Recovering 3-D Motion Parameters from Image Sequences with Gross Errors

Chung-nan Lee, Robert M. Haralick, and Xinhua Zhuang
Department of Electrical Engineering
University of Washington
Seattle, WA 98195

ABSTRACT

A robust algorithm to estimate 3-D motion parameters from a sequence of extremely noisy images is developed. The noise model includes correspondence mismatch errors, outliers, uniform, and Gaussian noise. More than one hundred thousand controlled experiments were performed. The experimental results show that the error in the estimated 3–D parameters of the linear algorithm almost increases linearly with fraction of outliers. However, the increase for the robust algorithm is much slower indicating its better performance and stability with data having blunders.

1. Introduction

The estimation of three-dimensional motion parameters of a rigid body is an important problem in motion analysis . Its applications include scene analysis, motion prediction, robotics vision, and on line dynamic industrial processing. There has been much literature contributed to 3D parameter estimation, but few of these contributions systematically discuss the effect of noise. Thompson (1959) developed the nonlinear equations using the form resulting from the correspondence of 2D perspective projection points on one image with 2D perspective projection points on another image. He gave a solution which determines a rotation matrix guaranteed to orthonormal. His method was to linearize the non–linear equations and iterate. Roach and Aggarwal (1980) developed a nonlinear algorithm and dealt with noisy data. Their results show that accuracy can be improved by increasing the number of corresponding point pairs; but the number of corresponding point pairs in their experiments is too few (15 corresponding point pairs). The linear motion parameters estimation algorithm was developed by Longuet -Higgins(1981) , extended by Tsai and Huang(1984), unified by Xinhua Zhuang, T.S. Huang, and R. M. Haralick(1986), and simplified by Xinhua Zhuang and R. M. Haralick. The linear algorithm has an advantage of being simple and fast over the nonlinear algorithm. Furthermore, it can always find a unique solution except in degenerate cases. The linear algorithm works very well when there is limited noise and no corresponding point matching errors. However, the algorithm is highly sensitive to noise and matching errors. Experiments show that when combined with real world image corresponding point data produced by a vision systems, a disaster occurs. Because the nature of linear algorithm, increasing the number of corresponding point pairs can only to some extent suppress the noise effect. The main problem in linear algorithm is the least squares estimation.

The method of least squares is based on evaluation of the magnitude of residuals and is sensitive to matching errors and outliers. Unlike the least squares estimator the robust estimator has good resistance and robustness to gross matching error and outliers. In section 2 a simplified linear algorithm presented by Zhuang and Haralick (1986) is used to get the baseline noise behavior of the linear algorithm. The principle of robust computation is presented in section 3. In the last section the experimental design is discussed and the results shows that robust algorithm has better performance and stability.

2. Simplified Linear Algorithm

As shown in Fig. 1 we assume that the coordinate system is the camera reference frame, the origin being the center of the lens. A rigid body is in

CH2716-9/89/0000/0046$01.00 © 1989 IEEE

motion in the half-space z $\leq$ 0. Let $P = (x, y, z)^t$ represent a set of object points coordinate before motion and $P' = (x', y', z')^t$ represent the same set of object points coordinate after motion. The superscript t means the transpose of matrix will be used in the paper. The point coordinate $[x_i, y_i, z_i] \in p$ is corresponding to $[x_i', y_i', z_i'] \in P'$. Let (X,Y), (X', Y') represent the perspective coordinate of P and P' onto the image plane z=1. These give

$$X = x/z$$
$$Y = y/z$$
$$X' = x'/z'$$
$$Y' = y'/z'. \quad (1)$$

The rigid body motion equation is given as follows:

$$P' = R_o P + T_o. \quad (2)$$

where R_o is an 3 x 3 rotation matrix (orthnormal) ; T_o is 3 x 1 translation vector. In terms of Euler angles ψ, θ, and ϕ the rotation matrix can be represented as follows:

$$R_o = \begin{pmatrix} \cos\psi\cos\theta & \sin\psi\cos\theta & -\sin\theta \\ -\sin\psi\cos\phi + \cos\psi\sin\phi\sin\theta & \cos\psi\cos\phi + \sin\psi\sin\theta\sin\phi & \cos\theta\sin\phi \\ \sin\psi\sin\phi + \cos\psi\sin\theta\cos\phi & -\cos\psi\sin\phi + \sin\psi\sin\theta\cos\phi & \cos\theta\cos\phi \end{pmatrix}$$

The problem is to estimate rotation matrix R_o and translation matrix T_o.

