted to the segment that belongs to the second object.

The above procedure can be very easily generalized to deal with even a larger number of motion components in the field of view of the camera, provided that the cardinality of the region belonging to the next dominant motion is larger than the cardinality of the rest of the objects, by simply reiterating the procedure on the remaining region.

4. EXPERIMENTAL RESULTS

4.1. Separation and estimation of multiple motions

A synthetic scene is used to illustrate the ability of our approach to separate and estimate multiple motions. Fig. 1 shows a screen shot of the 3D scene consisting of two objects. Texture is mapped on a planar surface (object 1) with

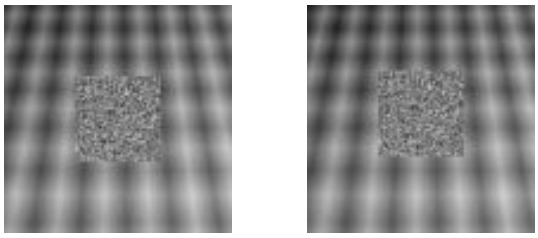

Figure 1. Two frames of a synthetic scene with 2 independently moving objects.

increasing depth from bottom to top in Fig. 1. Please note that in Fig. 1 object 1 is the background since it covers the entire field of view. The plane moves to the right (x direction). In front of object 1 a second object with different texture is moving independently in y direction. The scene is recorded before and after the motion and the two resulting images are used for motion estimation. The algorithm presented in this paper accurately estimates the dominant motion to be translational only in x direction (in this case the motion of object 1). The displacement vector for each pixel is obtained searching for the best match along the epipolar line for a rectangular measurement window around the pixel. Fig. 2 shows the recovered depth map applying the estimated dominant motion for all image pixels. It can be seen that the structure of object 1 is accurately recovered, but the image area covered by object 2

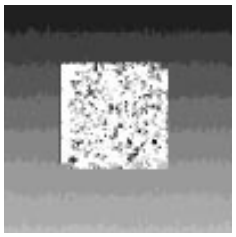

Figure 2. Recovered depth map using the dominant motion parameters for all image pixels.

leads to many outliers. Please note that dark values represent large depth values. We then sort the matches along the epipolar line found for each pixel by their magnitude as shown in Fig. 3. The separation of the two objects is achie-

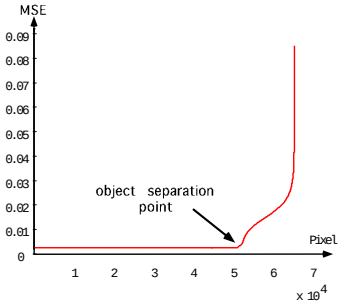

Figure 3. Sorted residual error (MSE) along the epipolar line for all image pixels using the dominant motion parameter set recovered with the LMedS estimator.

ved by searching for the point where we observe a sudden increase in matching error. Fig. 4 shows the depth map with separation of the objects. All pixels not belonging to object 1 are white. It can be seen that automatic segmentation is achieved. Please note that pixels that belong to uncovered background (object 1) are not classified as part of object 1. Pixels have to be visible in both frames in order to achieve correct segmentation. The algorithm now removes all pixels belonging to object 1 from the data set and the algorithm restarts from the beginning for the remaining pixels. In a second experiment we used the images reprodu-

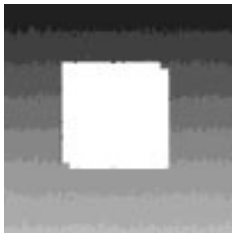

Figure 4. Recovered depth map after object separation with the LMedS estimator.

ced in Fig. 5 as input to our algorithm. In comparison to Fig. 1, object 2 is replaced by a textured 3D head model. The plane moves translationally to the right (x-direction), whereas the motion parameters of the head are rotation of 1^o around the y-axis ($R_x = 0^o$, $R_y = 1^o$, $R_z = 0^o$) and

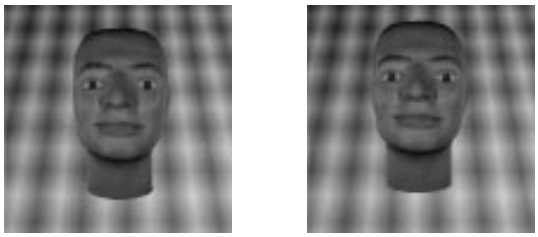

Figure 5. Second example of a Synthetic scene with two independently moving objects.

translation in y-direction. Fig. 6 shows the recovered depth map after object separation with the algorithm proposed in this paper. As in the previous experiment, the plane (object 1) represents the dominant motion. The motion parameters recovered for object 1 show translation in x-direction. After compensation of the dominant motion, the estimated rotation for object 2 is $R_x = -0.02^o$, $R_y = 0.993^o$, $R_z = 0.01^o$. The translation is correctly estimated to be in y-direction only.

