

ROBUST ESTIMATION OF RIGID BODY 3-D MOTION PARAMETERS FROM POINT CORRESPONDENCES

Th. Papadimitriou¹, K. I. Diamantaras², M. G. Strintzis³, M. Roumeliotis⁴

¹Educational Technology Lab

⁴Dept. Applied Informatics

University of Macedonia

156 Egnatia St.

GR-54006, Thessaloniki, Greece

²Dept of Informatics

Technological Education Institute

Sindos

GR-54101, Thessaloniki, Greece

³Information Processing Lab

Dept. Electrical and Computer Eng.

Aristotle Univ. Thessaloniki

GR-54006, Thessaloniki, Greece

ABSTRACT

The estimation of a rigid body 3-D motion parameters from perspective views is typically very sensitive to noise and also to the presence of outliers in the measurements. In this paper we present a robust 3-D motion estimation approach based on a previously proposed method using SVD analysis of the measurements matrix. On the introduction of noise and outliers the performance of the old method was seen to deteriorate rapidly. Here the problem is attacked by splitting the measurement set in smaller subsets and combining the properties of the resulting submatrices with the properties of the desired solution vector in order to obtain our estimate. The method is very robust and it has been successfully tested in both artificial datasets and real images with up to 50 % presence of outliers. In addition, the method is fast and more importantly, the estimate quality is independent of the percentage of outliers.

1. INTRODUCTION

The goal of rigid-body motion estimation is to determine the 3-D rotation and translation parameters of a rigid object moving relative to a fixed coordinate system. The estimates are based on 2-D displacement information as it appears in an image sequence. Alternatively, the motion parameters can be used in order to estimate the camera motion relative to a fixed, stationary scene. Generally a 3-D motion estimation algorithm involves two steps: (a) 2-D motion estimation represented by a displacement vector field and (b) derivation of 3-D movement from the displacement vec-

tors. Displacement can be represented by a dense vector field [6] or by a sparse vector field generated by point correspondences [7] or line correspondences [9].

There is a wide variety of techniques for estimating 3-D motion parameters from 2-D displacement fields. Some of the proposed methods are: to use more than two frames [2], to use quaternion representation [8], to assume orthographic projection [1], to use region alignment [13], to use neural networks [12], to use polar decomposition [8]. However, the accuracy of the 2-D motion field is of paramount importance for the accurate estimation of the 3-D motion parameters, as the latter process is very sensitive to noise and to the outliers present in the 2-D measurements.

In our previous work [4] we proposed a 3-D motion estimation method using point correspondences between 2 frames. The method is based on the SVD analysis of a system matrix which yields perfect results in the ideal case. However, as with all 3-D motion estimation methods this approach is very sensitive to measurement errors and to the presence of outliers. In this paper we present a novel algorithm that makes this method robust to outliers up to 50%, while the performance of the motion estimation is stable.

1.1. SVD-based motion estimation

Let us define a 3-D coordinate system so that the Z-axis is parallel to the camera optical axis and the X-, Y-axes coincide with the x -, y -axes of the image plane. The motion of a point $\mathbf{P} = [X, Y, Z]^T$ in 3-D space is described by the following equation

$$\mathbf{P}' = [X', Y', Z']^T = \mathbf{R}\mathbf{P} - \mathbf{t} \quad (1)$$

where the orthogonal matrix \mathbf{R} describes the rotation, and the vector \mathbf{t} describes the translation of \mathbf{P} .