2.1 The Two View Motion Equation

Choosing any nonzero vector T which is collinear with T_o and taking its cross-product with both sides of Eq.(2), we obtain

$$\frac{z'}{z} T \times (X', Y', 1)^t = T \times [R_o (X, Y, 1)^t]. \quad (3)$$

Taking inner product of both sides of Eq.(3) with $(X', Y', 1)$ yields

$$(X', Y', 1)(T \times R_o)(X, Y, 1)^t = 0. \quad (4)$$

where T x $R_o = [T \times r_1, T \times r_2, T \times r_3]$, and r_1, r_2, r_3 are the columns of R_o. Define the motion parameter matrix E by

$$E = T \times R_o. \quad (5)$$

For any image corresponding pair [(X,Y),(X', Y')] the matrix E satisfies the following linear homogeneous equations w. r. t. nine elements of E:

$$(X', Y', 1) E (X, Y, 1)^t = 0. \quad (6)$$

Relation of Eq.(6) was originally shown by Thompson(1959). Suppose that we have N correspondences. Let

$$A = \begin{pmatrix} X_1'X_1 & X_1'Y_1 & X_1' & Y_1'X_1 & Y_1'Y_1 & Y_1' & X_1 & Y_1 & 1 \\ X_2'X_2 & X_2'Y_2 & X_2' & Y_2'X_2 & Y_2'Y_2 & Y_2' & X_2 & Y_2 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_n'X_n & X_n'Y_n & X_n' & Y_n'X_n & Y_n'Y_n & Y_n' & X_n & Y_n & 1 \end{pmatrix}$$

$$E = \begin{pmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & h_9 \end{pmatrix}$$

$$h = (h_1, h_2, h_3, h_4, h_5, h_6, h_7, h_8, h_9)^t$$

Then the Eq.(6) can be transformed into the overconstraint linear equation for h

$$Ah = 0. \quad (7)$$

Solving Eq.(7) in the least squares sense we seek an estimator h which minimizes $\|Ah\|^2$. The 9 component vector h is found to be the right eigenvector of A having smallest singular value. Any $T \times R_0$ with $T \times T_0 = 0$ satisfies Eq.(6). Moreover, such a colinear vector T has one degree of freedom when $T_0 \neq 0$ or three degrees of freedom when $T_0 = 0$. Thus the general solution of the Two-View Motion Eq.(6) has at least one degree of freedom when $T_0 \neq 0$ or three degrees of freedom when $T_0 = 0$. When $T_0 \neq 0$, the nine elements of E must have a rank 8, and $T_0 = 0$ the nine elements of E must have rank 6. Under the surface assumption (Zhuang, Haralick, and Huang, 1986) the number of image corresponding point pairs must be at least 8 when $T_0 \neq 0$, or greater than or equal to 6 when $T_0 = 0$. The geometry interpretation we use assumes that the object is stationary and the camera is moving. Let the origin of the camera system be O and O' respectively before and after motion. Then the surface assumption holds if and only if the 3D points corresponding to the observed image points do not lie on a quadratic surface passing through O and O' when $T_0 \neq 0$ or a cone with its apex at O when $T_0 = 0$.

