Robust estimation of the fundamental matrix and stereo correspondences

Nuno Gracias and José Santos-Victor *

Instituto Superior Técnico & Instituto de Sistemas e Robótica
Av. Rovisco Pais,
1096 Lisboa Codex, Portugal

Abstract. This paper addresses the problem of robustly estimating correspondences between stereo images, together with the computation of the Fundamental Matrix - F. The fundamental matrix describes the epipolar geometry of a pair of stereo images, and a number of algorithms have been described in the literature to estimate this matrix. However, most of these algorithms assume that a number of correct point correspondences has been established a priori, which is hardly ever the case in practice.
In the paper, we focus on the joint estimation of point correspondences and F, in the practical situation where a number of points may correspond to erroneous matches. We compare various methods for the estimation of the fundamental matrix using a robust procedure to select good point correspondences from a larger data set. A new algorithm based on the Least Median of Squares is presented. It compares favourably to various existing ones, both in accuracy and computational effort.
The robust estimation of F and point correspondences is a determinant step for numerous applications like 3D reconstruction and video mosaicing, that are discussed in the paper.

1 Introduction

Estimating point correspondences from a pair of stereo images is a first step for many applications, such as 3D reconstruction. This problem can be simplified once the epipolar geometry has been recovered.
The fundamental matrix is the most powerful tool in the analysis of images pairs taken from uncalibrated cameras. It encapsulates all the available information on the camera geometry and can be computed from a set of point correspondences. Using the fundamental matrix it is possible to recover the structure of the scene up to a projective transformation, and to integrate available ground information for recovering of metric structure. One of the main problems on the estimation is the fact that the fundamental matrix can be very sensitive to error on the point locations.
Much effort has been taken on the study of the fundamental matrix over the last few years[10, 9, 4]. Several algorithms have been proposed which try to minimise the problems due to errors on the point locations and mismatches. Any automated stereo system which is not robust to this kind of errors is doomed to fail on real world applications.
This paper reports a set of implementations for the fundamental matrix computation, bearing in mind the effects of the most common types of point correspondence errors: small feature location errors and mismatches. Several algorithms are presented and tested. The uncalibrated reconstruction problem is also addressed and illustrated. All presented implementations are based on the linear least squares criteria. Non-linear techniques are left to a forthcoming paper.
In Section 2, we revise the main ideas related to the fundamental matrix and describe various estimation methods to determine the fundamental matrix and compare their sensitivity relative to numerical errors. In Section 3 we consider the existence of both pixel localization errors and mismatched points. Then, we introduce a method that robustly estimates the fundamental matrix and rejects bad correspondences from the input data. Results are described and compared. We assume that the images have enough structure information to allow the estimation of the unique fundamental matrix, i.e., there are no degenerate cases.

* Email: {ngracias,jasv}@isr.ist.utl.pt. The work described in this paper has been supported by the Portuguese PRAXIS XXI project, E.C. VIRTUOUS Inco-Copernicus Proj. 960174, and NARVAL Esprit-LTR Proj. 20185

When such occur, more than one fundamental matrix can explain the correspondences equally well. This is the case of planar scenes or pure camera rotation between images. An estimation problem arises from the fact that outliers can make degenerate data appear non-degenerate thus producing a wrong fundamental matrix. The problem of degeneracy will not be considered here. For robust detection of degenerate configurations refer to [14].
In Section 4 we tackle the problem of retrieving the cameras projection matrices from an estimated fundamental matrix. In this section we generalize to some extent the work of Hartley, by considering general projection matrices which include all the intrinsic parameters. Finally, in Section 5, we draw some conclusions and establish further directions of work.

2 Linear Estimation the Fundamental Matrix

The epipolar constraint is well known in the analysis of stereo images. Given a projection of a 3D point in one image, the corresponding projection in the second image is constrained to lie on the epipolar line. The epipolar lines depend both on the cameras intrinsic parameters and on the relative orientation and position between both view points. Thus, recovering the epipolar constraints is of the utmost importance for stereo matching and reconstruction. In projective coordinates, the epipolar constraint [4] can be written as :

$$u'^T F u = 0 \tag{1}$$

where $\{u, u'\}$ are the projective coordinates of two corresponding points in the two images and F is a (3x3) matrix, defined up to a scale factor, known as the Fundamental matrix [9]. From Equation (1), F can be estimated linearly, given a minimum of 8 corresponding points between two images, without involving camera calibration.
The fundamental matrix has rank 2, and since it is defined up to a scale factor, there are 7 independent parameters. The null space of F and its transpose F^T are the homogeneous representations of the epipoles. A set of parameters arranged to have geometric meaning are presented in [9]: the position of the epipoles and three coefficients of the homography between the two pencils of epipolar lines.

