Int. J Computer Vision, , 1–33 ()

The Development and Comparison of Robust Methods for Estimating the Fundamental Matrix

P H S TORR AND D W MURRAY

[phst,dwm]@robots.ox.ac.uk

Department of Engineering Science, University of Oxford, Parks Road, Oxford, OX1 3PJ, UK

Received 17th August 1995. Revised 10th July 1996.

Abstract.

This paper has two goals. The first is to develop a variety of robust methods for the computation of the Fundamental Matrix, the calibration-free representation of camera motion. The methods are drawn from the principal categories of robust estimators, viz case deletion methods, M-estimators and random sampling, and the paper develops the theory required to apply them to non-linear orthogonal regression problems. Although a considerable amount of interest has focussed on the application of robust estimation in computer vision, the relative merits of the many individual methods are unknown, leaving the potential practitioner to guess at their value. The second goal is therefore to compare and judge the methods.

Comparative tests are carried out using correspondences generated both synthetically in a statistically controlled fashion and from feature matching in real imagery. In contrast with previously reported methods the goodness of fit to the synthetic observations is judged not in terms of the fit to the observations *per se* but in terms of fit to the ground truth. A variety of error measures are examined. The experiments allow a statistically satisfying and quasi-optimal method to be synthesized, which is shown to be stable with up to 50 percent outlier contamination, and may still be used if there are more than 50 percent outliers. Performance bounds are established for the method, and a variety of robust methods to estimate the standard deviation of the error and covariance matrix of the parameters are examined.

The results of the comparison have broad applicability to vision algorithms where the input data are corrupted not only by noise but also by gross outliers.

1. Introduction

In most computer vision algorithms it is assumed that a least squares framework is sufficient to deal with data corrupted by noise. However, in many applications, visual data are not only noisy, but also contain outliers, data that are in gross disagreement with a postulated model. Outliers, which are inevitably included in an initial fit, can so distort a fitting process that the fitted parameters become arbitrary. This is particularly severe when the veridical data are themselves degenerate or near-degenerate with respect to the model, for then outliers can appear to break the degeneracy.

In such circumstances, the deployment of *robust* estimation methods is essential. Robust methods continue to recover meaningful descriptions of a statistical population even when the data contain outlying elements belonging to a different population. They are also able to perform when other assumptions underlying the estimation, say the noise model, are not wholly satisfied.

Amongst the earliest to draw the value of such methods to the attention of computer vision researchers were Fischler and Bolles (1981). Figure 1 shows a table of x, y data from their paper which contains a gross outlier (Point 7). Fit 1 is the result of applying least squares, Fit 2 is the result of applying least squares after one robust method has removed the outlier, and the solid line is the result of applying their fully robust RANSAC algorithm to the data. The data set can also be used to demonstrate the failings of naïve heuristics to remove outliers. For example, discarding

x	*y*
0	0
1	1
2	2
3	2
3	3
4	4
10	2

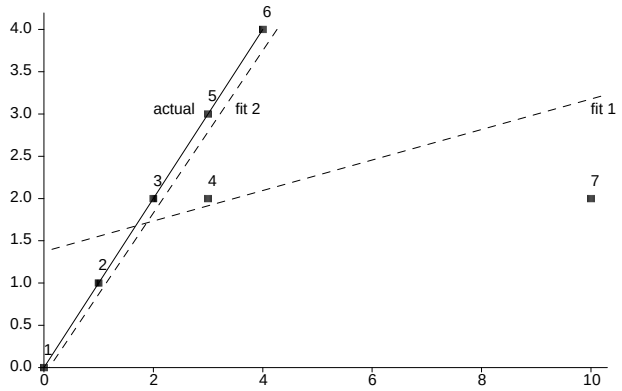

Fig. 1. A data set with an outlier comparing least squares and robust fitting.

the point with largest residual after least squares fitting removes Point 6 not Point 7. Indeed repeated application of this heuristic to convergence results in half the valid data being discarded, and Point 7 remaining as an inlier to a completely erroneous fit.

The statistical literature reports a wide variety of robust estimators (Maronna 1976; Mosteller & Tukey 1977, Cook & Weisberg 1980; Devlin *et al.* 1981; Fischler & Bolles 1981; Huber 1981; Critchley 1985; Hoaglin *et al.* 1985; Hampel *et al.* 1986; Rousseeuw 1987; Chaterjee & Hadi 1988, Roth 1993, Shapiro and Brady 1993, Torr and Murray 1993, Zhang *et. al.* 1994, Kumar and Hanson 1994, Stewart 1995). The first aim of this work is to develop a variety of methods from several categories — *viz* M-estimators, Case-deletion Diagnostics and Random Sampling — and apply them to the computation of the fundamental matrix. This in turn requires novel extensions of some of the robust estimation techniques to handle non-linear problems involving orthogonal regression.

The fundamental matrix provides a general and compact representation of the ego-motion captured in two views by a projective camera, requiring no knowledge of the camera calibration (Faugeras 1992; Hartley 1992). In the computation of the fundamental matrix, outliers typically arise from gross errors such as correspondence mismatches or the inclusion of movement inconsistent with the majority. The latter might be caused by features being on occluding contours, shadows or independently moving objects. Robust estimation impacts therefore not only on estimation, but also on data segmentation. Degeneracies in the fundamental matrix also occur frequently.

The second aim of the work is to compare the performance both of non-robust least squares methods and the range of robust methods, making both intra- and inter-category comparisons on large controlled data sets and data from real imagery. This has allowed the coupling of several robust techniques to arrive at an empirically optimal, and statistically satisfying method. The techniques used and conclusions drawn have applicability to the broad sweep of computer vision problems troubled by outlying data.

Recovery of motion, and then structure, using the fundamental matrix, and it calibrated analogue the essential matrix, has a long history. Spetsakis and Aloimonos (1991) divided research in the area into three epochs. The first was spent finding out whether the problem in the broadest terms had a solution. Once it was ascertained it had, the next epoch saw researchers devising constructive proofs of the uniqueness of the solution involving the minimum number of points (e.g. Longuet-Higgins 1981; Tsai & Huang 1984). Unfortunately these 'minimalist' algorithms were highly sensitive to noise, leading to an erroneous belief that recovery of structure and motion was essentially an ill-posed problem and that only qualitative solutions were possible. The third epoch was then directed towards minimizing the effects of noise by using more correspondences (Weng *et al.* 1989) and more images (Spetsakis & Aloimonos 1991; Weng *et al.* 1993). This has usually been done within a least-squares framework, with the concomitant difficulties highlighted

above. A fourth epoch is required, where the emphasis is on robust estimation that provide as output not only the inlying solution, but also a list of data that are in gross disagreement with it. Some previous work has been carried out using robust estimators within the context of structure from motion recovery by Torr and Murray (1993), Kumar and Hanson (1994) and Zhang *et al.* (1994). Their conclusions and ours will be discussed later.

The paper is organised as follows. Section 2 reviews best practice in least squares estimation methods for the fundamental matrix, and discusses the variety of error measures used. Section 3 describes our method for comparing both non-robust and robust estimators, together with the method of generating synthetic data. Theory for the M-estimators, Case deletion diagnostics, and Random Sampling is developed in Sections 4–6, and Section 7 comments briefly on the impracticability of Hough transforms for problems with large parameter spaces. A requirement of all robust methods is an estimation of the standard deviation, and a method to achieve this which is itself robust is given in Section 8.

The best estimators from the intra-category competitions are compared in an inter-category competition described in Section 9. Several techniques are blended into an quasi-optimal method, the results of which is demonstrated on real imagery in Section 10. Finally in Section 11 we discuss our results and draw conclusions.

2. Linear least squares methods

2.1. *The example application: the fundamental matrix*

Consider the movement of a set of point image projections from an object which undergoes a rotation and non-zero translation between views. After the motion, the set of homogeneous image points $\{\underline{\mathbf{x}}_i\}, i = 1, \ldots n$, as viewed in the first image is transformed to the set $\{\underline{\mathbf{x}}_i{}'\}$ in the second image, positions related by

$$\underline{\mathbf{x}}_i'^{\top}\mathbf{F}\underline{\mathbf{x}}_i = 0 \qquad (1)$$

where $\underline{\mathbf{x}} = (x, y, \zeta)^{\top}$ is a homogeneous image coordinate and

$$\mathbf{F} = \begin{bmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \\ f_7 & f_8 & f_9 \end{bmatrix}$$

is the Fundamental Matrix (Faugeras 1992). Although 3×3, the matrix has only seven degrees of freedom because only the ratio of parameters is significant and because $\det \mathbf{F} = 0$. Throughout, underlining a symbol $\underline{x}$ indicates the perfect or noise-free quantity, distinguishing it from $x = \underline{x} + \Delta x$, the value corrupted by noise (assumed Gaussian).

2.2. *Orthogonal Least Squares Regression: Method OR*

Ignoring for the moment the problem of enforcing the rank 2 constraint, given $n \geq 8$ correspondences this system appears an archetype for solution by linear least squares regression. Linear here refers to linearity in the parameters f_i — equation (1) written out is just

$$f_1\underline{x}_i'\underline{x}_i + f_2\underline{x}_i'\underline{y}_i + f_3\underline{x}_i'\zeta + f_4\underline{y}_i'\underline{x}_i + f_5\underline{y}_i'\underline{y}_i + f_6\underline{y}_i'\zeta + f_7\underline{x}_i\zeta + f_8\underline{y}_i\zeta + f_9\zeta^2 = 0 \ .$$

Because errors exist in all the measured coordinates x, y, x', y', orthogonal least squares (Pearson 1901) rather than ordinary least squares should be used, minimizing the sum of the squares of the distances shown in part (a) rather than (b) of Figure 2.

Consider fitting a hyperplane $\mathbf{f} = (f_1, f_2, \ldots, f_p)$ through a set of n points in $\mathcal{R}^p$ with coordinates $\mathbf{z}_i = (z_{i_1}, z_{i_2}, \ldots, z_{i_p})$, taking the centroid of the data as origin. (Centring is a standard statistical technique that involves shifting the coordinate system of the data points so that the centroid lies at the origin. The best fitting hyperplane passes through the centroid of the data (Pearson 1901).) Assuming that the noise is Gaussian and that the elements of $\mathbf{z}$ have equal variance[1], the hyperplane $\mathbf{f}$ with maximum likelihood is estimated by minimizing the perpendicular sum of Euclidean distances from the points to the plane (Pearson 1901; Kendall & Stuart 1983)

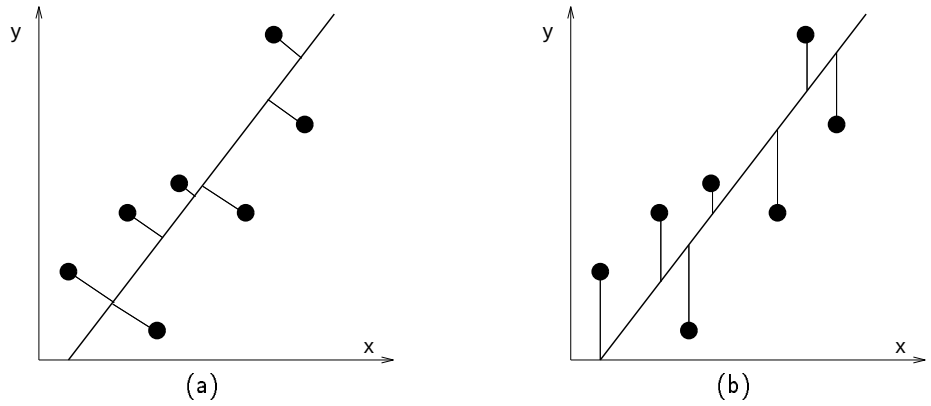

Fig. 2. The (a) orthogonal and (b) ordinary least squares distances.

$$\min_{\mathbf{f}} \sum_{i=1}^{n} (\mathbf{f}^\top \mathbf{z}_i)^2$$

subject to $\mathbf{f}^\top \mathbf{f} = 1$. The constraint ensures that the estimate will be invariant to equiform transformation — rotation, translation and scaling — of the inhomogeneous coordinates. For example the best fitting line to a 2 dimensional scatter (x_i, y_i), $i = 1 \ldots n$ is estimated by minimizing $\sum_{i=1}^{n} (ax_i + by_i + c)^2$ subject to the constraint $a^2 + b^2 = 1$ (Pearson 1901).