2.2 Decomposing E

E has two decompositions; $T \times R_0$ and $(-T) \times R_0$ with R_0 being an orthonormal matrix of the first kind. In order to determine the correct decomposition we note that $E = [T \times r_1, T \times r_2, T \times r_3]$. Hence, its three columns span a 2D space and also $\|E\| = \sqrt{2}\|T\|$. Therefore we can get three constraints as follows:

$$\begin{aligned} \text{Rank } (E) &= 2 \\ \|E\| &= 2\|T\|. \\ E^t T &= 0 \end{aligned} \tag{8}$$

We can use the least square method to solve Eq.(8) for T and obtain the value of the T vector from the other two constraints. Since T is colinear with T_0, T_0 should have the same orientation as T or $-T$. Taking a cross-product with both sides of Eq.(2) by $(X', Y', 1)^t$ we obtain

$$z_1(X', Y', 1)^t \times [R_0(X, Y, 1)^t] + (X', Y', 1)^t \times T_0 = 0 \tag{9}.$$

Since $z < 0$, it implies that T_0 has the same orientation as T or $(-T)$ if and only if $(X', Y', 1)^t \times [R_0(X, Y, 1)^t]$ has the same orientation as $(X', Y', 1)^t \times T$ or $[-(X', Y', 1)^t \times T]$. This implies that it has the same orientation if and only if

$$\sum_{i=1}^{n} (X_i', Y_i', 1)^t \times [R_0(X_i, Y_i, 1)^t](X_i', Y_i', 1)^t \times T \geq 0 or \leq 0. \tag{10}$$

Once the correct T is determined, the true R_0 could be uniquely determined through $E = T \times R_0$ as follows:

$$R_0 = [E_2 \times E_3, E_3 \times E_1, E_1 \times E_2] - T \times E \tag{11}$$

where $E = [E_1, E_2, E_3]$

3. The Robust Algorithm

As mentioned in the previous section the Eq.(7) can be solved by least-squares estimator. However, it is sensitive to gross errors. In this section the robust algorithm is presented. The robust algorithm is an iterative reweighted least squares estimation procedure where the weights are recomputed each iteration and are computed as a biweight. The difference between the biweight estimator and the least-squares estimator is briefly discussed.

3.1 Biweight Estimator

Let x_i be the i^{th} observation and $\hat{x}$ be estimated mean value of the observations. The least squares method minimizes the residual error

$$\epsilon^2 = \sum_{i=1}^{n} (x_i - \hat{x})^2. \tag{12}$$

and the object function, ρ, is expressed as follows

$$\rho(x_i; \hat{x}) = (x_i - \hat{x})^2. \tag{13}$$

To find the solution of the problem we differentiate ρ w.r.t. $\hat{x}$, the derivative ψ satisfies

$$\sum_{i=1}^{n} \psi(x_i; \hat{x}) = \sum_{i=1}^{n} (x_i - \hat{x}) = 0. \tag{14}$$

As discussed in Hoaglin the least-squares estimator is linear and unbounded with an increase in any of the observed values.

The ψ function of the biweight estimator can be represented as follows

$$\psi(u) = \begin{cases} u(1-u^2)^2 & |u| \leq 1 \\ 0 & \text{otherwise} \end{cases}. \tag{15}$$

where

$u_i = \frac{f_i(\epsilon)}{c s_n}$

$f_i(\epsilon)$: residual error function

s_n : median value of $f_i(\epsilon)$

c : tuning constant

Unlike the least-squares estimator, the ψ- function of the biweight estimator is bounded. When the value of tuning constant is small it will delete a lot of useful data. On the other hand, when the value is large the outliers can not be removed from the images. Hence, the tuning constant depends on the value of gross errors. A reasonable value range for tuning constant is from 4 to 12. In here we let c=4. Let $\psi(u)$ = w(u)u. Thus, the weight function $w(u)$ can be represented

$$w(u) = \begin{cases} [1-u^2]^2, & \text{if } |u| \leq 1; \\ 0, & \text{otherwise.} \end{cases} \tag{16}$$