We assume that the camera geometry is described by perspective projection with focal length f . We also

This work was supported in part by the NATO, Scientific and Environmental Affairs Division, Collaborative Research Grant SA.5-2-05(CRG.960201)424/96/JARC-501. and by the Greek Secretariat of Research and Technology.

assume small enough rotation and translation parameters so that the 2-D displacements $[u_i, v_i] = [x'_i - x_i, y'_i - y_i]$ for N points $i = 1, \dots, N$, can be approximated as

$$u_i = \frac{-ft_x + x_it_z}{Z_i} + \frac{1}{f}w_x x_i y_i - w_y \left(f + \frac{x_i^2}{f}\right) + w_z y_i \quad (2)$$

$$v_i = \frac{-ft_y + y_it_z}{Z_i} - \frac{1}{f}w_y x_i y_i + w_x \left(f + \frac{y_i^2}{f}\right) - w_z x_i \quad (3)$$

Eliminating Z_i from (2) and (3) leads to the following linear equation [3]

$$\mathbf{A}\theta = 0 \quad (4)$$

where the 9-dimensional vector

$$\theta^T = [t_x, t_y, t_z, t_x w_x, t_y w_y, t_z w_z, t_x w_y + t_y w_x, t_x w_z + t_z w_x, t_y w_z + t_z w_y] \quad (5)$$

involves all six free parameters of the problem, i.e. $t_x, t_y, t_z, w_x, w_y, w_z$. In the noise-free case the solution hinges on the SVD analysis of the $N \times 9$ system matrix

$$\mathbf{A} = \begin{pmatrix} f v_1 & \dots & f v_N \\ -f u_1 & \dots & -f u_N \\ u_1 y_1 - v_1 x_1 & \dots & u_N y_N - v_N x_N \\ -(f^2 + y_1^2) & \dots & -(f^2 + y_N^2) \\ -(f^2 + x_1^2) & \dots & -(f^2 + x_N^2) \\ -(x_1^2 + y_1^2) & \dots & -(x_N^2 + y_N^2) \\ x_1 y_1 & \dots & x_N y_N \\ f x_1 & \dots & f x_N \\ f y_1 & \dots & f y_N \end{pmatrix}^T \quad (6)$$

since θ is associated with the null space of \mathbf{A} [3].

In the noisy case the problem can be formulated as a Total Least Squares problem [4]. Still the SVD analysis employed for TLS is quite sensitive to noise and the results obtained are good for noise levels less than 10%.

2. ROBUST TLS MOTION ESTIMATION

As with most 3-D motion estimation methods the solution depends heavily on quality of the measurements. When the output of the feature matching process contains noise and outliers, the TLS method will give very poor results. In this section, a recursive robust scheme is introduced to reject outliers from the dataset based on the residuals information.

First we introduce different weighting on the rows of \mathbf{A} : we call $\mathbf{W} = \text{diag}[w_1, w_2, \dots, w_n]$ a diagonal weight matrix where $w_i \in [0, 1]$ reflects our confidence on the i -th measurement -0 means no confidence, 1 means perfect confidence. Then the linear system to solve is,

$$\mathbf{W}\mathbf{A}\theta = 0 \quad (7)$$

Our goal is to determine \mathbf{W} and θ . This is done recursively. First, we start with an estimate of \mathbf{W} which leads to an estimate of θ from Eq. (7). Then the residual error of our estimate leads to a new weight matrix \mathbf{W} and a new estimate of θ , and so on. Since we have no prior knowledge on the quality of our measurements we set initially all the weights equal to 1, so $\mathbf{W}^0 = \mathbf{I}$. An estimation of θ_i is made in every iteration using the TLS algorithm proposed in [4]. The residuals e_i of the estimation are used to calculate the weights for the next iteration using the equation:

$$w_i = \begin{cases} (1 - e_i)^2 & \text{if } |e_i| < 1 \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

In order to achieve better results, studentized residuals are used instead of simple residuals [16]. The studentized residuals are defined as:

$$e'_i = \frac{e_i}{\sqrt{1 - h_{ii}}} \quad (9)$$

where h_{ii} is the diagonal element of matrix $\mathbf{H} = \mathbf{W}\mathbf{A}(\mathbf{W}\mathbf{A})^+$ where the superposed + denotes pseudoinverse.