To reformulate as an eigen-problem, let $\mathbf{Z}$ be the $n \times p$ measurement matrix with rows $\mathbf{z}_i$, and let $\mathbf{M} = \mathbf{Z}^\top \mathbf{Z}$ be the $p \times p$ moment matrix, with eigenvalues, in increasing order, $\lambda_1 \ldots \lambda_p$ and with $\mathbf{u}_1 \ldots \mathbf{u}_p$ the corresponding eigenvectors forming an orthonormal system. The best fitting hyperplane is given by the eigenvector $\mathbf{u}_1$ corresponding to the minimum eigenvalue λ_1 of the moment matrix. It is evident that

$$\lambda_1 = \sum_{i=1}^{n} (\mathbf{u}_1^\top \mathbf{z}_i)^2 = \sum_{i=1}^{n} r_i^2$$

which is the sum of squares of residuals r_i, which in this case are the perpendicular distances to the hyperplane.

For the fundamental matrix,

$$\mathbf{z} = \left(x_i' x_i \;\; x_i' y_i \;\; x_i' \zeta \;\; y_i' x_i \;\; y_i' y_i \;\; y_i' \zeta \;\; x_i \zeta \;\; y_i \zeta \;\; \zeta^2\right)^\top$$

and the measurement matrix is

$$\mathbf{Z} = \mathbf{W} \begin{bmatrix} x_1' x_1 & x_1' y_1 & x_1' \zeta & y_1' x_1 & y_1' y_1 & y_1' \zeta & x_1 \zeta & y_1 \zeta & \zeta^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n' x_n & x_n' y_n & x_n' \zeta & y_n' x_n & y_n' y_n & y_n' \zeta & x_n \zeta & y_n \zeta & \zeta^2 \end{bmatrix}$$

where $\mathbf{W}$ is a diagonal matrix of the weights given to each feature correspondence, corresponding to the inverse standard deviation of each error. (This is assumed to be homogeneous at present. In the next section its estimation by iteratively reweighted least squares is explained.) If the variances of the image coordinates are different along the two axes, σ_x^2 and σ_y^2 say, the image coordinates are weighted by dividing them by their respective variances.

The estimate $\mathbf{f} = \mathbf{u}_1$ actually minimizes

$$\mathbf{f}^\top \mathbf{M} \mathbf{f}$$

but now subject to $\mathbf{f}^\top \mathbf{J} \mathbf{f} =$ constant, where $\mathbf{J} = \mathrm{diag}(1, 1, 1, \ldots, 1, 0)$ is the normalization chosen to realize a solution from the equivalence class of solutions with different scalings. Again, for best numerical stability, the origin of coordinate system should be placed at the data centroid. Centring the moment matrix is achieved by subtracting $\mathbf{1}\,\overline{z}_j$ from each column of $\mathbf{Z}$, where $\mathbf{1}$ is an n dimensional vector $\mathbf{1} = (1, 1, 1, \ldots, 1)^\top$ and $\overline{z}_j$ is the mean of column j.

2.3. Iteratively Re-weighted Least Squares: Methods S1 and S2

The orthogonal least squares method (OR) will in fact produce a sub-optimal estimate of $\mathbf{F}$ because the residuals that are minimized

$$r_i = f_1 x'_i x_i + f_2 x'_i y_i + f_3 x'_i \zeta + f_4 y'_i x_i + f_5 y'_i y_i + f_6 y'_i \zeta + f_7 x_i \zeta + f_8 y_i \zeta + f_9 \zeta^2 \quad (2)$$

are not Gaussianly distributed. The cause of this is now outlined, and different residuals which are more nearly Gaussianly distributed are discussed. These require a working knowledge of the solution, and so an iteratively re-weighted least squares minimization is required (Bookstein 1979; Sampson 1982).

The expression for r_i given in Equation (2) is known as the *algebraic* distance. It has no geometrical significance, and does not, for example, measure the perpendicular distance of a feature to the quadric variety represented by $\mathbf{F}$ in 4D image coordinate space (x, y, x', y'). The variance of r_i is said to be *heteroscedastic*, meaning that it depends on the location of the feature correspondences.

Sampson (1982) discovered a similar heteroscedascity when fitting algebraic residuals to conics. If each point is perturbed by Gaussian noise, minimization of the algebraic distance using the eigenvector of the moment matrix was found to be sub-optimal. It was shown by Kendall & Stuart (1983) that the best fitting, maximum likelihood, quadratic curve is such that the sum of squares of the perpendicular *geometric* distances of points to the curve is a minimum[2], as illustrated in Figure 3(b). Furthermore, this solution is invariant to Euclidean transformations of the coordinate system. The reason this problem does not arise when fitting a hyperplane to residuals that are linear in the measurements is that the algebraic and geometric distances coincide (Figure 3(a)). The joins of the different points to the conic in Figure 3(b) are neither parallel nor unique, and a closed form solution is unobtainable. In his work, Sampson proposed using a first order approximation to the distance.

Method S1. Noting that the expression for the residuals for the fits to a conic and to the fundamental matrix were both bilinear in the measurements, Weng *et al.* (1989) adapted Sampson's method to the computation of the fundamental matrix. They estimated $\mathbf{f}$ by computing

$$\mathbf{f} = \min_{\mathbf{f}} \sum_{i=1}^{n} (w_{S_i} \mathbf{f}^\top \mathbf{z}_i)^2$$

where w_{S_i} is the optimal weight, the variance of the residual. Dropping subscripts i, and following Sampson and Weng *et al.*, the optimal weighting is

$$w_S = \frac{1}{\nabla r}$$

where the gradient is

$$\nabla r = (r_x^2 + r_y^2 + r_{x'}^2 + r_{y'}^2)^{1/2}$$

and where the partial derivatives r_x, and so on, are found from Equation (2) as $r_x = f_1 x' + f_4 y' + f_7 \zeta$, and so on. This is a first order approximation to the standard deviation of the residual.

Because calculation of the weights requires a value for the fundamental matrix, and vice versa, an iterative method is called for. We have modified the method proposed for conics by Sampson (1982), exploiting the fact that the fundamental matrix defines a quadratic in the image coordinates. The method computes an algebraic fit to $\mathbf{f}$ by an eigenvalue method, then re-weights the algebraic distance from each sample point $\{\mathbf{x}, \mathbf{x}'\}$ by $1/\nabla r^-(\mathbf{x}, \mathbf{x}')$, where ∇r^- is the gradient computed at the previous iteration, using unit weights on the first iteration.

The fundamental matrix should have zero determinant, but in the presence of noise this constraint has to be imposed on the minimization. If it were not, the epipolar lines would not all intersect at a unique epipole. To force the estimated fundamental matrix to be rank 2, at each iteration $\mathbf{F}$ is replaced by the nearest rank 2 matrix *before* calculating the weights. The procrustean[3] approach adopted here proceeds as follows. Let the singular value decomposition (Golub & van Loan 1989) of the recovered $\mathbf{F}$ be

$$\mathbf{F} = \mathbf{V} \mathbf{\Lambda} \mathbf{U}^\top \ .$$

Due to noise $\mathbf{F}$ will have full rank with non zero singular values: $\mathbf{\Lambda} = \mathrm{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \sqrt{\lambda_3})$. To approximate $\mathbf{F}$ by a rank two matrix, let $\mathbf{\Lambda}^+ = \mathrm{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, 0)$ whence the reduced rank approximation of the fundamental matrix is

$$\mathbf{F} = \mathbf{V} \mathbf{\Lambda}^+ \mathbf{U}^\top \ .$$

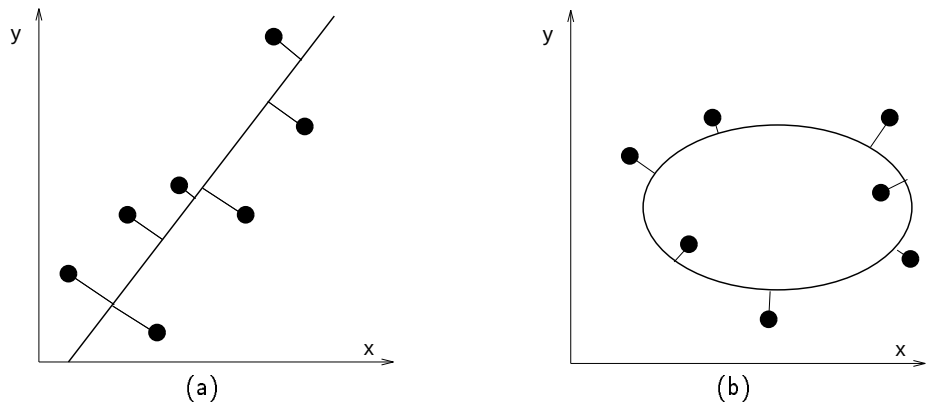

Fig. 3. For both line and conic fitting minimizing the perpendicular geometric distances is optimal. For the line this is equivalent to minimizing the algebraic residuals, but for the conic, as for the quadric surface of **F**, it is not.

The optimal weights convert the algebraic distance of each point into the statistical distance in noise space, which is equivalent to the first order approximation of the geometric distance, as shown by Sampson (1982) and Pratt (1987). The weighting breaks down at the epipole, as the numerator and the denominator both approach zero, indicating that there is less information about correspondences the closer they are to the epipole. In practice to remove unstable constraints all points within a pixel of the estimated epipole are excluded from that iteration of the calculation. We note that Kanatani (1996) recently proposed a modification to Sampson's distance for estimating the essential matrix, the calibrated analogue to **F**. We have found both distances yield almost identical results for the fundamental matrix on our test data, and we continue to use Sampson's distance.

Method S2. Luong and Faugeras (1993) examined Sampson's weighting, $w_S = 1/\nabla r$, and suggested that marginally better results could be obtained by using the distance of a point to its epipolar line as the error to be minimized. As already noted, the fundamental matrix **F** defines the epipolar geometry, and any point corresponding to $\underline{\mathbf{x}}$ in image one must lie on the epipolar line $\mathbf{F}\underline{\mathbf{x}}$ in image two. Noisy measurements will however not lie on their associated epipolar lines exactly. The perpendicular distance of a point (x, y) to the epipolar line in the first image is

$$e_1 = \frac{x r_x + y r_y + r_\zeta}{(r_x^2 + r_y^2)^{1/2}} = \frac{r}{(r_x^2 + r_y^2)^{1/2}}.$$

and the distance of a point (x', y') to the epipolar line in the second image is

$$e_2 = \frac{r}{(r_{x'}^2 + r_{y'}^2)^{1/2}}.$$

In order to minimize the geometric distance of each point to its epipolar line within an image the root mean square over the two image planes of each point to its epipolar line is used. This ensures that each image receives equal consideration. The distance $e = (e_1^2 + e_2^2)^{1/2}$ is referred to as the epipolar distance. It is assumed that the errors in the measured location of each point are Gaussian with variance σ^2. If the coordinate system were rotated so that one axis aligned with the epipolar line, the distance of each point to its (true) epipolar line has Gaussian distribution with variance σ^2 equal to the variance on the image point locations, hence e^2 approximately follows a χ^2 distribution with two degrees of freedom.

As in S1, Method S2 uses iterative re-weighted least squares to estimate the solution, and again **F** must be forced to be rank 2. The quantity minimized is equivalent to weighting the algebraic distance r by w_E, so that $e = r w_E$. Effectively then the epipolar weighting is

$$w_E = \left(\frac{1}{r_x^2 + r_y^2} + \frac{1}{r_{x'}^2 + r_{y'}^2} \right)^{1/2},$$

which is not dissimilar to the Sampson/Weng weighting

$$w_S = \left(\frac{1}{r_x^2 + r_y^2 + r_{x'}^2 + r_{y'}^2} \right)^{1/2} .$$

In summary, three least squares methods with different error terms have been discussed:

1. Method OR uses the sum of squares of *algebraic distances*:
$$R^2 = \sum r_i^2, \text{ where } r_i = \mathbf{x}_i'^{\top} \mathbf{F} \mathbf{x}_i.$$
2. Method S1 uses the sum of squares of the Sampson/Weng distances:
$$D^2 = \sum (w_{Si} r_i)^2 = \sum (r_i / \nabla r_i)^2 = \sum d_i^2.$$
3. Method S2 uses the sum of squares of the epipolar distances:
$$E^2 = \sum (w_{Ei} r_i)^2 = \sum (e_{1i}^2 + e_{2i}^2) = \sum e_i^2.$$

In terms of probability distributions, the first corresponds to something intractable, the second is a first order approximation to a χ^2 distribution, and the third is a χ^2 distribution.

We do not show all the comparisons made between least squares methods (Torr 1995) here. In summary, the error criteria D^2 and E^2 of methods S1 and S2 are indeed similar in performance and superior to OR, the more so when D^2 and E^2 were used as cost function in a non-linear gradient descent minimization, such as that of Gill and Murray (1978). More important is the result that even the best least-squares method performs feebly in the presence of outliers.