2.1 Eight Point Algorithm and Least-Squares

A linear 8-point algorithm was introduced by Longuet-Higgins [8] which can be used for computing the fundamental matrix. The linear constraint imposed by Equation (1) on the elements of F can be expressed as

$$a^T f = 0 \tag{2}$$

where a is a vector containing the coordinates of a matched pair of points, and f are the 9 entries of F organised as 9- vector. The set of eight linear constraints can be arranged to form the system $Af = 0$, where A is an (8x9) matrix. The solution vector, f, is the null space of the columns of A. In real-world applications, eight matched points will not suffice, as the resulting matrix F is bound to have large errors, due to inaccuracies on the point locations [9, 7]. Using more than eight point leads to an over-determined set of equations which can be formulated as the classical minimisation problem

$$\min_f \| Af \| \ \text{ constrained to } \| f \| = 1 \tag{3}$$

The solution to this problem is the eigenvector associated with the smallest singular value of A. A suitable algorithm for finding the eigenvector is the Singular Value Decomposition [12]. This solution has the advantage of being non iterative, allowing very fast implementation. The main drawback is that it is extremely sensitive to noise.
Except for very accurate point locations, the F matrix will not have rank two. Since most applications rely on this property, one has to force the rank constraint afterwards. A simple way to do this [7], is to replace F by the rank deficient matrix F', which minimises the Frobenius distance. The F matrix can be easily computed by zeroing the smallest singular value of the SVD decomposition of F, as suggested in [15]. Care has to be taken when enforcing this constraint. Since it affects all entries of F by approximately the same amount, the smallest values of F (which are the most important) tend to undergo the largest relative changes, thus affecting the estimated geometry.

2.2 Normalisation of the Input

The error sensitivity of the linear criteria can be reduced by applying an input coordinate transformation, prior to the computation of F. On fundamental matrix estimation procedures, the coordinate origin is often placed on the top- left corner of the image. This does not promote coordinate homogeneity, and affects the condition of the system matrix A. A simple procedure of coordinate transformation is presented in [7] which performs coordinate translation and scaling. By moving the origin to the point mass centre on both images, and by scaling the axes so to have unitary standard deviation, one can improve the problem condition and stability. Moreover, the transformation has the advantage of better balancing the elements of F. The singularity enforcement is therefore less disruptive.
Experimental evidence indicates that the normalised version of the least-squares can sometimes outperform iterative techniques, while maintaining a much lower level of algorithmic complexity. The results presented on this paper, which are all based on the linear criteria, were obtained using input normalisation.

3 Robust Matching Selection and Outlier Detection

This section describes several implemented algorithms and tests. A comparison on their robustness under the presence of small pixel localization error and gross mismatches is presented. The two types of errors are treated separately.

3.1 Pixel localization errors

When estimating the fundamental matrix from a number of point correspondences, one has to consider the existence of pixel localization errors which will influence the quality of the estimates. These errors are due to the limited resolution of the images and the feature extractors. Here we study the influence of the localization uncertainty on the estimation of the fundamental matrix.
We have performed an experiment with synthetic 3D points and the corresponding image pixels, given two camera projection matrices. All the correspondences are correct, in the sense that they are projections of the same 3D points, and Gaussian noise was added to simulate the localization noise. In Figure 1, we show the average distance to the epipolar line, as a merit criterion, against the standard variation of the localization noise, when different procedures for data normalization are used.

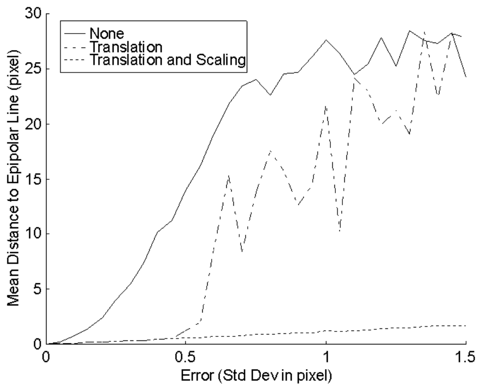


Fig. 1.: Sensitivity to feature localization errors, for various estimation methods, without data normalization (None), translating the data centroid to the origin (Translation), and translating and scaling the point coordinates (Translation and scaling)

In Figure 2, we show results using real image data. In this case, to evaluate the accuracy of the estimated fundamental matrix, we present the set of epipolar lines, when using least-squares or normalised least-squares. The correct epipolar lines are along the image scanlines.