3.2 Robust Estimation of E

From Eq. (16)we can see that the biweight estimator is a weighted least square estimator. With the weight matrix we rewrite the Eq.(7)

$$WAh = 0. \tag{17}$$

To find the value of h which minimizes $\|WAh\|^2$ the singular value decomposition can be used

$$WA = U \sum V^t. \tag{18}$$

where

$$\sum_{m \times n} = \begin{pmatrix} s_1 & 0 & \dots & 0 \\ 0 & s_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \ddots & 0 & s_9 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 \end{pmatrix}$$

$$V^t_{n \times n} = [v_1, v_2 ..., v_n]$$
$$U_{m \times m} = [u_1, u_2 ..., u_m]$$

The index n is 9 and m is the number of corresponding point pairs. The right eigenvector of V which corresponds to the smallest nonzero eigenvalue in $\sum$ is the solution of weighted least squares. Here it will be denoted by v_9. Multiplying the current solution for h by A to get the new residual. Gross errors are not necessarily accompanied large residuals as explained in Huber(1981). Hence, the residual errors need to be adjusted according to the following

$$f_i(\epsilon) = \frac{\epsilon_i}{1 - h_{ii}}. \tag{19}$$

where h_{ii} is the diagonal element of the projection matrix H

$$H = (WA)((WA)^t(WA))^{-1}(WA)^t. \tag{20}$$

We can simplify the above equation by substituting $U \sum V^t$ for WA. After some linear algebra manipulation Eq.(20) becomes

$$H = U_a U_a^t. \tag{21}$$

where

$$U_{am \times 9} = [u_1, u_2 ...., u_9]$$

It is trivial to obtain $h_{ii} = \sum_{k=1}^{9} u_{ik}^2$. Once h_{ii} are obtained, then they can be substituted into Eq.(19) to get the new residual error function and to update the weight matrix. The initial weight matrix is identity matrix. The iterations continue until some criteria are satisfied. In our experiments when the error ϵ^2 is less than $0.001\epsilon^2$ of first iteration or the iteration number is larger than 25, then the iteration process stops. Usually it will converge after a few iterations. The value of v_9 at the last iteration is the robust solution.

4. Simulation Result and Discussion

In this section we discuss the experimental results of a large number of controlled experiments using the linear algorithm and the robust algorithm under a varying amount of noise, gross errors and corresponding point pairs. As shown in Fig. 1, the image frame is located at $z = 1$. By mapping 3D spatial coordinates into image frame, and then adding noise to the points before and after motion, we obtain

$$\begin{pmatrix} X(t) \\ Y(t) \end{pmatrix} = \begin{pmatrix} 1/z(t) & 0 & 0 \\ 0 & 1/z(t) & 0 \end{pmatrix} \begin{pmatrix} x(t) \\ y(t) \\ z(t) \end{pmatrix} + \begin{pmatrix} n_x(t) \\ n_y(t) \end{pmatrix} \tag{22}$$

Signal is related to object image size, and noise may come from camera error, digitization, or corresponding point extraction error. Define SNR $= 20 \log \frac{signal}{\sigma}$db, where σ is the standard deviation. In the simulation experiments , the 3-D spatial coordinates before motion (x,y,z), true matrix R_o, and true translation vector T_o are generated by a random number generator. The 3-D data are generated within the (-2,-2,-2) to (2,2,2) cube. The rotation angles ϕ,θ,ψ are generated within the the range of [-15, 15] degree and translation vectors are chosen within the range (-0.5,-0.5,-0.5) to (0.5,0.5,0.5) cube. Then the 3-D spatial coordinates after motion (x', y', z') can be calculated in the natural way. Projecting the 3-D spatial coordinates into the image frame we get perspective coordinates. Noisy image data is obtained by adding Gaussian or Uniform noise with zero mean to the image coordinates. Outliers are generated by randomly moving some corresponding points position in image frame after motion. The number of outliers are chosen as a percent of corresponding point pairs. Following the linear algorithm or the robust algorithm as described above we can get the calculated rotation matrix and translation vector. From the calculated rotation matrix the calculated ϕ, θ, ψ are obtained. We compare the difference between the calculated ϕ, θ, ψ and the true ϕ, θ, ψ in terms of mean

absolute error. For each experimental condition a thousand trials are done. Mismatching noise is simulated by randomly swapping one component from a pair of corresponding points. The percent of mismatch is the ratio of mismatching points to number of corresponding points.