Given that both the mean square Sampson/Weng distance D^2 and the mean square epipolar distance E^2 are both popular measures accuracy of a solution, both measures will be evaluated in the comparison of the robust estimators. But which measure is more justifiable? Although the epipolar distance has the merit of being image-based and physically intuitive, the adoption of the Sampson/Weng distance has several things to recommend it.

First it represents the sum of squares of the algebraic residuals divided by their standard deviations, whereas the standard deviations of the epipolar distances are unknown. Secondly, Kendall & Stuart (1983) suggested that the set of parameters that minimize the orthogonal distance of each point to a curve/surface are the maximum likelihood solution. This distance turns out to be intractable, but D^2 provides a first order approximation to it.

Thirdly, as discussed more fully in (Torr 1996), $d = w_S r$ is the first order approximation of the distance of a correspondence in the 4D space defined by (x, y, x', y') to the manifold defined by $\mathbf{F}$ in that space, an approximation good to 4 or 5 significant figures. It is also shown there that d is a similarly good approximation of the distance of point to its optimally estimated correspondence as given by Hartley and Sturm (1994) .

Fourthly, the value $D^2/(n-7)$ provides a maximum likelihood estimate of the variance of the error on each coordinate. Experiment has confirmed this. If the data are perturbed by noise $\sigma = 1.0$, the estimate of σ provided is near 1. Given the ground truth, which is known in our tests using synthetic data, the r.m.s. distance D of the *noise-free* points $\underline{\mathbf{x}}$ from $\mathbf{F}$ estimated using the *noise-corrupted* points $\mathbf{x}$ tends to zero as the fit improves.

3. The method for comparing robust estimators

Although considerable work appears in the statistical literature on the detection of outliers in the context of ordinary, non-orthogonal, regression (see Chaterjee & Hadi (1988) for a review), little work has been done on outlier detection in orthogonal regression — the work of Shapiro and Brady (1993) on hyperplane fitting appears an exception both in the statistical and computer vision literature. Moreover, it appears that no large scale comparative studies have been reported on the robust estimation of general hyper-surfaces, into which category the estimation of the fundamental matrix falls now that we are to minimize the geometric rather than algebraic distance.

We have therefore developed and evaluated a number of robust methods to this last problem. Our evaluation places emphasis on two performance criteria: (i) relative efficiency and (ii) breakdown point, defined as follows.

(i) The *relative efficiency* of a regression method is defined as the ratio between the lowest achievable variance for the estimated parameters (the Cramér-Rao bound (Kendall & Stuart

1983)) and the actual variance provided by the given method. An empirical measure of this is achieved by calculating the distance (either d_i or e_i) of the *actual* noise free projections of the synthetic world points to $\mathbf{F}$ provided by each estimator. Traditionally the goodness of fit has been assessed by seeing how well the parameters fit the *observed* data. But we point out that this is the wrong criterion as the aim is to find the set of parameters that best fit the (unknown) *true data*. The parameters of the fundamental matrix themselves are not of primary importance, rather it is the structure of the corresponding epipolar geometry. Consequently it makes little sense to compare two solutions by directly comparing corresponding parameters in their fundamental matrices; one must rather compare the the difference in the associated epipolar geometry weighted by the density of the given matching points. The inadequacy of using the fit to the *observed* data to assess efficiency, in the presence of outliers, will be demonstrated in the results section.

(ii) The *breakdown point* of an estimator is the smallest proportion of outliers that may force the value of the estimate outside an arbitrary range. For a normal least squares estimator one outlier is sufficient to arbitrarily alter the result, therefore it has a breakdown point of $1/n$ where n is the number of points in the set. An indication of the breakdown point is gained by conducting the tests with varying proportions of outliers.

The overall plan for comparison involves two levels — an initial simple test to weed out methods that are completely ineffective, followed by a more detailed testing of the remaining methods, involving evaluation on a range of real and synthetic data. We return to characterize the tests below, but first describe the generation of synthetic data.

Data $\underline{\mathbf{X}}$ are randomly generated in the region of $\mathcal{R}^3$ visible to two positions of a synthetic camera having intrinsic coordinates

$$\mathbf{C} = \begin{bmatrix} 1.00 & 0.00 & 0.36 \\ 0.00 & 1.50 & 0.36 \\ 0.00 & 0.00 & 0.0014 \end{bmatrix},$$

equivalent to an aspect ratio of 1.5, an optic centre at the image centre $(256, 256)$, and a focal length of $f = 703$ (notionally pixels), giving a field of view of $40°$, and giving $0 \leq x, y \leq 512$. These values were chosen to be similar to the camera used for capturing real imagery. The projection of a point $\underline{\mathbf{X}}$ in the first position is $\underline{\mathbf{x}} = \mathbf{C}[\mathbf{I}|\mathbf{0}]\underline{\mathbf{X}}$ and in the second is $\underline{\mathbf{x}}' = \mathbf{C}[\mathbf{R}|\mathbf{t}]\underline{\mathbf{X}}$ where the camera makes a rotation $[\mathbf{R}]$ and translation $\mathbf{t}$. The motion is random and different in each test. In order to simulate the effects of the search window commonly employed in feature matchers, and to limit the range of depths in 3D, correspondences were accepted only if the disparity lay between $4 \leq \delta \leq 30$ pixels. (Some notion of the limits depth Z can be obtained for pure translation as $|\mathbf{t}|f/\delta_{\max} \leq Z \leq |\mathbf{t}|f/\delta_{\min}$.)

In Figure 4 we show a typical set of point correspondences as image motion vectors arising from some arbitrary random motion. The blob end is the position (x, y) and the other end is at (x', y'). Overlaid are the epipolar lines computed in image one using the motion, camera intrinsics and positions (x', y').

The initial weeding was achieved by testing on 10 sets of 200 synthetic correspondences. In each set, 180 point correspondences were generated in accordance with the synthetic camera motion and an additional 20 correspondences were outliers. Each image point $\underline{\mathbf{x}}$ was perturbed to $\mathbf{x}$ by Gaussian noise with standard deviation 1.0. The standard deviation of the *actual* noise free projections of the synthetic world points to the estimated epipolar lines were calculated. If the standard deviation exceeded 4.0 (four times the noise on the point positions), that method was rejected outright.

Those methods that passed this initial test were then tested more exhaustively using increasing outlier contamination, up to 50% in steps of 5%. The outliers (or mismatches) were generated so as to be in a random direction on the image plane between minimum and maximum allowable disparity in pixels from their position in image one. Each experiment, with a different percentage of outliers, was repeated on 100 different data sets, each of size 200, giving $20,000$ correspondences for each proportion of outliers, and $200,000$ correspondences over all.

As mentioned at the end of §2, to assess the performance of a method, the variance of the

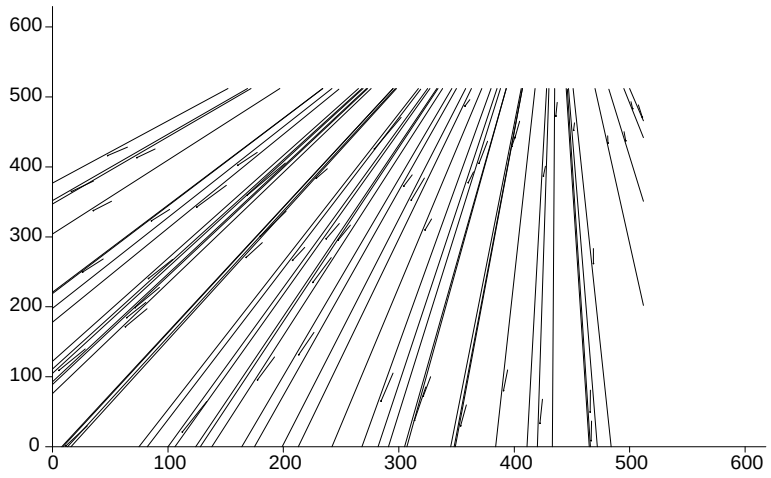

Fig. 4. A set of synthetically generated correspondences perturbed by noise; superimposed is the *true* epipolar geometry of the synthetic camera pair.

weighted distances (d_i, e_i etc) from the noise-free (ie ground truth) projections of the synthetic world points was derived as a function of the fraction of outliers.

4. Category I: M-estimators

We now turn to the first category of robust estimators, that of *M-estimators*. Our description will be for the Sampson/Weng distance measure — to use the epipolar distance measure, D and d are replaced by the E and e derived earlier.

Given a set of correspondences $\mathbf{x}_i \leftrightarrow \mathbf{x}'_i$, suppose we wish to find $\mathbf{F}$ which has maximum likelihood given the data. If there is no preferred *a priori* value of $\mathbf{F}$ this is equivalent to maximizing $\Pr(D|\mathbf{F})$. This joint probability is identical to the noise distribution, which assuming the noise is Gaussian, has zero mean, is independent at each datum, and has the same standard deviation σ, is

$$\Pr(D|\mathbf{F}) = \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left(-D^2/2\sigma^2\right).$$

Maximizing this is equivalent to minimizing the negative of its logarithm

$$\frac{D^2}{2\sigma^2} + \text{constant terms} ,$$

which of course is the proof that least squares is the maximum likelihood estimator *when the errors are Gaussian.*

Under real conditions this Gaussian assumption is rather poor. The aim of M-estimators (Maronna 1976; Huber 1981; Hampel *et al.* 1986) is to follow maximum-likelihood formulations by deriving optimal weighting for the data under non-Gaussian conditions. Outlying observations have their weights reduced rather than being rejected outright. The estimators minimize the sum of a symmetric, positive-definite function $\rho(d_i)$ of the errors d_i, with a unique minimum at $d_i = 0$. That is, the parameters are sought that minimize

$$\sum_i^n \rho(d_i) \quad .$$

The form of ρ is derived from the particular chosen density function in the manner shown for the case of Gaussian errors. Usually the density function is chosen so that ρ is some weighting, $\rho(d_i) = (\gamma_i d_i)^2$, of the squared error that reduces the effects of outliers on the estimated parameters. A typical weighting scheme in the statistics literature is that proposed by Huber (1981):

$$\gamma_i = \begin{cases} 1 & d_i < \sigma \\ \sigma/|d_i| & \sigma < d_i < 3\sigma \\ 0 & d_i > 3\sigma. \end{cases} \qquad (3)$$

The standard deviation of the error (scale) σ is either known *a priori* or is found as a maximum likelihood estimate using the median

$$\sigma = \frac{\text{med}_i d_i}{0.6745} \quad . \qquad (4)$$

Further discussion of the estimation of σ is deferred until §8.

Within computer vision M-estimation has been used by Luong (1992), Olsen (1992), and Zhang *et al* (1994) for the estimation of epipolar geometry. Within the statistics literature most of the analytical work on M-estimators has been applied to ordinary least squares. The only work done for orthogonal regression has been in the area of Principal Component Analysis where Devlin *et al.* (1981) conducted a comparison of a number of robust techniques for estimating the principal components of data (ie, the eigenvectors). Of the several weightings explored, they concluded that Huber's weighting and a weighting due to Maronna (1976)

$$\gamma_i = \frac{1 + d_i}{1 + d_i^2}$$

gave the best estimates of the principal components. For large and small d_i this tends to weighting by the inverse residual and to unity, respectively.

Numerical calculation of M-estimators is problematic at best and, so far, no closed form solutions exist — indeed the problem appears intractable. Note that the weights cannot be computed without an estimate of the residuals, which in turn requires knowledge of the solution. Huber (1981) suggests an iterative computational scheme in which the weights are held constant at values equal to those found at the last iteration, whilst the set of parameters are estimated. Huber proves that if these iterations are repeated a local (possibly global) minimum of the objective function (4) is reached. The algorithm, presented in the Appendix, is a modification of the iterative least squares method of Sampson (S1), combining the Sampson/Weng distance weighting w_S with the robust weighting of the M-estimators. An initial solution is obtained using orthogonal regression (Method OR).

Figure 5 gives the results of test using synthetic data on two M-estimators, using the weightings of Huber and Maronna, and assuming σ is known (i.e. the best possible case). Figure 5 (a) shows the results when the sum of squares of Sampson's distance D is minimized and the graph shows the variance of Sampson's distance for the *true or noise free* correspondences to the *estimated* epipolar geometry. Although Huber wins, both M-estimators are similar and typically poor for more than $20 - 25\%$ outliers in the data set with variance on the error term in excess of 4.0 (recall the initial Gaussian noise used has variance 4.0). Figure 5 (b) shows the results when the distance E is minimized, showing the distance of the *true* points to their predicted epipolar lines obtained from the *estimated* $\mathbf{F}$. It can be seen that Huber provides a more graceful degradation with outliers than Maronna which totally fails beyond 20% outliers. This is because M-estimators are highly vulnerable to poor starting conditions — the algorithm converges to a local minimum. Unfortunately, the linear least squares estimator used for initialization will almost certainly produce poor starting conditions in the presence gross outliers! The observed variance of the estimate is very high when compared with what we would expect given that the added Gaussian noise has a variance of $\sigma^2 = 4.0$. In fact many of the trials produced a much lower standard deviation, but the overall standard deviation of the error was inflated by a small proportion of estimates that totally failed to converge.