(a) (b) (c)

Fig. 2.: Effect of normalisation using real data. (a) - Least Squares, (b) - Normalised Least Squares (translation and scaling), (c) - Correct epipolar geometry

These examples show the significant difference of the estimation results when pre-conditioning the data prior to the application of a least squares method. In fact, by normalising the input data coordinates, the sensitivity to noise is greatly reduced, as shown in Figure 1. Nevertheless, in real data, the pencil of epipolar lines that is obtained is not entirely satisfactory, as shown in Figure 2.

3.2 Mismatch errors

We will now consider the effect of mismatched points, which happen in real applications, on the fundamental matrix estimation. These errors can be due to occluding objects or severe light conditions changes. The tested algorithms can be classified under three categories: Smoothing, Case deletion and Random Sampling. Only the latter two exhibit a degree of robustness. The only smoothing algorithm used is a version of the normalised least-squares, included here for comparison purposes. The case deletion methods are based on the elimination of the data points according to some evaluation criterion, and two algorithms of this type were tested:

LSRes: F is estimated using all the point correspondences; then we select the 8 pairs of points that correspond to the least residual in Equation (2) and re-estimate the fundamental matrix.

CDDist: This method consists of $N-8$ iterations where we estimate F using least squares (LS). In each iteration we discard the correspondence pair that correspond to the largest distance to the epipolar line.

The Random Sampling Consensus (**RANSAC**) paradigm was introduced in the Computer Vision literature in 1981 by Fischler and Bolles [5]. The main idea is to find a significant set of data points consistent with an instantiation of the followed model, i. e., to find a consensus group. The remaining data points are rejected as outliers. The RANSAC algorithm is best illustrated by the following steps :

1. Sample all the matched points for a set of the minimum number of pairs required for the model instantiation (thus 8 for a linear computation), and compute the matrix F.
2. For a given distance threshold d_T, select all the pairs whose distance to the respective epipolar lines is smaller than d_T. This is the consensus group for this F.
3. If the number of selected pairs is greater than an initial estimate of the number of correct data points, then compute the F matrix based on all the selected pairs, using a smoothing technique (like the least squares).
4. Otherwise, repeat steps 1 to 3 up to a specified number of iterations, and return F computed with the largest consensus group found.

This paradigm, as originally proposed, requires the specification of three parameters: the error tolerance used to discard data points; the number of subsets to test and the threshold d_T. Another type of robust regression, the Least Median of Squares (LMedS) was proposed by Rousseeuw [13, 11] in 1984. Instead of finding a large consensus group, the LMedS estimates the model parameters by minimizing the median of

the squared residuals computed for the entire data set. Since the space of the possible estimates for the fundamental matrix is too large, it is common practice to constrain the analysis to a randomly sampled subspace. Results on the use of LMedS for the estimation of the fundamental matrix can be found in [2].
We will now propose a variant to LMedS - **MEDSERE**[2] - using case deletion and random sampling, with the following algorithm :

1. Randomly sample the complete set of matched points S_{total} for a set of 8 pairs.
2. Estimate the F matrix and compute the median of the point-to-epipolar line distance for S_{total}. If the median is below a given threshold, return F and exit.
3. Repeat 1. and 2. for a specified number of samples N_1.
4. Select the F matrix for which the minimal median was found, and sort the matched points by their point- to-epipolar line distance, using F.
5. Create the set S_{best} with the elements of S_{total} whose distance is below the median.
6. Repeat 1. and 2. on S_{best} for a N_2 number of samples.
7. Return the minimal median matrix found.