In our experiments, the number of corresponding point pairs varies from the 8-point pairs to 110-point pairs in 4 steps. Figure 2-5 show the translation error and rotation degree error, which can define an average of mean absolute error of three Euler angles, versus the signal to noise ratio for different numbers of corresponding point pairs for both Gaussian noise and Uniform noise. When noise-free, the linear algorithm has excellent performance with zero error for all cases and the error increases as the noise level increases. Furthermore, depending on kind of noise and number of corresponding point pairs, the error increases very rapidly when the signal to noise ratio gets below a knee value. Table 1 shows the minimum signal to noise ratio to guarantee a less than 1 degree error as a function of numbers of corresponding point pairs and kind of noise distribution.

The robust experiments show that the robust estimators can protect from outliers almost up to a fraction of 50 percent. The linear algorithm breaks down when only a small percent of outliers is present. Similar results occur in the mismatch experiments. Fig. 6 a.b.c.d. shows the effect of outliers to both the linear and robust algorithm. The error of the linear algorithm almost increases linearly, but the robust algorithm shows much better performance and stability. The error of ψ is approximately twice less than the error for θ and ϕ The azimuth and tilt angle are more vulnerable to noise than swing angle.

In Fig. 7 a.b.c.d. we fix the percent of outliers and increase the number of corresponding points. Because the outlier percentage is constant, the mean error is approximately constant as the number of corresponding points increase. The mismatch error results are shown The mismatch error results are shown in Fig. 8 a.b.c.d. They show results similar to the outlier results. Fig. 9 shows the standard deviation of the points plotted in Fig. 6, Fig. 7, and Fig. 8. The behaviors of the standard deviation of the three rotation angles are similar, hence we put them together and take average.

5. Summary of Robust algorithm

Step 0. Use the identity matrix for initial weight matrix.

Step 1. Use singular value decomposition to solve Eq.(17)

Step 2. Update the weight matrix by Eq.(16) and Eq.(19)

Repeat Step 1. and 2. until the criteria satisfied.

Step 3. Determine the translation vector from Eq.(8), Eq(9), and Eq.(10)

Step 4. Obtain true R_o from Eq.(11)

6. Conclusion

The noise behavior for the general linear motion algorithm and its robust version was determined from over hundred thousand experimental trials. The experimental results indicated that the robust algorithm can extract the 3-D motion parameters with one degree rotation mean absolute error from image sequences which contain 30 percent of outliers. This is much better than the linear algorithm which has more than ten degree rotation mean absolute error. The robust algorithm can detect the outliers, mismatching errors, and blunders up to 30 % of observed data. Therefore, it can be an effective tool in estimating 3-D motion parameters from multiframe time sequence imagery. It should prove equally effective when applied to image flow data.

Table 1. SNR (db) for mean absulute error in 1 degree.

	Rotation Angles				Translation Vector			
No. of Point Pairs	8	20	50	110	8	20	50	110
Gaussian	75	57	52	50	105	78	73	68
Uniform	74	56	52	49	106	78	72	68

References

[1] Hoaglin, C.F. Mosteller, and J.W.Tukey, "Understanding Robust and Exploratory Data Analysis," Johy Wiley & Sons, Inc., 1983 pp. 348-349.

[2] Fang, J.Q., and T.S. Huang, "Some experiments on estimating the 3D motion parameters of a rigid body from two consecutive image frames," IEEE Trans. Pattern Analy. Mach. Intell., PAMI-6, 1984, pp. 547-554.

[3] Peter J. Huber, "Robust Statistics," John Wiley & Sons, 1981.

[4] Robert V. Hogg, "An Introduction to Robust Estimation," Robustness in Statistics edited by R.L. Launer & G.N. Wilkinson, Academic press, 1979.