Some other weighting functions have been explored, notably the biweight function of Mosteller and Tukey (1997):

$$\rho(d_i) = \begin{cases} d_i^2/2 - d_i^4/2a^2 + d_i^6/6a^4 & \text{if } d_i < a \\ a^2/6 & \text{otherwise} \end{cases}$$

with $a = 1.96\sigma$, as used by Kumar and Hanson (1994). This error function has the advantage of performing locally like a Gaussian for small errors, and tapering off to a constant for large errors, thus limiting their effect. Our tests failed to reveal any significant improvement for the use of this error function, and indeed in most cases Huber performed marginally better.

Two other implementations were tried, both producing poorer results. First, M-estimator routines for *ordinary* least squares (with the error presumed in one variable) were taken from the Numerical Algorithm Group's (1988) library. The performance of these was poorer than Huber based on orthogonal regression. Second, gradient descent methods were tried with the Huber error function, using OR as the staring point. The

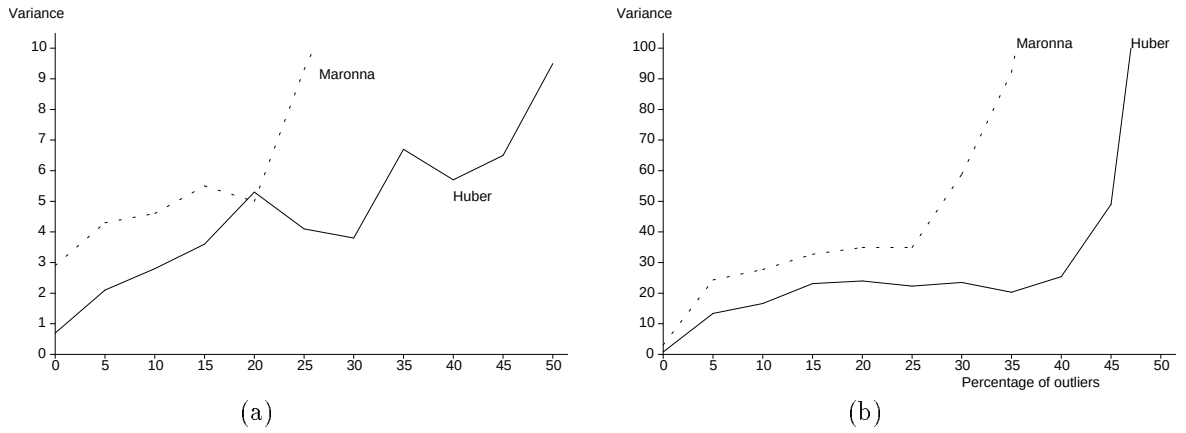

Fig. 5. Variance of the distance measures (measured from projections of the noise free points) as a function of the percentage of outliers for the M-estimators of Huber and Maronna. (a) the distance measure is Sampson's d (b) it is the distance to predicted epipolar line e. Each is derived from 100 trials each with 200 data.

frequency with which the algorithm became entrapped within local minima led to this approach being rejected outright using a non-robust initialization, but a robust initialization provided exceptionally good results, as will be seen.

In conclusion, it seems that the computation of M-estimators is highly intractable involving the solution of n non-linear equations, where n is the number of correspondences. Even the Huber's suggested approach to computing them, iterated least squares, is only suitable if there is *a priori* knowledge of the parameters or if there are a few gross outliers which are easily identified.

5. Category II: Case Deletion Diagnostics

This section describes the methods based on *influence measures*, particularly *case deletion diagnostics* (Chaterjee & Hadi 1988). The basic concept underlying influence is simple. Small perturbations are introduced into some aspect of the problem formulation and an assessment made of how much these change the outcome of the analysis. The important issues are, first, the determination of the type of perturbation scheme; secondly, the particular aspect of the analysis to monitor; and, thirdly, the method of assessment.

Case deletion methods in particular monitor the effect on the analysis of removing data. For instance, we might ask how the epipolar geometry would change given the deletion of one of the correspondences. Several different measures of influence have been proposed within the statistical literature for case deletion. They differ in the particular regression result on which the effect of the deletion is measured, and the standardization used to make them comparable over observations. For the methods discussed below, the influence can be computed within the regression process, and are inexpensive relative to the cost of the regression itself.

In the case of ordinary least squares the interested reader is referred to Chaterjee & Hadi (1988) for a rigorous analytical coverage of the theory and methods. Much less attention has been given to orthogonal regression. Critchley (1985) suggested the use of eigen-perturbation to arrive at influence functions assessing the first or higher order effect on the principal eigenvalues and eigenvectors. Torr and Murray (1993b) extended Cook's D to orthogonal regression, and this is discussed in more detail below. Shapiro and Brady (1993) proposed an influence measure that monitors the effects of the deletion on the minimum eigenvalue. Torr and Murray (1992, 1993a) gave examples of the use of a variety of such diagnostics for the estimation of affine instantaneous flow.

In the next section the case deletion diagnostic is developed for orthogonal regression, and then it is adapted to estimate the fundamental matrix.

5.1. Extending Cook's D to the case of orthogonal regression

We now derive a formula for the influence of a point on orthogonal regression is derived, extending the works of Cook and Weisberg (1980), Critchley (1985) and Shapiro and Brady (1993). An early version was presented in (Torr & Murray 1993b).

Consider the set of points lying on a hyperplane $\underline{\mathbf{f}}$ and let their measured values be $\mathbf{z}_i$, $i = 1 \ldots n$, after perturbation by Gaussian noise with uniform[4] standard deviation σ. If $\mathbf{Z}$ is the matrix whose rows are $\mathbf{z}_i^\top$ then the least squares estimate $\mathbf{f}$ of $\underline{\mathbf{f}}$ is given by the eigenvector of the moment matrix $\mathbf{M} = \mathbf{Z}^\top\mathbf{Z}$ corresponding to the minimum eigenvalue. By analogy with Cook's D (Cook & Weisberg 1980) which was developed for ordinary least squares, we monitor the effect of deleting an observation on the estimated parameters $\mathbf{f}$. As exact solutions cannot be found in closed form, the change in the solution is calculated using eigenvector perturbation theory (Torr 1995, Golub & van Loan 1989). (As noted above, examination of the effects of perturbation on some other aspect of the model could be made, e.g. the covariance matrices or the minimum eigenvalue (Shapiro & Brady 1993), but as our interest is in $\mathbf{f}$ it appears best to test eigenvector perturbations directly.)

Let $\mathbf{M}$ be the p-dimensional symmetric moment matrix, having eigenvalues, in increasing order, $\lambda_1 \ldots \lambda_p$ with $\mathbf{u}_1 \ldots \mathbf{u}_p$ the corresponding eigenvectors forming an orthonormal system. If matrix $\mathbf{M}$ is perturbed into

$$\mathbf{M}' = \mathbf{M} + \delta\mathbf{M},$$

and if the multiplicity of λ_j is 1, i.e. the data is not degenerate, then the eigenvector $\mathbf{u}_j$ is perturbed into (Golub & van Loan 1989)

$$\mathbf{u}_j' = \mathbf{u}_j + \sum_{k \neq j} \frac{\mathbf{u}_k^\top \delta\mathbf{M}\mathbf{u}_j}{\lambda_j - \lambda_k}\mathbf{u}_k + O(\delta\mathbf{M})^2.$$

In this case the deletion of the ith observation means that $\delta\mathbf{M} = -\mathbf{z}_i\mathbf{z}_i^\top$. This allows calculation of what would have been the estimate of the parameters with the ith element had been excluded

$$\mathbf{f}_{(i)} = -\mathbf{z}_i^\top\mathbf{u}_1 \sum_{k \neq 1} \frac{\mathbf{u}_k^\top\mathbf{z}_i}{\lambda_1 - \lambda_k}\mathbf{u}_k + \mathbf{u}_1, \qquad (5)$$

where $\mathbf{f}_{(i)}$ is the estimate of $\mathbf{f}$ with the ith element deleted. As $\mathbf{f} = \mathbf{u}_1$ then

$$\mathbf{f}_{(i)} - \mathbf{f} = -\mathbf{z}_i^\top\mathbf{u}_1 \sum_{k \neq 1} \frac{\mathbf{u}_k^\top\mathbf{z}_i}{\lambda_1 - \lambda_k}\mathbf{u}_k. \qquad (6)$$

In §5.4 some comments are made on how to improve the estimate of $\mathbf{f}_{(i)}$, but the above form is used for the analysis below as it provides some intuition into the nature of how the outliers effect the solution.

To be most useful, influence should be a scalar quantity. It is therefore necessary to use a norm to characterize influence; this norm will map the p vector, $\mathbf{f}_{(i)} - \mathbf{f}$, to a scalar. The norm is defined in terms of a symmetric, positive definite $p \times p$ matrix $\mathbf{L}$, to give an influence measure

$$T_i(\mathbf{L}) \stackrel{\text{def}}{=} \left(\mathbf{f}_{(i)} - \mathbf{f}\right)^\top \mathbf{L} \left(\mathbf{f}_{(i)} - \mathbf{f}\right).$$

Using Equation (6) this is

$$T_i(\mathbf{L}) = \left(\mathbf{z}_i^\top\mathbf{u}_1 \sum_{k \neq 1} \frac{\mathbf{u}_k^\top\mathbf{z}_i}{\lambda_1 - \lambda_k}\mathbf{u}_k\right)^\top \times \mathbf{L} \left(\mathbf{z}_i^\top\mathbf{u}_1 \sum_{l \neq 1} \frac{\mathbf{u}_l^\top\mathbf{z}_i}{\lambda_1 - \lambda_l}\mathbf{u}_l\right)$$

and noting that $\mathbf{z}_i^\top\mathbf{u}_1 = r_i$, the residual for the ith observation,

$$T_i(\mathbf{L}) = r_i^2 \left(\sum_{k \neq 1} \frac{\mathbf{u}_k^\top\mathbf{z}_i}{\lambda_1 - \lambda_k}\mathbf{u}_k\right)^\top \times \mathbf{L} \left(\sum_{l \neq 1} \frac{\mathbf{u}_l^\top\mathbf{z}_i}{\lambda_1 - \lambda_l}\mathbf{u}_l\right). \qquad (7)$$

Contours of constant $T_i(\mathbf{L})$ are ellipsoids of dimension equal to the rank of $\mathbf{L}$, centred at $\mathbf{f}$ or equivalently at $\mathbf{f}_{(i)}$. Clearly the character of $T_i(\mathbf{L})$ is determined by $\mathbf{L}$, which may be chosen to reflect specific concerns. The norm is chosen to make T_i both scaleless and invariant to non-singular linear transformations of the data.