The required parameters are the number of samplings on each part N_1 and N_2, and the median threshold. Since the first two directly determine the number of operations, they are best defined by processing time constraints.
Figure 3 shows the performance of the described algorithms on the presence of mismatch errors, ranging from 0 to 60% of the total number of pairs. The images used are projections of a synthetic shape with 100 3D points. Averages of 30 runs for each algorithm were taken. The evaluation criteria on the first plot is the median distance of each point pair to the corresponding epipolar lines, for the whole set including mismatched data. To better assess the performance, the second and third plots show the mean (b) and the variance (c) of the point-to-epipolar line distance, for the original error-free data set.

One can see that the best performant algorithms are MEDSERE and LMedS. MEDSERE is capable of good results even under conditions as severe as 60% of mismatches, slightly outperforming LMedS. The RANSAC also presents good results up to 50% mismatches, but fails shortly afterwards. The other methods perform poorly.

Another relevant aspect when comparing the performance of different algorithms is to characterize the required computational effort. Figure 4 show the approximate number of floating point operations of LMedS and MEDSERE in the presence of a variable number of mismatched points. The maximum number of samples and the median threshold are the same for the two methods. In this aspect, MEDSERE compares favourably with LMedS. For a more complete survey on robust methods, refer to [11] and [13].

4 Uncalibrated Reconstruction

We will now consider the problem of uncalibrated reconstruction. There are many applications where the camera calibration parameters are not available, thus preventing metric reconstruction. It is known [3, 4] that if the cameras intrinsic parameters information is available, then it becomes possible to determine the camera placements from the fundamental matrix, and thus achieve Euclidean reconstruction.
Let us consider a generic camera projection matrix $P = (M \mid -MT)$, where M is the 3x3 matrix on the left side of P, and T is a 3x1 vector. If the camera is not placed at the infinity then M is invertible [6]. The vector T contains the 3D coordinates of the camera centre expressed on the world coordinate system. Therefore it is straightforward to determine the camera centre from P, if the camera is not placed at infinity. The F matrix can be directly related to a pair of camera matrices $P_1 = (M_1 \mid -M_1T_1)$ and $P_2 = (M_2 \mid -M_2T_2)$ according to :

$$F \approx M_2^* M_1^T [M_1(T_1 - T_2)]_\times \tag{4}$$

where $\approx$ stands for equality up to a scale factor; $[t]_\times$ is a 3x3 skew-symmetric matrix implementing the vector product: $t \times p = [t]_\times p$; and M_2^* is the adjoint of M_2. If only finite camera locations are considered then $M_2^* = M_2^T$. For more details refer to [6]. An important result (Hartley, 1992), is that two pairs of camera matrices $\{P_1, P_1\}$ and $\{P_1', P_1'\}$ relate to the same F, if and only if there is a 4x4 non singular matrix H so

[2] MEDSERE stands for MEDian SEt REduction.