[5] Thompson,E.H. "A Rational Algebraic Formulation of The Problem of Relative Orientation," The Photogrammetric Record, Vol.III No.14 1959 pp. 152-159

[6] Granshaw S.I. "Relative Orientation Problems," Photogrammetric Record, 9(53) 1979, pp.669-674 Pope, J.A.,1970 An Advantageous, Alternative Parametrization of Rotations for Analytical Photogrammetry. ESSA Tech. Rep., C and GS 39.

[7] Longuet - Higgins C., "A computer Algorithm for Reconstructing a scence from two projectives". Nature 293, 133 - 135(1981)

[8] Tasi R.Y. and T.S. Huang, "Uniqueness and Estimation of 3-D Motion Parameters of Rigid Objects With Curved Surfaces," IEEE Trans. Pattern Anal. Mach. Intell. PAMI-6, 13-17(1984)

[9] Roach, J.W., and J.K. Aggarwal, "Determining the movement of objects from a sequence of images," IEEE Trans. Pattern Analy. Mach. Intell., PAMI-6, 1980, pp. 554-562.

[10] Xinhua Zhuang, Thomas S. Huang, and Robert M. Haralick, " Two-View Motion Analysis: A Unified Algorithm," J. Opt. Soc.Am. A/Vol. 3, No.9 1986 1492 - 1500

[11] Xinhua Zhuang and R. M. Haralick " A Simplification to Linear Two View Motion Algorithms," To be Published

[12] Huber, P.J., "Robust Statistics" New York: John Wiley and Sons, Inc.. 1981

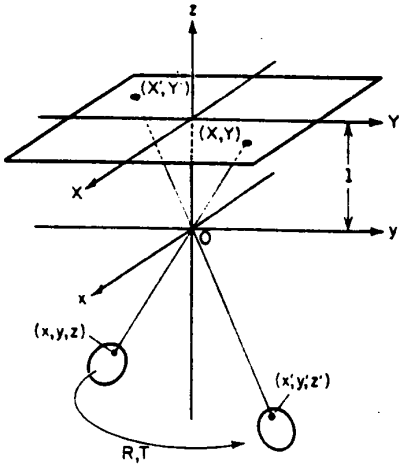

Figure 1: Imaging Geometry

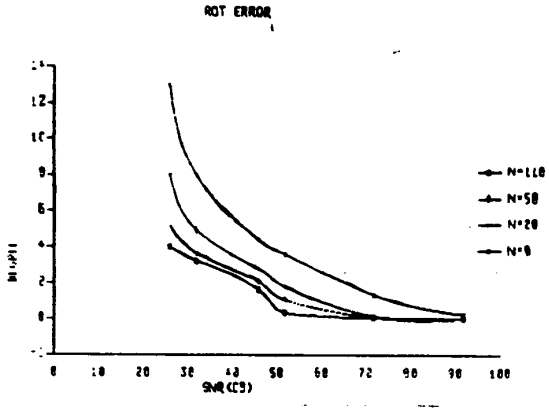

Figure 2: Mean angle error between the estimated rotation angles and the true rotation angles versus the Gaussian noise level for four corresponding point data set sizes of 8 to 110 pairs. Each point on the graph represents 1,000 trials.

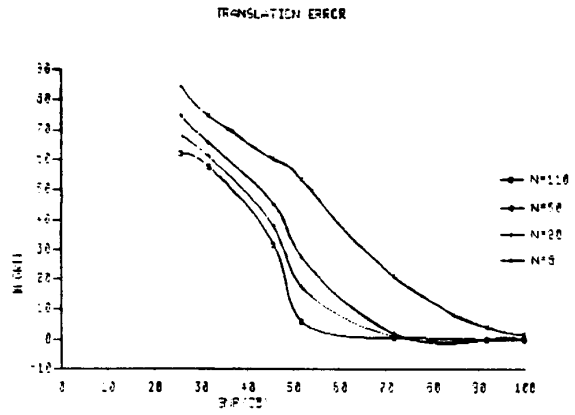

Figure 3: Mean angle error between the calculated translation vector and the true translation vector versus the Gaussian noise level for four corresponding point data set sizes of 8 to 110 pairs. Each point on the graph represents 1,000 trials.