Now suppose the matrix $\mathbf{M}$ is used for $\mathbf{L}$. In (Torr 1995) it is shown that $\mathbf{M}\sigma^{-2}$ is approximated by the pseudo inverse of $\mathbf{\Gamma}_f$, the covariance matrix of the parameter estimate:an improved estimate is given in §5.4. Choosing the covariance matrix as the norm in which to measure change in the solution allows the alterations in each element of the parameter vector to be given equal weight, so that changes in the parameters approximate the changes in the error measure D^2. Furthermore, when $\mathbf{L} \leftarrow \mathbf{M}$, $T_i(\mathbf{L})$ defines a conic in parameter space with principal axes determined by the eigenvalues and eigenvectors of $\mathbf{M}$. That is, if $\mathbf{f}_{(i)}$ is $\mathbf{f}$ computed without $\mathbf{z}_i$, then from Equation (5)

$$\begin{aligned} T_i(\mathbf{M}) &= \mathbf{f}_{(i)}^\top \mathbf{M}\mathbf{f}_{(i)} - 2\mathbf{f}^\top \mathbf{M}\mathbf{f}_{(i)} + \mathbf{f}^\top \mathbf{M}\mathbf{f} \\ &= \mathbf{f}_{(i)}^\top \mathbf{M}\mathbf{f}_{(i)} - 2\lambda_1 \mathbf{f}^\top \mathbf{f}_{(i)} + \lambda_1. \end{aligned}$$

Equation (7) leads to

$$T_i(\mathbf{M}) = \frac{1}{\sigma^2} \left(\mathbf{z}_i^\top \mathbf{u}_1 \sum_{k\neq 1} \frac{\mathbf{u}_k^\top \mathbf{z}_i}{\lambda_1 - \lambda_k} \mathbf{u}_k \right)^\top \times \mathbf{M} \left(\mathbf{z}_i^\top \mathbf{u}_1 \sum_{l\neq 1} \frac{\mathbf{u}_l^\top \mathbf{z}_i}{\lambda_1 - \lambda_l} \mathbf{u}_l \right).$$

Noting that $\mathbf{u}_j^\top \mathbf{Z}^\top \mathbf{Z}\mathbf{u}_k = \lambda_k \mathbf{u}_j^\top \mathbf{u}_k = \lambda_k \delta_{jk}$, the influence measure becomes

$$\begin{aligned} T_i(\mathbf{M}) &= \frac{1}{\sigma^2} \left(\mathbf{u}_1^\top \mathbf{z}_i\right)^2 \sum_{k\neq 1} \left(\frac{\mathbf{u}_k^\top \mathbf{z}_i}{\lambda_1 - \lambda_k} \sqrt{\lambda_k} \right)^2 \\ &= \frac{r_i^2}{\sigma^2} \sum_{k\neq 1} \left(\frac{\mathbf{u}_k^\top \mathbf{z}_i}{\lambda_1 - \lambda_k} \sqrt{\lambda_k} \right)^2. \end{aligned} \tag{8}$$

The singular value decomposition (Golub & van Loan 1989; Thisted 1988; Teukolsky *et al.* 1988) is used to derive a computationally simple form for Equation (8). Let the singular value decomposition of $\mathbf{Z}$ be

$$\mathbf{Z} = \mathbf{V}\mathbf{\Lambda}\mathbf{U}^\top,$$

where $\mathbf{V}$ is a $n \times p$ matrix whose columns are the left hand singular vectors of $\mathbf{Z}$, $\mathbf{U}$ is a $p \times p$ matrix whose columns are the right hand singular vectors of $\mathbf{Z}$ and $\mathbf{\Lambda}$ is the diagonal matrix of the corresponding singular values of $\mathbf{Z}$: $\mathbf{\Lambda} = \mathrm{diag}\left(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \ldots, \sqrt{\lambda_p}\right)$ in ascending order such that $\sqrt{\lambda_1}$ is the smallest singular value. Then if $\mathbf{V}_{ik}$ is the ikth element of $\mathbf{V}$, it is easy to see that

$$\mathbf{u}_k^\top \mathbf{z}_i = \mathbf{V}_{ik} \sqrt{\lambda_k},$$

and so

$$T_i(\mathbf{M}) = \frac{r_i^2}{\sigma^2} \sum_{k\neq 1} \left(\frac{\mathbf{V}_{ik}\lambda_k}{\lambda_1 - \lambda_k} \right)^2.$$

The leverage factor is defined as

$$l_i \stackrel{\text{def}}{=} \sum_{k\neq 1} \left(\frac{\mathbf{V}_{ik}\lambda_k}{\lambda_1 - \lambda_k} \right)^2.$$

This leverage factor will be large only if the orthogonal projection of the observation $\mathbf{z}_i$ onto $\mathbf{u}_k$, $k \neq 1$, is large and the corresponding eigenvalue λ_k is small. The leverage gives a measure of the influence of each point and is large for outliers even when the residual is small. The leverage is a key factor distinguishing use of $T_i(\mathbf{M})$ from consideration of the algebraic residual r_i alone. For an outlier r_i might be small but $T_i(\mathbf{M})$ will tend to be large.

Note that the scale σ need not be known to calculate the relative values of T_i. Temporarily ignoring the scaling, the measure T_i has both a revealing and computationally efficient form

$$T_i = r_i^2 l_i$$

which is the residual multiplied by the leverage factor. The statistic $T_i(\mathbf{M})$ gives the relative influence of each point in the regression, and to remove outliers the point with maximal influence is deleted and the regression recomputed, repeating this procedure until the data falls below a χ^2 threshold determined by σ.

5.2. A worked example

Here, the case deletion outlier detection scheme is applied to the x, y data set of Fischler and Bolles (1981) given in the introduction.

The data is first centred, giving $\mathbf{u}_1 = (u_{11}, u_{12}) = (0.18, -0.98)$ and sum of squares $\lambda_1 = 8.19$. The singular value decomposition for the centred data matrix is

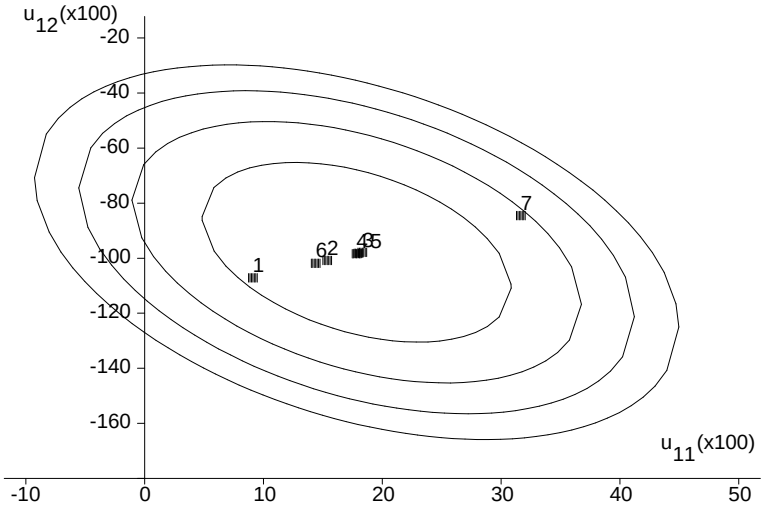

Fig. 6. Ellipses of constant T_i in parameter space which are concentric about $\mathbf{u}_1 = (0.178136, -0.984006)$. The ellipses are determined by the covariance matrix of the parameter estimate. The perturbed solutions that have arisen from the deletion of points are also plotted. It can be seen that points 1 and 7 lie furthest from the centre of the ellipses.

$$\mathbf{Z} = \mathbf{V}\mathbf{\Lambda}\mathbf{U}^{\top}$$

which on inserting values is

$$\begin{bmatrix} -3.28 & -2.00 \\ -2.28 & -1.00 \\ -1.28 & 0.00 \\ -0.28 & 0.00 \\ -0.28 & 1.00 \\ 0.71 & 2.00 \\ 6.71 & 0.00 \end{bmatrix} = \begin{bmatrix} 0.48 & -0.44 \\ 0.20 & -0.30 \\ -0.08 & -0.15 \\ -0.017 & -0.03 \\ -0.36 & -0.01 \\ -0.64 & 0.13 \\ 0.41 & 0.81 \end{bmatrix} \times \begin{bmatrix} 8.08 & 0.00 \\ 0.00 & 2.86 \end{bmatrix} \begin{bmatrix} 0.18 & -0.98 \\ 0.98 & 0.18 \end{bmatrix}.$$

Using Equation (5) with these values gives the the perturbed results $(u_{11}(i), u_{12}(i))$ which are given in Table 1 and plotted in Figure 6.

Also shown in Figure 6 are concentric ellipses corresponding to increasing values of the influence measure T_i determined from the covariance matrix of the parameter estimate. In parameter space ellipses of constant T_i are

$$63.43u_{11}^2 + 20u_{11}u_{12} + 10u_{12}^2 - 2.91u_{11} + 16.04u_{12} = T_i - 8.19$$

where the values (from inwards out) $T_i = 1, 2, 3,$ and 4. Figure 6 shows clearly that Point 7 is exerting undue influence on the fit.

In Table 2 we compare the T_i diagnostic with the naïve residual diagnostic and the diagnostic of Shapiro and Brady which measures the perturbation of the smallest eigenvalue. The generalized distance T_i correctly identifies Point 7 as outlying, whereas the algebraic residual r_i and the perturbation of the smallest eigenvalue $\delta\lambda_1(i)$ indicate Point 6 as most outlying.

Table 2 also shows the leverage

$$l_i = \left(\frac{\mathbf{V}_{i2}\lambda_2}{\lambda_1 - \lambda_2} \right)^2 ,$$

Table 1. The perturbed results arising from the deletion of each point estimated to a first order approximation using Equation 5.

Point i	$u_{11}(i)$	$u_{12}(i)$
1	0.092530	-0.999503
2	0.153986	-0.988378
3	0.183134	-0.983101
4	0.178383	-0.983961
5	0.179975	-0.983673
6	0.144509	-0.990094
7	0.314437	-0.959331

Table 2. The T_i measure correctly identifies point 7 as an outlier, but the algebraic residual r_i and the smallest eigenvalue perturbation $\delta\lambda_1(i)$ do not. Columns 4 and 5 show values of the leverages h_{zii} and l_i.

Point	r_i	$\delta\lambda_1(i)$	T_i	h_{zii}	l_i
1	1.38	-2.34	0.49	0.43	0.26
2	0.58	-0.37	0.039	0.13	0.12
3	-0.23	-0.054	0.0017	0.031	0.032
4	-0.051	-0.0026	0.0	0.0015	0.0016
5	-1.035	-1.071	0.00023	0.13	0.00021
6	-1.84	-3.45	0.076	0.43	0.022
7	1.20	-2.53	1.25	0.84	0.87

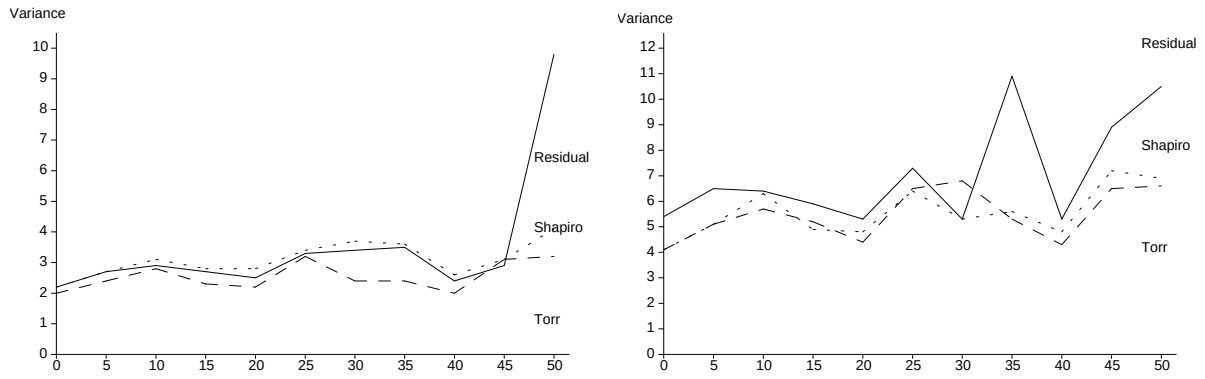

Fig. 7. Variance of the distance measures (measured from projections of the noise free points) as a function of the percentage of outliers for the case deletion methods. (a) the distance measure is Sampson's, (b) it is the distance to predicted epipolar line. Each is derived from 100 trials each with 200 data.

and that for Point 7 is large. (Note that the rather simpler expression for leverage $h_{zii} = \mathbf{V}_{i1}^2 + \mathbf{V}_{i2}^2$ which is sometimes used is not as discriminating.) It might be noted that while not indicating outliers directly, leverages do give a good indication of what points are influential in the regression.

To understand why the diagnostic based on the change in the eigenvalue fails in this case, note that the eigenvectors are much more sensitive to perturbations than the eigenvalues, especially when the eigenvalues might be quite close. This leads to the speculation that the diagnostic T_i might be good for detecting degeneracy within small data sets.

Shapiro and Brady (1993) overcame this problem by explicitly recomputing the regression for each point deleted when there are only a few outliers left, to determine which gives a minimal sum of squared residuals. This approach is prohibitively expensive for large datasets.

5.3. Application to computing the Fundamental matrix

The method above was derived under the assumption of linear regression. However the fundamental matrix gives rise to a quadratic in the image coordinates. In order to minimize the correct measure a modification to the ILS iterative least squares method described in Section 2.3 is used. At each iteration, as well as re-weighting all the data to convert the algebraic residuals into the correct statistical distance in noise space, the point with maximum influence is deleted. Furthermore the fundamental matrix is projected onto the nearest singular fundamental matrix using the singular value decomposition, as described in Section 2.3.

All the case deletion algorithms successively delete points until the sum of squares of residuals lies below a χ^2 test. An outline of the new algorithm is presented in the Appendix.

Figure 7 (a) (b) gives the average sum of squares of distance of actual points to estimated epipolar geometries for three influence measures — those of Torr, Shapiro and the "deleting the largest residual at each iteration", and the two error measures D E. Overall the performances are similar, with no significant statistical difference between the estimators. The data sets tested here are large with $n = 200$, on smaller data sets ($n \leq 40$) T_i was found to give better results than just deleting the largest residual, T_i also gave a better performance for linear data (e.g. when fitting hyperplanes).