The SVD algorithm needs at least a dataset of nine points to estimate a solution. The robust algorithm ends when a number (larger than nine) of weights w_k remain larger than a threshold L . The corresponding points are used to the TLS estimation. In our experiments, we were searching for 12 - 15 points, while the threshold L was 0.8 - 0.95.

The above improvement of the original algorithm is significantly more robust but it can be further improved. The idea is to generate a number of smaller submatrices \mathbf{A}_i by randomly selecting rows from \mathbf{A} . For each submatrix we use the same approach outlined above. Using more submatrices we increase the probability that two of them contain sufficiently low percentage of outliers so that the algorithm will yield estimates $\hat{\theta}$ close to the true θ .

In order to take advantage of this idea the algorithm must be able to judge the quality of the estimate $\hat{\theta}_i$ for each submatrix \mathbf{A}_i . We achieve this by making use of two properties inherent to the data model. First, let $\sigma_1, \sigma_2, \dots, \sigma_9$, be the singular values of \mathbf{A} sorted in decreasing order. In the noiseless case it can be shown [3] that $\sigma_9 = 0$. In the noisy case this is not true but we may assume that the ratio between the smallest and the largest singular value should remain very small: $\sigma_9/\sigma_1 \approx 0$. Second, note that good estimates are those that cluster in a small region around the true θ , whereas bad estimates tend to disperse in \mathbb{R}^6 . Therefore, if two estimates from two different submatrices lie close to each other it is more likely that they lie close to

the true value θ , than that they lie in some random neighborhood of \mathfrak{R}^6 .

Based on the above discussion our proposed algorithm aims at constructing two submatrices \mathbf{A}_i and \mathbf{A}_j such that (a) $\frac{\sigma_a}{\sigma_1} \approx 0$ for both submatrices, and (b) the estimates $\hat{\theta}_i, \hat{\theta}_j$ are close to each other. Then our estimate $\hat{\theta}$ is the mean of $\hat{\theta}_i$ and $\hat{\theta}_j$. The complete algorithm is presented below.

1. Create A_i .
2. Robust Estimation.
 - (a) Initialisation $W^0 = I$.
 - (b) Estimate $\hat{\theta}_i^t$ of $W^t A_i \theta = 0$ using TLS.
 - (c) Calculate studentized residuals e'_k from Equation (9)
 - (d) Calculate the new weights w_k^{t+1} using e'_k
 - (e) If twelve different w_k are larger than L goto step 3. L is a predefined threshold. Else goto step (b).
3. If $\frac{\sigma_a}{\sigma_1} < 10^{-4}$ store θ_i .
4. Goto Step 1 unless there are two estimates θ_i and θ_j such as $\hat{\theta}_i^T, \hat{\theta}_j \approx 1$.
5. $\hat{\theta} = (\hat{\theta}_i + \hat{\theta}_j)/2$

In the next section, the performance of the algorithm is presented for artificial data and real images.

3. RESULTS

3.1. Artificial Data

For the artificial tests, 100 3D points were generated. Correspondences of these points were generated using different motion models. Then a percentage of outliers was injected in the dataset. The performance of the TLS method was tested by introducing quantization error p to the data. The quantization is measured using the factor $s = \frac{f}{q}$, where f is the focal length of the camera, and q is the size of the square pixel. Typical values are $f = 40mm$ and $q = 0.01mm$, so $s \simeq 4000$. Quality of the dataset is calculated using SNR, defined as:

$$SNR = 10 \log \left(\frac{\sum_i u_i^2 + v_i^2}{\sum_i (\bar{u}_i - u_i)^2 + (\bar{v}_i - v_i)^2} \right) \quad (10)$$

Our choice of parameters yields realistical, average quality data (measurement SNR between 19 and 20db).