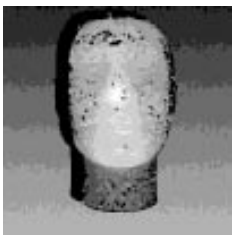

Figure 6. Recovered depth map after object separation with the LMedS estimator.

4.2. Animation of an artificial 3D object in a real image sequence

In the following we describe how the algorithm presented in this paper has been applied to the automatic insertion of an artificial 3D model into a real image sequence. The virtual camera recording the artificial object is animated using the motion parameters estimated from the real sequence. Fig. 7 shows the first frame of the *Flowergarden* sequence and the rendered artificial 3D object, a windmill, to be inserted. The artificial object is superimposed ma-

Figure 7. Left: frame 1 of *Flowergarden* sequence. Right: an artificial 3D object rendered at initial position.

nually at the position in 3D space such that the projection

in the image plane corresponds to the desired starting position. Fig. 4.2. shows the estimated motion parameters (translation) for the first 100 frames of the real image sequence. From monocular image sequences the translation

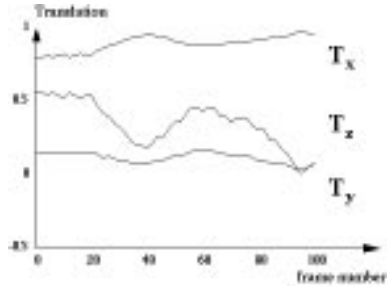

Figure 8. Estimated translation parameters for 100 frames of the *Flowergarden* sequence.

vector can be reconstructed only up to a common scale factor. We therefore have to appropriately scale the recovered translation vector from frame to frame. In order to initialize the length of the translation vector we average the recovered translation parameters for the first 20 frames of the sequence and manually determine the required length of the translation vector for correct position of the rendered object in the image plane. We then compute the scaling factor between successive depth maps comparing the average inverse depth value of the scene. At frame 50 we reinitialize the length of the translation vector in order to avoid drift of the artificial object. In Fig. 4.2. we show frames 1, 20, 40, 60, 80, and 100 of the artificially created image sequence. When viewed as a motion sequence, the synthetic object appears to be rigidly connected to the (natural) ground.

5. CONCLUSIONS

We have presented a robust method to recover multi-component motion in a scene using the epipolar constraint in conjunction with an LMedS estimator. The tasks of motion segmentation and depth recovery are performed simultaneously in a computationally efficient way. Experimental results show that the approach presented in this paper is capable of estimating the motion of multiple objects. We applied the algorithm to the problem of automatic insertion of artificial 3D models into a real image sequence.

Figure 9. Frames 1, 20, 40, 60, 80, 100 of the artificially created image sequence.

REFERENCES

[1] A. Rognone, M. Campani and A. Verri, "Identifying Multiple Motions from Optical Flow", *Proc. 2nd ECCV*, pp 258-266, Santa Margherita Ligure, Italy, May 1992.

[2] S. Ayer, "Sequential and Competitive Methods for Estimation of Multiple Motions", *Doctoral Dissertation*, École Polytechnique Fédérale de Lausanne, 1995.

[3] M.J. Black and P. Anandan, "The Robust Estimation of Multiple Motions: Parametric and Piecewise-Smooth Flow Fields", *Computer Vision and Image Understanding*, vol. 63, no. 1, January, pp. 75-104, 1996.

[4] W. Wang and J.H. Duncan, "Recovering the Three-Dimensional Motion and Structure of Multiple Moving Objects from Binocular Image Flows", *Computer Vision and Image Understanding*, vol. 63, no. 3, May, pp. 430-446, 1996.

[5] E. Steinbach and B. Girod, "Estimation of Rigid Body Motion and Scene Structure from Image Sequences Using a Novel Epipolar Transform",*Proc. ICASSP '96*, pp. 1911-1914, Atlanta, 1996.

[6] P.J. Rousseeuw and A.M. Leroy, "Robust Regression and Outlier Detection", John Wiley, New York, 1987.

[7] R.Y. Tsai and T.S. Huang, "Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces, "*IEEE Trans. Pattern Anal. Machine Intell.*, vol. 6, no. 1,pp. 13-27, 1984.

[8] J.K. Aggarwal and N. Nandhakumar, "On the Computation of Motion from Sequences of Images - A Review," *Proc. IEEE*, vol. 7b,no. 8, pp. 917-935, August 1988.

[9] J. Weng, N. Ahuja, T.S. Huang, "Optimal Motion and Structure Estimation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 15, no. 9, pp. 864-884, September 1993.

[10] B. Girod and P. Wagner, "Displacement Estimation with a Rigid Body Motion Constraint," *Proc. International Picture Coding Symposium*, Cambridge, Mass., USA, March 1990.

[11] S. Chaudhuri, S. Sharma, and S. Chatterjee, "Recursive Estimation of Motion Parameters", *Computer Vision and Image Understanding*, November 1996.