Convergence of the case deletion diagnostics was superior to the convergence of the M-estimators, and the solution was typically more accurate. This is because whereas in the M-estimation process all the data are reweighted at each iteration, in the case deletion schemes only one datum is considered at each iteration. This leads to increased accuracy at the expense of more iterations (generally one per outlier). The disadvantage of the case deletion schemes is that they require a fairly good estimate of σ.

5.4. On improving the estimate of T_i

Although not used in the experiments reported in this paper, we have recently adopted three techniques that have been found to increase the accuracy of the estimation process. The first is an improvement in the estimate of $\mathbf{f}_{(i)}$ using iterative methods. The second improvement uses the calculated $\mathbf{f}_{(i)}$ to remove bias in the solution, and the third improvement uses the calculated $\mathbf{f}_{(i)}$ to re-estimate the covariance matrix.

Improving the estimate of $\mathbf{f}_{(i)}$. Although a closed form solution cannot be obtained for $\mathbf{f}_{(i)}$ the first order approximation can be improved upon. Golub (1973) provides a method for computing the eigenvalues and eigenvectors of a matrix $\mathbf{C} = \mathbf{D} + \mathbf{u}\mathbf{u}^\top$ in $O(n^2)$ operations, where $\mathbf{D}$ is diagonal. A more general analysis is presented but Gu & Eisenstat (1995), and the methods have been implemented and used for case deletion by Shapiro and Brady (1994), providing a marked improvement in the result.

Non-parametric Removal of Bias. Luong *et al.* (1993) show that linear methods produce a biased solution for the fundamental matrix, analogous to that in conic fitting found by Kanatani (1991). Kanatani (1996) gives the bias for linear estimation of the essential matrix. Kanatani follows a parametric approach for the removal of bias under the assumption of Gaussian noise.

Here however we suggest a non-parametric approach that should be more robust to outliers or the failure of Gaussian assumptions. The *jackknife* is a well-known and extensively studied statistical non-parametric technique to gain an unbiased estimate of $\mathbf{f}$ together with its covariance. If $\mathbf{f}_J$ is the jackknife estimate it can be shown that its bias decreases as a polynomial function of the number of observations n (Kendall & Stuart 1983). Using the original sample of n data, all n subsamples of $n-1$ data are formed, by systematically deleting each observation in turn. The jackknife is given by

$$\mathbf{f}_J = n\mathbf{f} - \frac{n-1}{n}\sum_{i=1}^{n}\mathbf{f}_{(i)} \ ,$$

the bias in each parameter being (Sprent 1989)

$$\left(1 - \frac{1}{n}\right)\sum_{i=1}^{n}(\mathbf{f}_{(i)} - \mathbf{f}) \ .$$

It can be seen that there is only a small amount of extra computation necessary to get the bias-free estimates, as the quantities $\mathbf{f}_{(i)}$ have already been calculated. One problem is that the removal of bias might sometimes increase the error, whereas the biased solution has a lower error. A full analysis of bias removal is complicated and beyond the scope of this paper. Our tests found on average about 1% reduction in the error of fit to the true points, dependent on the type of motion (the nearer the image to orthographic conditions the less bias there is to remove, due to the fact that the problem becomes linear in the orthographic case). This average hides the fact that some correspondences (such as those near high curvature points of the fundamental matrix consider as a quadric manifold in the 4-space of the image coordinates) have much greater bias than others.

Non-parametric Estimation of the Covariance. Our further experimentation has revealed that the estimation of the covariance by the pseudo-inverse of the moment matrix to be a poor one. An improved estimate (really beneficial only for large values of n) may be gained at little extra computational cost by the jackknife estimate of the covariance matrix

$$\mathbf{\Gamma}_J = \frac{1}{n-7}\sum_{i=1}^{n}(\mathbf{f}_{(i)} - \mathbf{f})(\mathbf{f}_{(i)} - \mathbf{f})^\top \ .$$

6. Category III: Random Sampling Algorithms

An early example of a robust algorithm is the random sample consensus paradigm (RANSAC) of Fischler and Bolles (1981). Given that a large proportion the data may be outlying, the approach is the opposite to conventional smoothing techniques. Rather than using as much data as is possible to obtain an initial solution and then attempting to identify outliers, as small a subset of the data as is feasible to estimate the parameters is used (e.g. two point subsets for a line, seven

correspondences for a fundamental matrix), and this process is repeated enough times on different subsets to ensure that there is a 95% chance that one of the subsets will contain only good data points. The best solution is that which maximizes the number of points whose residual is below a threshold. Once outliers are removed the set of points identified as non-outliers may be combined to give a final solution.

An initial exploration of the RANSAC method to estimate the epipolar geometry was reported in Torr and Murray (1993b). To estimate the fundamental matrix seven points are selected to form the data matrix $\mathbf{Z}$:

$$\mathbf{Z} = \mathbf{W} \begin{bmatrix} x'_1x_1 & x'_1y_1 & x'_1\zeta & y'_1x_1 & & & \\ \vdots & \vdots & \vdots & \vdots & & & \\ x'_7x_7 & x'_7y_7 & x'_7\zeta & y'_7x_7 & & & \\ & & & & y'_1y_1 & y'_1\zeta & x_1\zeta & y_1\zeta & \zeta^2 \\ & & & & \vdots & \vdots & \vdots & \vdots & \vdots \\ & & & & y'_7y_7 & y'_7\zeta & x_7\zeta & y_7\zeta & \zeta^2 \end{bmatrix}$$

The null space of the moment matrix $\mathbf{M} = \mathbf{Z}^\top\mathbf{Z}$ is dimension two, barring degeneracy ($\mathbf{Z}$ is 7×9). It defines a one parameter family of exact fits to the 7 correspondences: $\alpha\mathbf{F}_1 + (1-\alpha)\mathbf{F}_2$. Introducing the constraint $\det|\mathbf{F}| = 0$ leads to a cubic in α:

$$\det|\alpha\mathbf{F}_1 + (1-\alpha)\mathbf{F}_2| = 0 \tag{9}$$

which has 1 or 3 real solutions for α. The total number of consistent features for each solution is recorded.

In order to determine whether or not a feature pair is consistent with a given fundamental matrix, the Sampson/Weng distance for each correspondence in the image is compared to a threshold, which will be described later in Section 8.

Table 3. The number m of subsamples required to ensure $\Upsilon \geq 0.95$ for given p and ϵ, where Υ is the probability that all the data points selected in one subsample are non-outliers.

Features	Fraction of Contaminated Data, ϵ						
p	5%	10%	20%	25%	30%	40%	50%
2	2	2	3	4	5	7	11
3	2	3	5	6	8	13	23
4	2	3	6	8	11	22	47
5	3	4	8	12	17	38	95
6	3	4	10	16	24	63	191
7	3	5	13	21	35	106	382
8	3	6	17	29	51	177	766

Ideally every possible subsample of the data would be considered, but this is usually computationally infeasible, so an important question is how many subsample of the dataset is required for statistical significance. Fischler and Bolles (1981) and Rousseeuw (1987) proposed slightly different means of calculation, but both give broadly similar numbers. Here we follow the latter's approach. The number m of samples is chosen sufficiently high to give a probability Υ in excess of 95% that a good subsample is selected. The expression for this probability Υ is

$$\Upsilon = 1 - (1 - (1-\epsilon)^p)^m, \tag{10}$$

where ϵ is the fraction of contaminated data, and p the number of features in each sample. Table 3 gives some sample values of the number m of subsamples required to ensure $\Upsilon \geq 0.95$ for given p and ϵ. Generally it is better to take more samples than are needed as some samples might be degenerate. It can be seen from this that, far from being computationally prohibitive, the robust algorithm may require fewer repetitions than there are outliers, as it is not directly linked to the number but only the proportion of outliers. It can also be seen that the smaller the data set needed to instantiate a model, the fewer samples are required for a given level of confidence. If the fraction of data that is contaminated is unknown, as is usual, an educated worst case estimate of the level of contamination must be made in order to determine the number of samples to be taken, this can be updated as larger consistent sets are found e.g. if the worst guess is 50% and a set with 80% inliers is discovered, then ϵ could be reduced from 50% to 20%.

In general if the seven correspondence sample has an insufficient spread of disparities then the estimate of $\mathbf{F}$ obtained from that sample might not be unique. This is an example of degeneracy. Consider the seven correspondences shown in Figure 8(a). Two epipolar geometries that fit this data are shown, for one view, in (b) and (c). The veridical epipolar geometry is (b) and (c) is erroneous solution nonetheless consistent with the cubic in Equation (9).

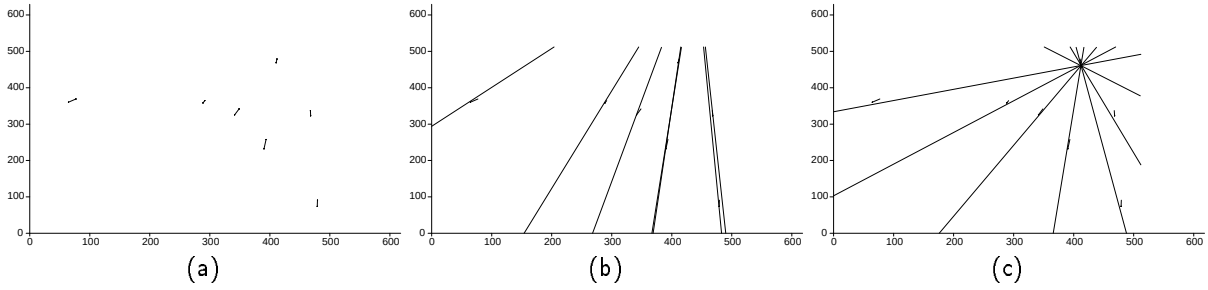

Fig. 8. (a) A typical set of seven points selected during RANSAC. alongside two epipolar geometries that exactly fit the data. (b) is the true epipolar geometry, and (c) is spurious.

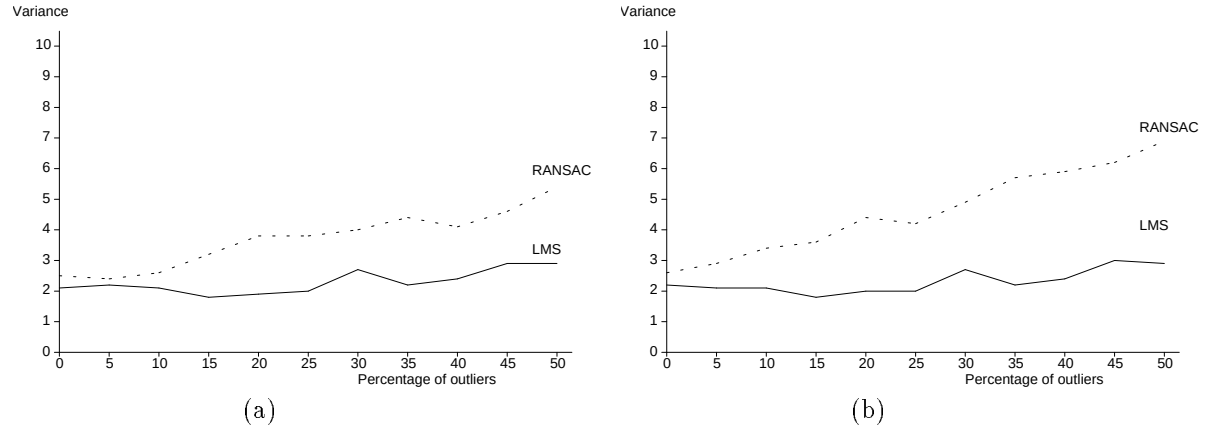

Fig. 9. Variance of the distance measures (measured from projections of the noise free points) as a function of the percentage of outliers for the RANSAC and LMS random sampling method. (a) the distance measure is Sampson's, (b) it is the distance to predicted epipolar line. Each is derived from 100 trials each with 200 data. The results show that generally LMS gives a slightly better result.

Clearly the result estimated from this sample will not have many other consistent correspondences that conform to the underlying motion. It is desirable to devise a scheme to determine whether any subsample is degenerate. The detection of degeneracy within RANSAC is the subject of (Torr *et al.* 1995a).

RANSAC originated in work on computer vision, and it was some years until a similar highly robust estimator was developed independently in the field of statistics, namely Rousseeuw's least median of squares (LMS) estimator (Rousseeuw 1987). The algorithms differ slightly in that the solution giving least median is selected as the estimate in (Rousseeuw 1987). Both algorithms have been implemented with the Sampson/Weng and epipolar distances, not on the algebraic distance, and are presented in the Appendix.

The variances of the estimators, for D and E, run over the correspondences stored in the data base are presented in Figure 9. Both estimators perform similarly very well, with LMS giving a slightly better performance for under 50% contamination.