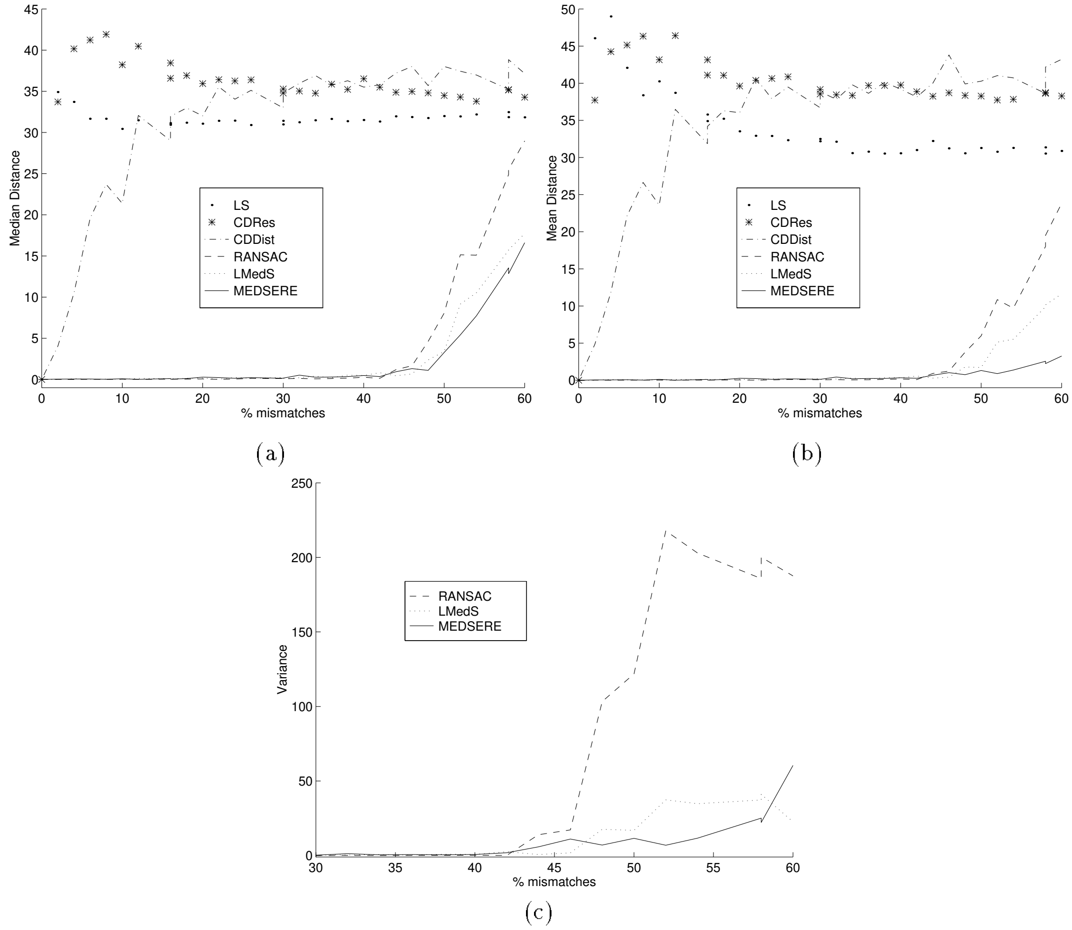


Fig. 3.: Performance of the algorithms under mismatch error conditions

that $P_1 H \approx P_1'$ and $P_2 H \approx P_2'$.
Let us consider now any two projection matrices P_1' and P_2' agreeing with an estimated F. Using P_1' and P_2', it is easy to compute the coordinates of 3D points x_i' from the image projections. In a general case these points differ from the original 3D points x_i, by the H transformation,

$$P_1' x_i' = P_1 H x_i' = P_1 x_i \quad \Rightarrow \quad x_i = P_1 x_i' \tag{5}$$

The H matrix is a 3D general perspective transformation [1] accounting for linear geometric rigid (rotation and translation) and non-rigid operations (scaling and skewing). It is defined up to a scale factor, and has 15 independent entries on the general case. The transformation H can be recovered from a set of points x_i', whose ground-truth 3D coordinates are known. We can now present an Euclidean reconstruction procedure, based on the use of ground points:

1. Estimate the fundamental matrix F from a set of matched points.
2. Determine some P_1' and P_2' agreeing with F.
3. Recover the 3D structure using P_1' and P_2'.
4. Estimate the H matrix by the use of ground points.
5. Apply H to the points recovered in 3.

The problem remaining to be solved is to determine a pair of camera projection matrices agreeing with the estimated F. We will now present a lemma which can be used directly to obtain such pair.

4.1 Recovering the projection matrices from F

In [6, 7] a method is presented to recover two projection matrices in accordance to an estimated fundamental matrix, F. Although not stated clearly, there is the assumption that such cameras have normalized intrinsic

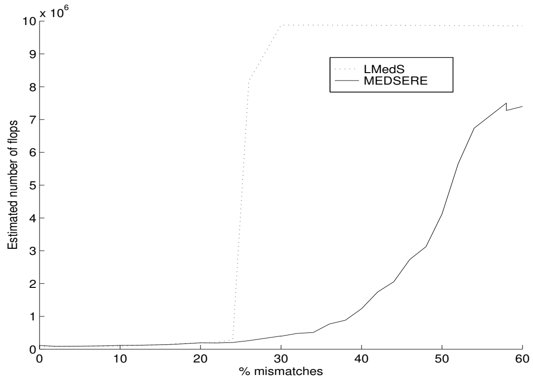


Fig. 4.: Evolution of the computational cost for the LMedS and MEDSERE algorithms

parameters and thereby the matrix M, described in equation (4) is a rotation matrix. In this section we generalize these results for the case of arbitrary cameras.