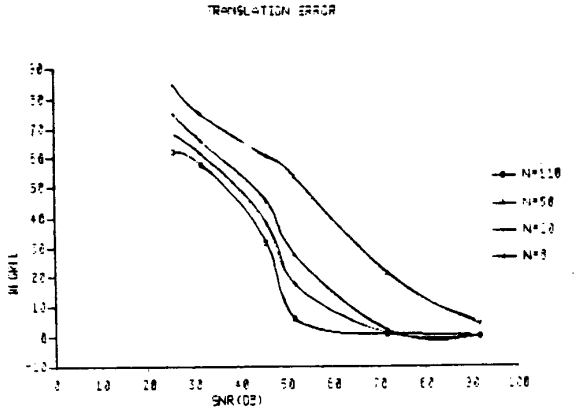

Figure 5: Mean angle error between the estimated translation vector and the true translation vector versus the Uniform noise level for four corresponding point data set sizes of 8 to 110 pairs. Each point on the graph represents 1,000 trials.

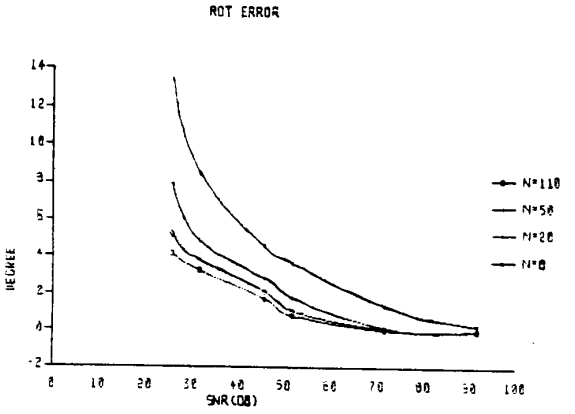

Figure 4: Mean angle error between the estimated rotation angles and the true rotation angles versus the Uniform noise level for four corresponding data set sizes 8 to 110 pairs. Each point on the graph represents 1,000 trials.

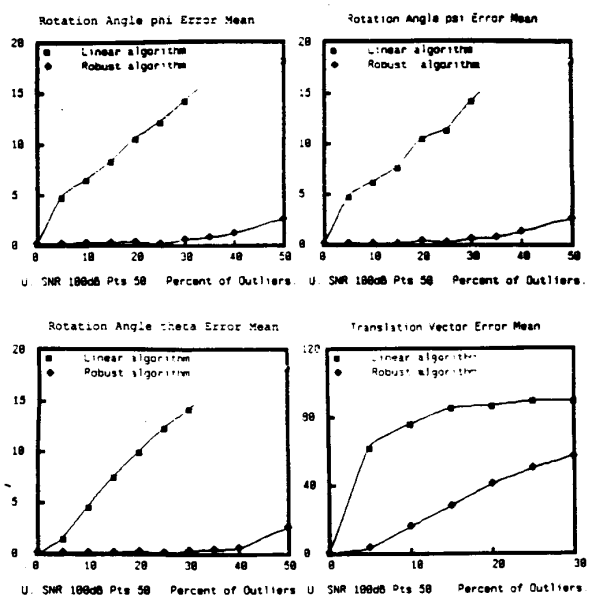

Figure 6: a.b.c.d. Compares the ϕ, ψ, θ angle error and translation angle error between the linear algorithm and robust algorithm for different percent of outliers. The noise is uniform with 100dB SNR. The number of points is 50. Each point on the graph represents 1,000 trials.