A more recent random sampling algorithm is MINPRAN (minimize probability of randomness), described by Stewart (1995). Like LMS, this does not require *a priori* knowledge of the variances. However, we do not use it here as it appears to make assumptions about the error distribution that are inappropriate for estimation of the fundamental matrix. This is discussed in (Torr *et al.* 1996).

7. A note on Hough Transforms

In our study, techniques that fall into the three categories described already have proved the most successful. It is worth however mentioning the Hough transform (eg Ballard & Brown 1982) as it

has long history of valuable service to computer vision. The parameter space is divided into cells, and each datum adds a vote to every cell of the parameter space whose parameters are consistent with that datum. After voting, cells in the parameter space that have a number of votes greater than a given threshold are marked as representing possible solutions.

The Hough transform runs into problems when the dimension of the parameter space is high, because its space requirement is exponential in the dimensionality and the computational expense of even the Fast Hough Transform (Li *et al.* 1986) rises exponentially with the dimension of the parameter space (McLauchlan 1990). The dimensionality of the Fundamental Matrix is 7, and even the coarsest quantization of the parameter space, say into 10 cells per dimension, would demand 10^7 cells!

8. Robust determination of the standard deviation

Robust techniques to eliminate outliers are all founded upon some knowledge of the standard deviation σ of the error, as outliers are typically discriminated from inliers using

$$i \in \begin{cases} \text{set of inliers} & \text{if } d_i \leq 1.96\sigma \\ \text{set of outliers} & \text{otherwise,} \end{cases} \quad (11)$$

where we recall that $d_i = w_{Si} r_i$ is the Sampson/Weng distance. This section describes a robust method for estimating σ.

The standard deviation is related to the characteristics of the image, the feature detector and the matcher. Often the value of σ is unknown, in which case it must be estimated from the data. If there are no outliers in the data the σ can be estimated directly as the standard deviation of the residuals of a non-linear least squares minimization. If there are outliers and they are in the minority, a first estimate of the variance can be derived from the median squared error of the chosen parameter fit (Rousseeuw 1987). It is known that $\text{med}_i |d_i| / \Phi^{-1}(0.75)$ is an asymptotically consistent estimator of σ when the d_i are distributed like $N(0, \sigma^2)$, where Φ is the cumulative distribution function for the Gaussian probability density function.

It was shown empirically by Rousseeuw (1987) that when $n \approx 2p$ (recall that n is the number of data, and p the number of parameters) the correction factor of $\left(1 + \frac{5}{n-p}\right)$ improves the estimate of the standard deviation. Noting $1/\Phi^{-1}(0.75) = 1.4826$ the estimate of σ is then

$$\sigma = 1.4826 \left(1 + \frac{5}{n-p}\right) \sqrt{\text{med}_i |d_i|} \, .$$

The LMS algorithm is used to obtain the estimate of the median and a first estimate of the standard deviation. Once the final result is obtained (after non-linear minimization) the estimate of the standard deviation can be improved by the EM algorithm (Dempster 1977). If the inlier and outlier distribution are Gaussian but with different parameters, the EM algorithm is guaranteed to increase the likelihood of the estimated standard deviation given the data.

The standard deviation can be estimated between each pair of images and the results filtered over time. Under Gaussian assumptions it can be shown (Bar-Shalom & Fortmann 1988) that about 800 correspondences are required to ensure that there is a 95% chance that the variance is within 10% of its actual value. Image pairs that give rise to unusually high standard deviations might possess independently moving objects, the detection of which is discussed in (Torr *et al.* 1994,1995b).

Given random perturbations of the image correspondences with unit standard deviation then the estimate of the standard deviation of $\mathbf{F}$ was found to be 1.07, this being a conflation of the image error and the error in the estimator.

9. Comparison of Robust Categories I-III

Figure 10 is a graph comparing the best method from the Least Squares category and the robust categories I-III described in this paper.

As expected, the least squares method is non-robust, giving a standard deviation of 4.7 when only 5% of the data are outliers. The M-estimator using the Huber error provides rather inaccurate results up to figures of 35% outliers and then breaks down. The case deletion diagnostics work

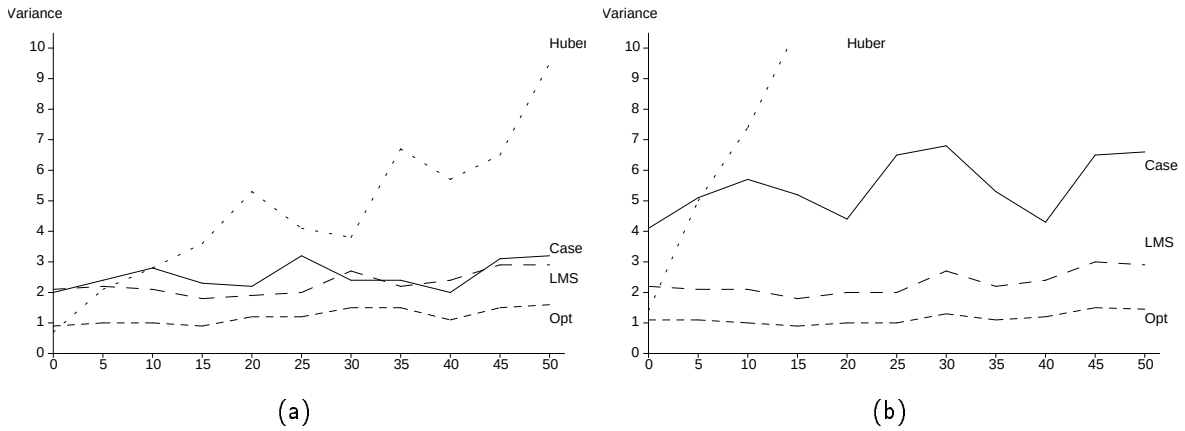

Fig. 10. (a) (b) Variance of the noise free points to the estimated **F** for the best of each category of estimator, using Sampson's distance D and the epipolar distance E respectively. The values were derived from over 100 tests on 200 points. It can be seen that Random Sampling (here using LMS) give the best result, the case deletion method performs well but requires an exact estimate of σ to achieve such a good result, limiting its use in practise. The near optimal estimator that we suggest is shown, using the LMS method to initialize and M-estimator is significantly better than the use of the Random Sampling method alone.

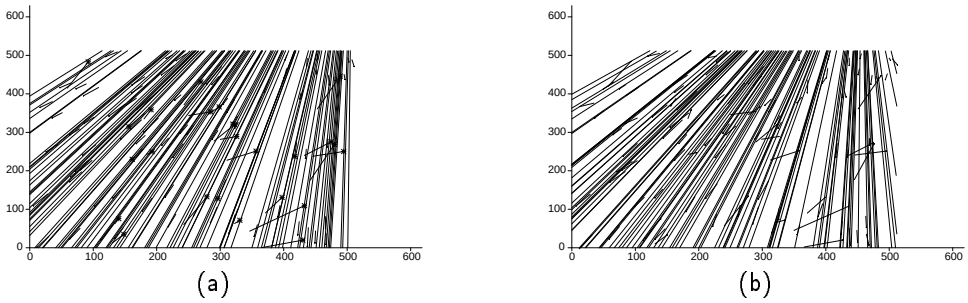

Fig. 11. A comparison of epipolar geometry estimated by LMS alone (a) with LMS followed by Huber (b). The data are corrupted from Figure 4, where the true epipolar geometry can be seen.

very well when provided with an accurate estimate for the standard deviation, which we assumed was known for these experiments. But if the standard deviation is unknown they perform very badly. The LMS algorithm gives the best performance for both error measures. RANSAC gave an equivalent or slightly worse performance when the standard deviation of the error term was known, but it has been shown that RANSAC can perform well even when there are 90% outliers (Roth 1993). This tallies with our experience in using RANSAC for motion segmentation (Torr & Murray 1994).

Earlier it was noted that iterative estimation of the M-estimators is only successful if the starting estimate was good. By using the output of the random sampling rather than linear regression as the starting estimate for M-estimation, here using an iterative Huber algorithm, a further improvement can be made, as shown for a range of contaminations in Figure 10. Although the improvement appears small, it has a significant effect on the computed epipolar geometry, as shown in Figure 11. Part (a) of the figure shows the epipolar geometry estimated by the LMS method, and (b) shows that estimated after iterative improvement of the result using Huber's M-estimator. The latter is closer to the veridical geometry derived from the uncorrupted data shown earlier in Figure 4.

The previous experiment highlights the sensitivity of the recovered epipolar geometry to fitting differences between robust methods. It is worth showing the difference between the best non-robust and best robust methods. Figure 12(a) shows the 80 correct correspondences and the 20 mismatches used as data. Figure 12(b) shows the veridical epipolar geometry, (c) shows that recov-

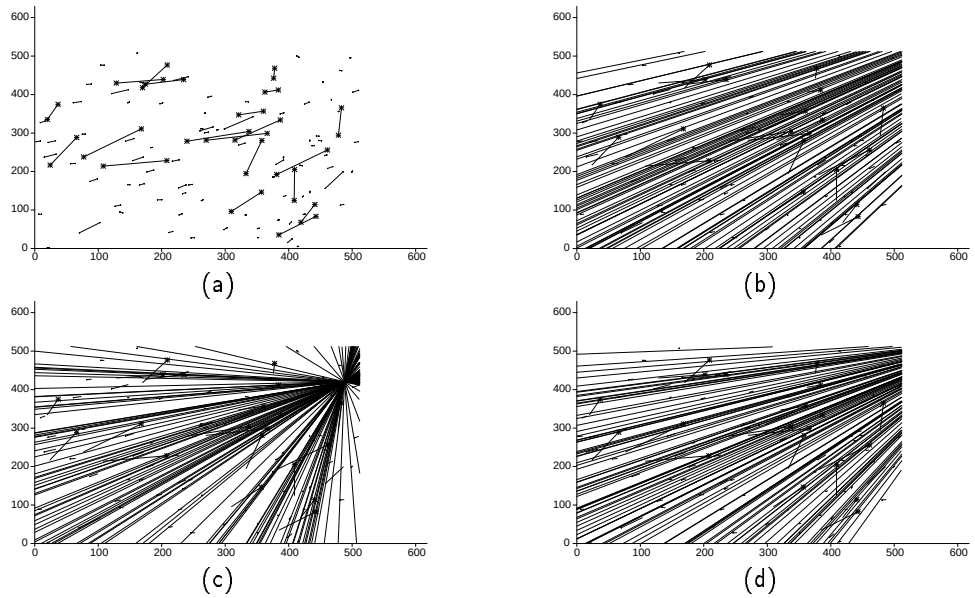

Fig. 12. Synthetically generated correspondences (a) with 80 inliers and 20 outliers. The veridical epipolar geometry is shown in (b). Part (c) Shows the epipolar geometry recovered by the non-robust algorithm **S2** and (d) shows that from the robust LMS plus Huber combination.

ered by iterative least square method S2 and (d) gives that recovered by the robust symbiotic combination of LMS with Huber. The difference is obviously substantial.

10. Towards an empirically optimal algorithm

Our key conclusions thus far are that

1. Of the major categories of robust estimator, random sampling gives the best results.
2. It is possible further to improve performance by mixing robust methods. We have found that the use of random sampling to initialize iterative M-estimators yields an empirically optimal combination.

These observations have allowed us to create an empirically optimal combination of algorithms, as summarized below.

The algorithm first determine the set of putative correspondences using "unguided matching" — that is, image-based matching without using the epipolar geometry — storing for each point an ordered list of the more likely correspondences.

The correspondences are supplied to RANSAC, which is initialized with an approximate guess at the standard deviation. For instance on the calibration example shown later the initial estimate of the standard deviation σ is made at 0.7. After LMS is run, the algorithm provides an updated estimate of $\sigma = 0.213$ from the median, running EM reduces this estimate to 0.204. After the iterated M-estimator the EM algorithm gives $\sigma = 0.1682$, and after the non-linear gradient descent part of the algorithm $\sigma = 0.1146$.

The best estimates of $\mathbf{F}$ and σ are handed on to the M-estimation algorithm for refinement. Here we use iterative re-weighted least squares with Huber's robust weighting function. In practice, around five iterations are adequate.

Finally, a non-linear gradient descent algorithm replaces the least-squares algorithm in the M-estimation scheme. The non-linear minimization is conducted using the method described in Gill and Murray (1978). This minimization uses a parameterization that enforces the $\det \mathbf{F} = 0$ condition.

Note first that *all* the correspondences are included at every stage. By stages 2 and 3, gross outliers are effectively removed as the Huber function places a ceiling on the value of their errors, but if the parameters move during the iterated search, marginal outliers can be re-classified as an inliers. This avoids RANSAC unduly biasing the M-estimation stages.