Lemma 1. *Let F be a fundamental matrix with SVD decomposition $F = U\,diag(r,s,0)V^T$, and $r \geq s > 0$. Let $P_1 = (M_1 \mid -M_1T_1)$ be an arbitrary camera matrix with $det(M_1) \neq 0$. Then the matrices,*

$$P_1 = (M_1 \mid -M_1T_1) \qquad\qquad P_2 = (M_2 \mid -M_2T_2)$$

correspond to the fundamental matrix F, where M_2 and T_2 are given by :

$$M_2 = [U\,diag(r,s,\gamma)EV^T]^{-T}M_1 \qquad \gamma \in \Re^+$$

$$T_2 = M_1^{-1}V[0,0,1]^T + T_1 \qquad E = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Proof. It is known that the null space of the columns of F is a 3x1 vector corresponding to the projective coordinates of the right epipole. According to equation (4),

$$F \approx M_2^{-T}M_1^T[M_1(T_1 - T_2)]_\times$$
$$F \approx U\mathrm{diag}(r,s,\gamma)EV^TM_1^{-T}M_1^T[M_1(M_1^{-1}V[0,0,1]^T + T_1 - T_1)]_\times = U\mathrm{diag}(r,s,\gamma)EV^T[V[0,0,1]^T]_\times$$

By construction of the SVD, the last column of V is a base vector for the null space of UDV^T, i.e., the null space of F, as required. The columns form a orthonormal basis spanning the 3D space. Bearing this is mind, it is easy to verify by inspection that

$$[V[0,0,1]^T]_\times = VZV^T, \text{ with } Z = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Performing this replacement, we will have,

$$F \approx U\mathrm{diag}(r,s,\gamma)EV^TVZV^T \approx U\mathrm{diag}(r,s,0)EV^T = SVD(F)$$

thus validating the lemma.

A simpler version of this lemma, on which this was based, is presented in [6] where M is a rotation matrix, hence simplifying the derivation process. A number of tests were conducted, using the reconstruction procedure on synthetic and real data. The synthetic data consists of a set of 3D points that are projected on two images given a pair of camera projection matrices. Gaussian noise is then added to the image coordinates, as localization noise. Then the fundamental matrix is estimated as described in the previous sections; P_1 and P_2 are computed and the 3D points reconstructed. Figures 5(a) and (b) show the 3D shape used for tests, with superimposed reconstructed points. The left image was obtained with no noise, and the right one with added zero-mean, 1.2 pixel standard deviation Gaussian noise. The effects of increasing amounts of point location noise are shown in Figure 5(c), again using zero mean Gaussian noise. The real data experiment reported here was conducted on the images of Figure 6. A calibration grid was used to provide the ground points with known locations. No explicit camera calibration was required or computed. The Euclidean reconstruction result is presented on Figure (b). Using 21 ground points and coordinate normalisation prior to the computation of the H matrix, the scene was recovered with a maximum error of 10% on the 3-D coordinates.