In Fig. 1 the tested motion model was for $\mathbf{t} = [2.9, 1.8, 3.3]$ and $\mathbf{w} = [3, -2, 0.5]$ (the angles are measured in degrees). The quality of the estimation is independant of the outliers percentage, while computational cost is propotional to the outliers level. An iteration of the algorithm in a Pentium II 266MHz processor running Matlab-5 needs between 0.9 and 1.6 seconds.

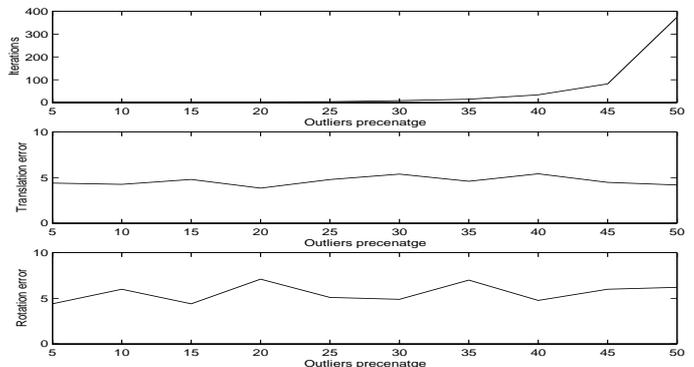


Figure 1: The top figure shows the necessary iterations until the end of the algorithm. The two bottom figures shows the average error percentage in the translation and the rotation parameters for different outliers levels

3.2. Real Data

The real experiments were made using the train calendar sequence. Feature points of the ball and the train were found. Motion of these points was estimated using TLS. Different levels of outliers were injected in the dataset to test the robust scheme. In Table 1 we show the mean and the standard deviation of the results for 100 noisy datasets. The rotation parameters w are in degrees. Robust TLS estimation is very close to the one of the TLS as shown by the means. The region of the estimated parameters as seen from the standard deviation is bigger in the robust estimation, due to the injected outliers, than in the simple TLS.

3.3. Discussion

The estimation of the proposed method is independent of the outliers percentage in the input dataset. The motion estimation quality is stable for data containing up to 50% outliers. Moreover, the computational cost is low and proportional to the outliers percentage. In the case of a dataset infected with more than 50% outliers, the robust TLS motion estimation will end without solution. Bad estimates are unable to pass the security thresholds described in section 2.

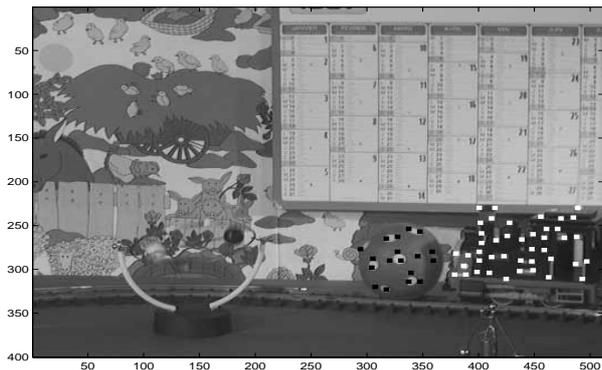


Figure 2: The first image of the calendar train sequence. Tracked features points of the train are in white and the tracked features points of the ball are in black

4. CONCLUSION

In this article a novel robust scheme for 3-D motion estimation is introduced. The method is based on the SVD analysis of a linear system. The method is tested in both artificial data and real video sequences. The results show that our method is robust in datasets with up to 50% of outliers. Our robust scheme is attractive since the computational cost is proportional to the level of outliers injected in the datasets, while the estimation quality is independent of the outliers percentage.