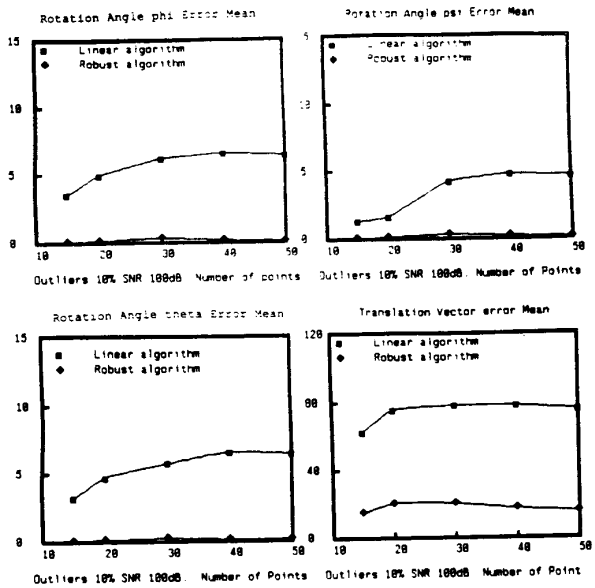

Figure 7: a.b.c.d. Compares the ϕ, ψ, θ angle error and translation angle error between the linear algorithm and robust algorithm for different number of points. The noise is uniform with 100dB SNR . The percent of outliers is 10 %. Each point on the graph represents 1,000 trials.

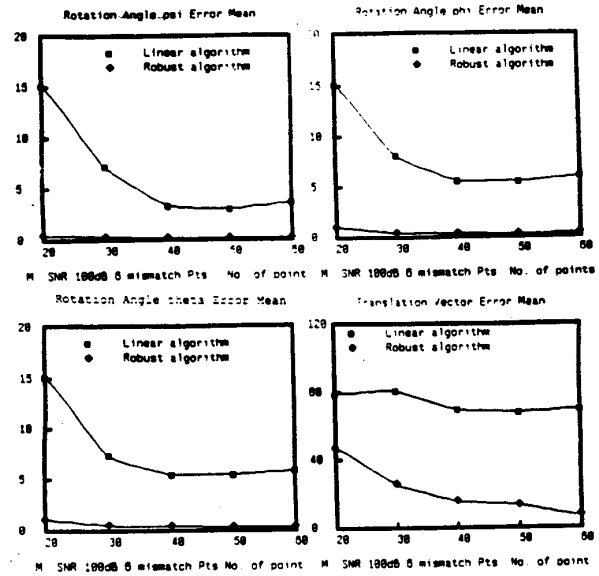

Figure 8: a.b.c.d. Compares the ϕ, ψ, θ angle error and translation angle error between the linear algorithm and robust algorithm for different number of points. The noise is uniform with 100dB SNR and is added six points of mismatch. Each point on the graph represents 1,000 trials.

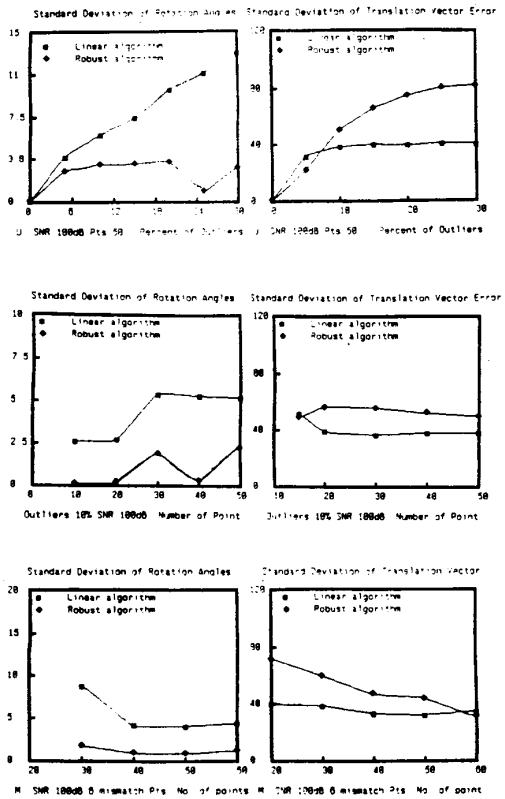

Figure 9: a.b.c.d.e.f. The standard deviation of the points plotted in Fig.6, Fig.7, Fig.8