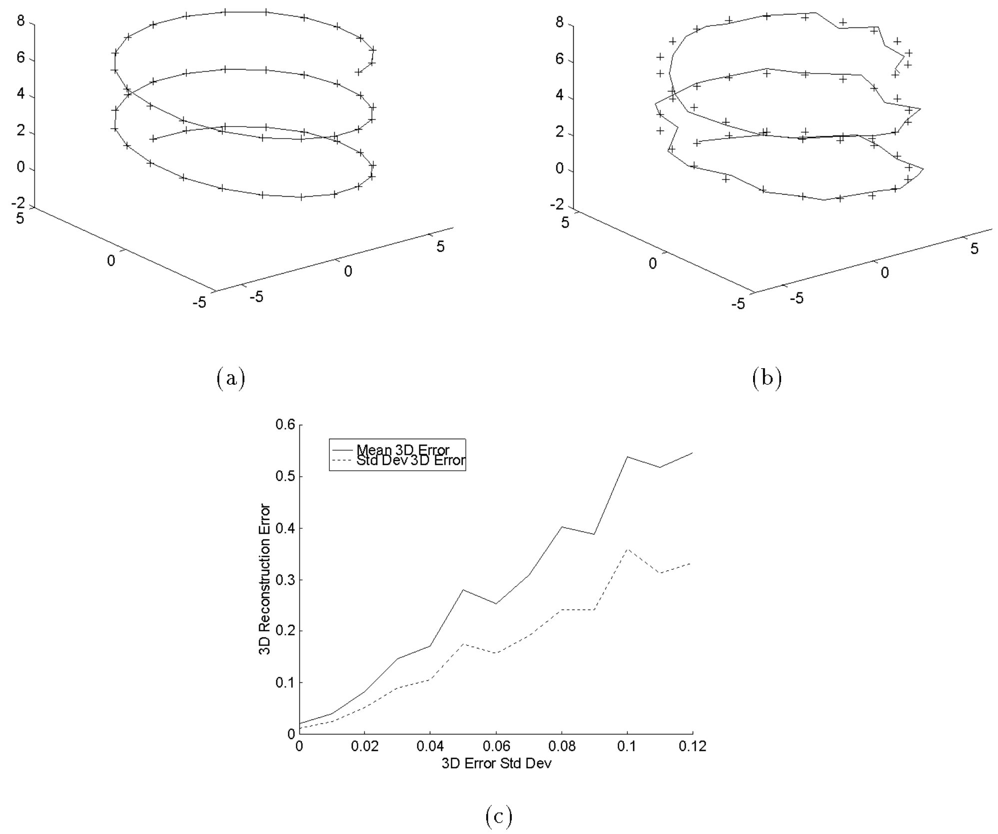
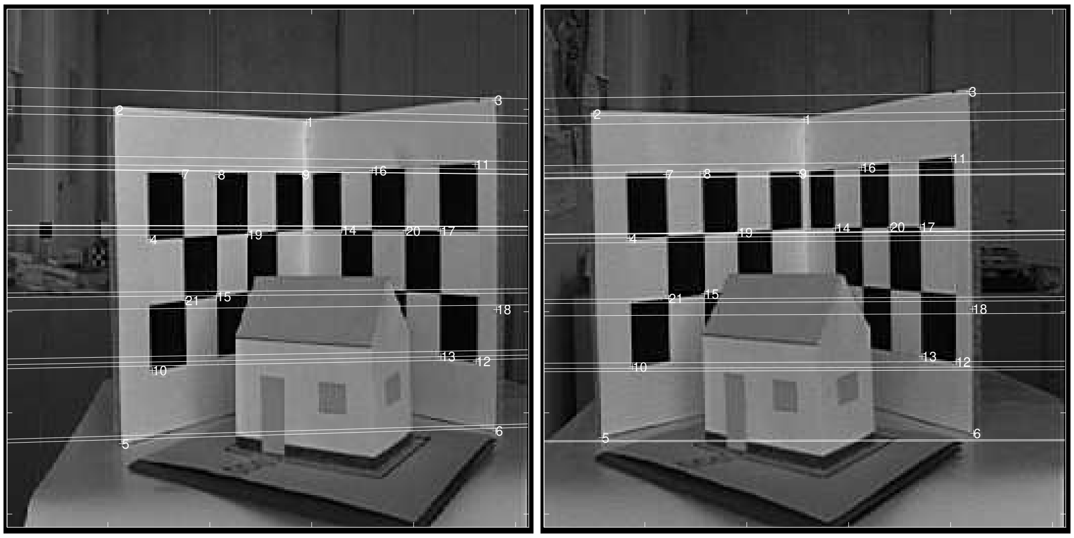


Fig. 5.: Euclidean reconstruction examples, with no added noise(a) and with additive Gaussian noise(b). The + marks indicate the original shape points. In (c) we show the effect of the noise in the reconstructed points.