5. REFERENCES

- [1] K. S. Arun, T. S. Huang and S. D. Blostein, "Least-squares fitting of two 3-d point sets", IEEE Trans. Pattern Anal. and Machine Intell., Vol. PAMI-9(5), 1987.
- [2] T. Broida and R. Chellappa, "Estimation of object motion parameters from noisy images", IEEE Trans. Pattern Anal. and Machine Intell., Vol. PAMI-8(1), pp. 90-99, 1987.
- [3] K. I. Diamantaras and M. G. Strintzis, "Camera motion parameter recovery under perspective projection", In Proc. ICIP-96, Vol 3, p 783-786, Lausanne, Switzerland, 16-19 September, 1996.
- [4] K. I. Diamantaras, Th. Papadimitriou, M. G. Strintzis and M. Roumeliotis, "Total Least Squares 3-D Motion Estimation", In Proc. ICIP-98, Chicago(IL), U.S.A., 4-7 October 1998.
- [5] G. H. Golub and C. F. Van Loab, Matrix Computation, Johns Hopkins University Press, 1989.
- [6] D. Heeger and A. Jepson, "Subspaces methods for recovering rigid motion : algorithm and implementation" International Journal of Computer Vision, Vol 7, p. 95-117, 1992.
- [7] R. Y. Tsai and T. S. Huang, "Uniqueness and estimation of 3-d motion parameters of rigid bodies with

	Translation			Rotation		
train	t_x	t_y	t_z	w_x	w_y	w_z
Mean	0.899	0.394	0.184	0.046	0.120	-0.160
STD	0.020	0.049	0.008	0.034	0.017	0.074
Mean	0.903	0.373	0.190	0.029	0.120	-0.103
STD	0.040	0.088	0.014	0.252	0.069	0.212
ball	t_x	t_y	t_z	w_x	w_y	w_z
Mean	0.529	0.775	0.357	0.836	1.845	0.132
STD	0.042	0.030	0.011	0.080	0.074	0.132
Mean	0.511	0.792	0.320	1.002	1.931	0.114
STD	0.076	0.053	0.032	0.610	0.521	0.681

Table 1: Estimation of the train and ball motion. First, the motion of the train is estimated from a clear dataset and a dataset injected with 12% of outliers, using TLS and the robust scheme respectively. Next the motion of the ball is estimated using the same procedure. The ball dataset is injected with 20% of outliers to test the robust TLS.

curved surfaces", IEEE Trans. Pattern Anal., and Machine Intel., Vol. PAMI-6(1), p. 13-27, 1984.

- [8] Kenishi Kanatani, "Analysis of 3-D Rotation Fitting", IEEE Trans. Pattern Anal., and Machine Intel., Vol. PAMI-16(5), p. 543-549, 1994.
- [9] Zhengyou Zhang, "Estimating Motion and Structure from Corresponding of Line Segments between Two Perspective Images", IEEE Trans. Pattern Anal., and Machine Intel., Vol. PAMI-17(12), p. 1129-1139, 1995.
- [10] Y. Huang, K. Palapaniappan, X. Zhuang and J. E. Cavanaugh, "Optic Flow Field Segmentation and Motion Estimation Using a Robust Genetic Partitioning Algorithm", IEEE Trans. Pattern Anal., and Machine Intel., Vol. PAMI-17(12), p. 1177-1190, 1995.
- [11] Y. Liu and T.S. Huang, "A sequence of stereo image data of a moving vehicle in an outdoor scene", Tech. report, Bechman Insitute, Univ of Illinois at Urbana-Champaign, 1990.
- [12] Y. Sun and M. M. Bayouni, "A Simple Feedforward Neural Network Architecture for 3-D Motion and Structure Estimation", In Proc. ICIP-96, Vol 3, p 783-786, 1996
- [13] M. Irani, B. Rousso and S. Peleg, "Recovery of Ego-Motion Using Region Alignment", IEEE Trans. Pattern Anal., and Machine Intel., Vol. PAMI-19(3), p. 268-272, 1997.
- [14] O. Faugeras, Three-dimensional computer vision: a geometric viewpoint MIT Press, 1993.
- [15] P.J. Huber, "Robust Statistics", John Wiley & Sons, New York, 1981.
- [16] Robert G. Staudte and Simon J. Sheather, "Robust Estimation and Testing", John Wiley & Sons, 1990.