5 Conclusions

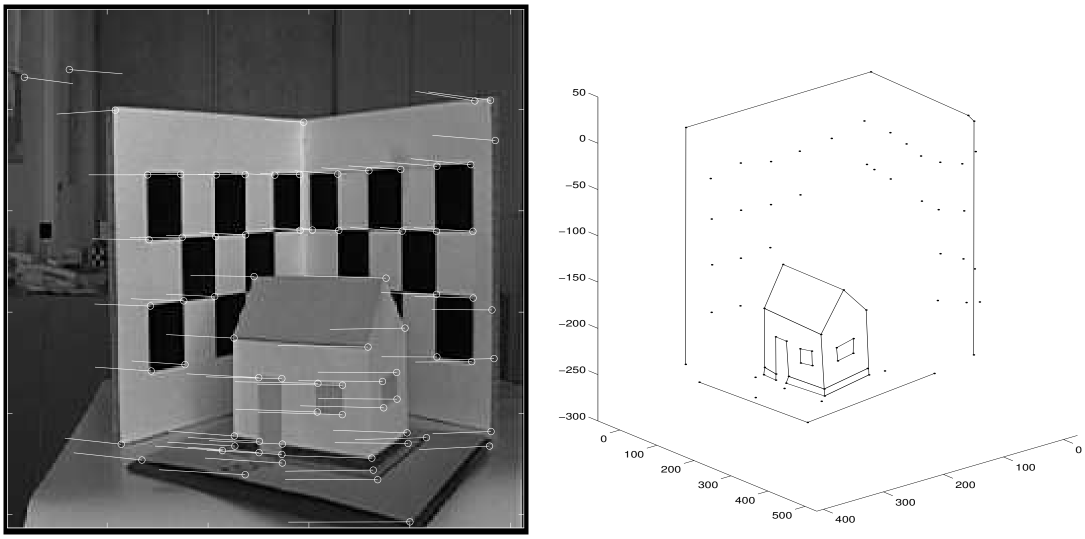
In the paper, we have described a procedure to estimate the fundamental matrix and point correspondences in a robust way. Rather than considering that a number of correct correspondences are available, we assume that the matching results contain a percentage of erroneous points, which is always the case in a practical system. We compare various algorithms to estimate F, in presence of pixel noise and wrong matches. Additionally we propose a new algorithm which shows encouraging results both in terms of the achieved accuracy and the required computational effort. For the reconstruction, we show how to determine two projection matrices from an estimated fundamental matrix, which to some extent generalises the framework proposed by Hartley.

These techniques are intended to be used in the future, for 3D reconstruction and video mosaicing from a



(a) (b)

Fig. 6.: Toy house setup used for reconstruction with real images. The matched points ares showned with the corresponding epipolar lines.



(a) (b)

Fig. 7.: Euclidean reconstruction with real images: disparity(a) and reconstruction(b) with superimposed lines.

collection of images acquired by an underwater vehicle.

References

1. D. Ballard and C. Brown. *Computer vision.* Prentice-Hall, London, 1982.
2. R. Deriche, Z. Zhang, Q.-T. Luong, and O.D. Faugeras. Robust recovery of the epipolar geometry for an uncalibrated stereo rig. In *Proc. of the European Conference on Computer Vision*, pages 567–576, Stockholm, Sweden, 1994.
3. O. Faugeras. What can we see in three dimensions with an uncalibrated stereo rig ? In *Proc. of the 2nd. European Conference on Computer Vision*, Santa Margherita, Italy, May 1992.
4. O. Faugeras. *Three Dimensional Computer Vision.* MIT Press, 1993.

5. M. Fischler and R. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 6(24):381–395, 1981.
6. R. Hartley. Stereo from uncalibrated cameras. In *Proc. of the IEEE Int. Conference on Computer Vision and Pattern Recognition CVPR*, Urbana Champaign, USA, June 1992.
7. R. Hartley. In defence of the 8-point algorithm. In *Proc. of IEEE Int. Conference on Computer Vision*, Cambridge MA, USA, June 1995.
8. H.C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.
9. Q. Luong, R. Deriche, O. Faugeras, and T. Papadopoulo. On determining the fundamental matrix: an analysis of different methods and experimental results. Technical Report 1894, INRIA, 1993.
10. Q. Luong and O. Faugeras. The fundamental matrix: Theory, algorithms, and stability analysis. *International Journal of Computer Vision*, 17(1):43–75, January 1996.
11. P. Meer, D. Mintz, A. Rosenfeld, and D. Kim. Robust regression methods for computer vision: a review. *Int. Journal of Computer Vision*, 6(1):59–70, 1991.
12. W. Press, B. Flannery, S. Teukolvsky, and W. Vetterling. *Numerical Recipes in C: The art of Scientific Computing.* Cambridge University Press, 1988.
13. P. Rousseeuw and A. Leroy. *Robust Regression and Outlier Detection.* John Wiley & Sons, 1987.
14. P. H. Torr, A. Zisserman, and S. J. Maybank. Robust detection of degenerate configurations for the fundamental matrix. In *Proc. of the 5nd. IEEE International Conference on Computer Vision*, 1995.
15. R. Tsai and T. Huang. Uniqueness and estimation of three dimensional motion parameters of rigid objects with curved surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6):13–27, 1984.

This article was processed using the TEX macro package with SIRS97 style