Note too that at each stage, and at each iterated step, the algorithm re-assesses the putative correspondences, using the epipolar geometry implicit in the running estimate of **F**. If it found that the point correspondence is an outlier, but that another reasonable match on the point's list of potential matches is an inlier, the algorithm alters the match so as to locally minimize d_i.

Empirically Optimal Algorithm

1. Generate matches: using unguided matching generate for each point feature in image 1 an ordered list of the best matching points in image 2, and similarly for image 2.
2. Random sampling: apply the estimator to the best set of matches from image 1 to 2 and vice versa. Use RANSAC if σ is known otherwise LMS.
3. Re-assess the matches, and re-estimate σ using the EM algorithm.
4. M-estimation (1): refine using iterative least squares incorporating the Sampson/Weng weights modified by the Huber robust weighting. iteratively re-weighted least squares.
5. Re-assess the matches, and re-estimate σ using the EM algorithm.
6. M-estimation (2): replace the iterative least squares by a non-linear method, using a parameterization that ensures the $\det \mathbf{F} = 0$ (see Luong 1992).
7. Re-assess the matches, and re-estimate σ using the EM algorithm.

Table 4. The performances of the best of each class of estimator, as well as our suggested combined method, on the real images, given in terms of the number of inliers found and the standard deviation of the error on the inlying set. The initial estimate of the standard for inliers was 0.5.

Fig	M Estimators		Case Deletion		RANSAC		Combined Method	
	σ	inliers	σ	inliers	σ	inliers	σ	inliers
13/14	0.36	143	0.32	143	0.30	145	0.30	137
Fig 15	0.29	450	0.34	435	0.34	451	0.34	459
Fig 16	0.43	168	0.40	149	0.45	148	0.43	164

10.1. Real Images

We now demonstrate the performance of the combination of random sampling and M-estimators on real imagery, using the results to aid corner matching. Whereas at step 4 we used the LMS algorithm for the synthetic data, here instead we use the RANSAC method with an initial estimate of $\sigma = 0.5$, unless it is certain that there are under 50% outliers.

Guided feature matching. Figure 13 shows two images of a calibration grid. The similarity of the features makes matching difficult, and Figure 13 (c) shows 187 correspondences postulated by a feature matcher based purely on intensities. There are a considerable numbers of mismatches. Part (d) of the figure shows the results of using RANSAC with Huber to eliminate outliers from computation of the fundamental matrix, and thence to use the associated epipolar geometry to guide matching. Part (e) of the the figure shows the estimated epipolar geometry for the calibration grid, along with initial set of matches.

The standard deviations and number of inliers for each method are summarized in Table 4, for this example and those in Figures 15 and 16. It can be seen that random sampling performs best, followed by Case Deletion, both of which are provide much better estimates than M-estimation *when poorly initialized* (here by OR).

The results of poorly-initialized M-estimation and Case Deletion Diagnostics are compared in Figure 14, where, using Table 4, it can be seen that although the standard deviations of the inliers are similar for the two estimators, the M-estimator has clearly failed to eliminate many outliers. This result demonstrates that use of the goodness of fit to the observed data is not always a good criterion with which to judge the estimator, bearing out our earlier decision to reject this as the measure of relative efficiency.

Football sequence. Figures 15(a) and (b) show two images of a sequence taken at a football match, with the motion computed by matching image corner features shown in the (b). The dominant image motion is the result of camera panning. Figure 15(c) shows the inliers, and (d) shows the outliers, lying predominantly on the indepen-

dently moving footballers; (e) shows the results of the M-estimator, note that the results are fairly good except for two major outliers. The reason that the M-estimators do not perform too badly in this case is because the outliers are only a small proportion of the data. (f) shows the inliers for the case deletion diagnostic. The data here is almost degenerate, most of the football supporters lie approximately on a plane and are thus consistent with many solutions. Hence different solutions have almost the same goodness of fit but with different correspondences indicated as inliers off the plane. As observed by Kumar and Hanson (1994), RANSAC performs less well when the data are near degenerate, but the subsequent use of an M-estimator after RANSAC helps stabilise the situation.

Walking sequence. Figures 16 (a) and (b) are two images from a sequence showing a person towards the camera as the camera moves to keep him in view. Figure 16 (c) shows the inliers and (d) the outliers. By inspection it can be seen that most of the grossly incorrect correspondences have been classified as outlying, (e) and (f) show the corresponding results. Again because the data are near degenerate, containing several large planes, the fit is not stable, and some outliers are included. Thus robust estimation is not the whole story, unless stability is also considered. This will be discussed below.

11. Discussion and conclusions

11.1. Actual versus expected performance

Although the robust estimation techniques are far superior to non-robust methods such as least squares, they are still of course imperfect. The use of synthesized data allows an objective measure of performance to be obtained, and this is shown for the combination of random sampling and Huber M-estimation in Figure 17. There are two types of error possible: Type I — an outlier is wrongly classified as an inlier; and Type II — an inlier is wrongly classified as an outlier. It can be seen that over 90% of outliers are correctly classified for contaminations as great as 50% — ie there are less than 10% Type I errors.

Is this better or worse than expected? The estimation of the standard deviation σ obviously plays a key rôle here, as it determines the threshold at which an error might be considered outlying. In estimating σ it is necessary to steer between Charybdis and Scylla: a higher estimate of σ will increase the number of Type I errors whilst decreasing the number of Type II errors, and vice versa. The expected bounds on the proportion of each type of error is estimated as follows.

Under Gaussian assumptions, 95% of the population lie within 1.96σ of the mean. If this is the confidence interval established on the error, it follows that at most 95% of the inliers would be identified correctly and that that there would be $\geq 5\%$ Type I errors.

A higher percentage of Type II errors is expected. This is because the epipolar constraint only allows disambiguation in one dimension, and so a mismatch that happens to lie along the epipolar line cannot be identified. Here we present an argument for distance e, a similar argument can be constructed for d. If the search window size is $l \times l$ pixels then the maximum area that is swept out within a distance ϵ of an epipolar line is below $2\sqrt{2}\epsilon l$. The chance of a mismatch being within this area at random is thus $2\sqrt{2}\epsilon/l$. Here, this value is approximately 8%, and so we expect at most 92% of the inliers would be correctly identified.

It can be seen in Figure 17 that the algorithm comes close to attaining these values over a wide range of contaminations.

For real data, it was found, after the EM algorithm was applied, that the difference between the inlier and outlier standard deviations was substantial, allowing a clear discrimination between inliers and outliers in most cases. For the images in Figure 13 the initial estimate of the standard deviation was 0.5, after application of the EM algorithm the standard deviation of inliers was 0.146 and that of outliers was 19.306. Generally the RANSAC algorithm was robust to the initial estimate of σ because the inlier and outlier distributions are so distinct.

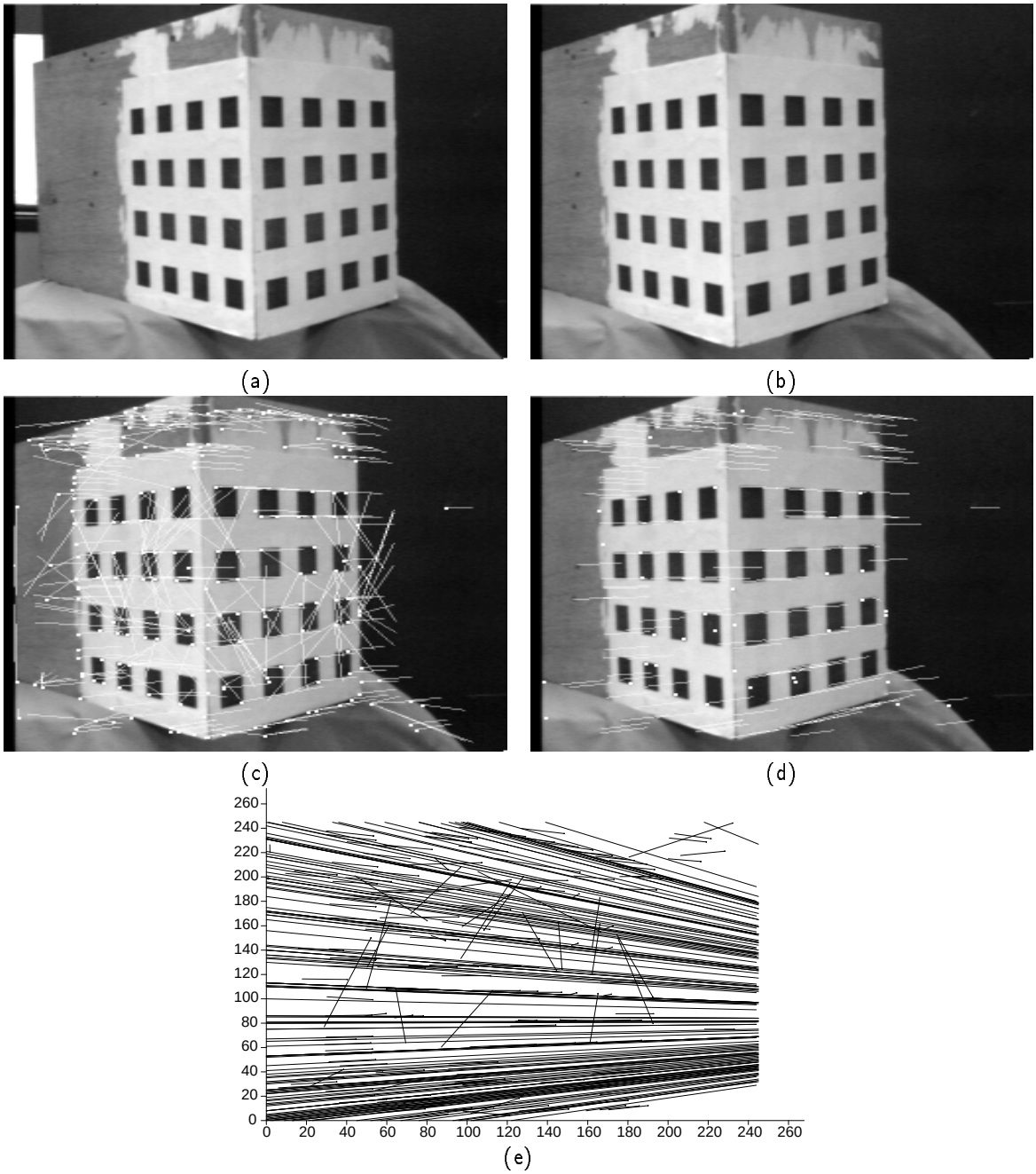

Fig. 13. (a) and (b) are two view of an object which creates matching difficulties for an impoverished matcher, as shown by the number of mismatches in (c). (d) shows the matches consistent with the epipolar geometry after eliminating outliers. Part (e) shows the estimated epipolar geometry, together with the matches that have been rejected and re-matched.

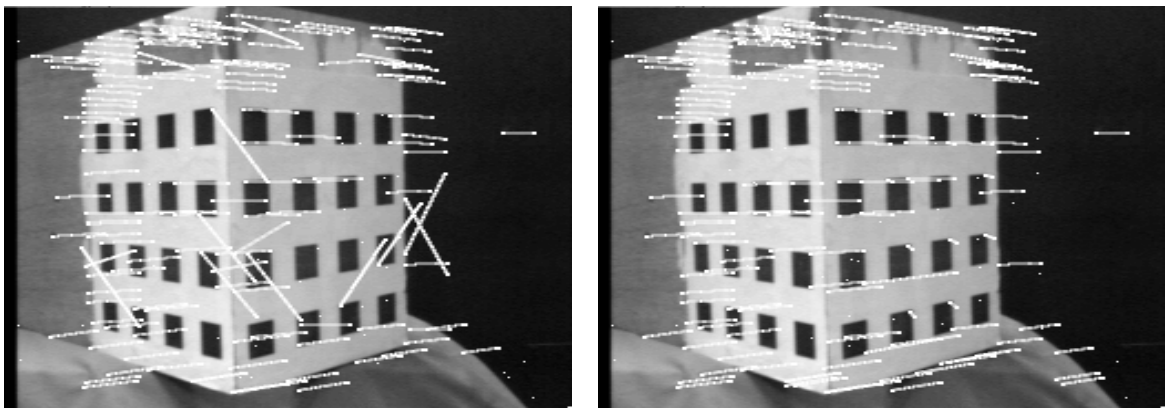

Fig. 14. (a) and (b) shows the matches consistent with the epipolar geometry after eliminating outliers. (a) is for the Huber M-estimator, initialized using OR, which has converged to an incorrect solution. (b) is for the case deletion diagnostic which, despite gross outliers has converged to a reasonable solution.

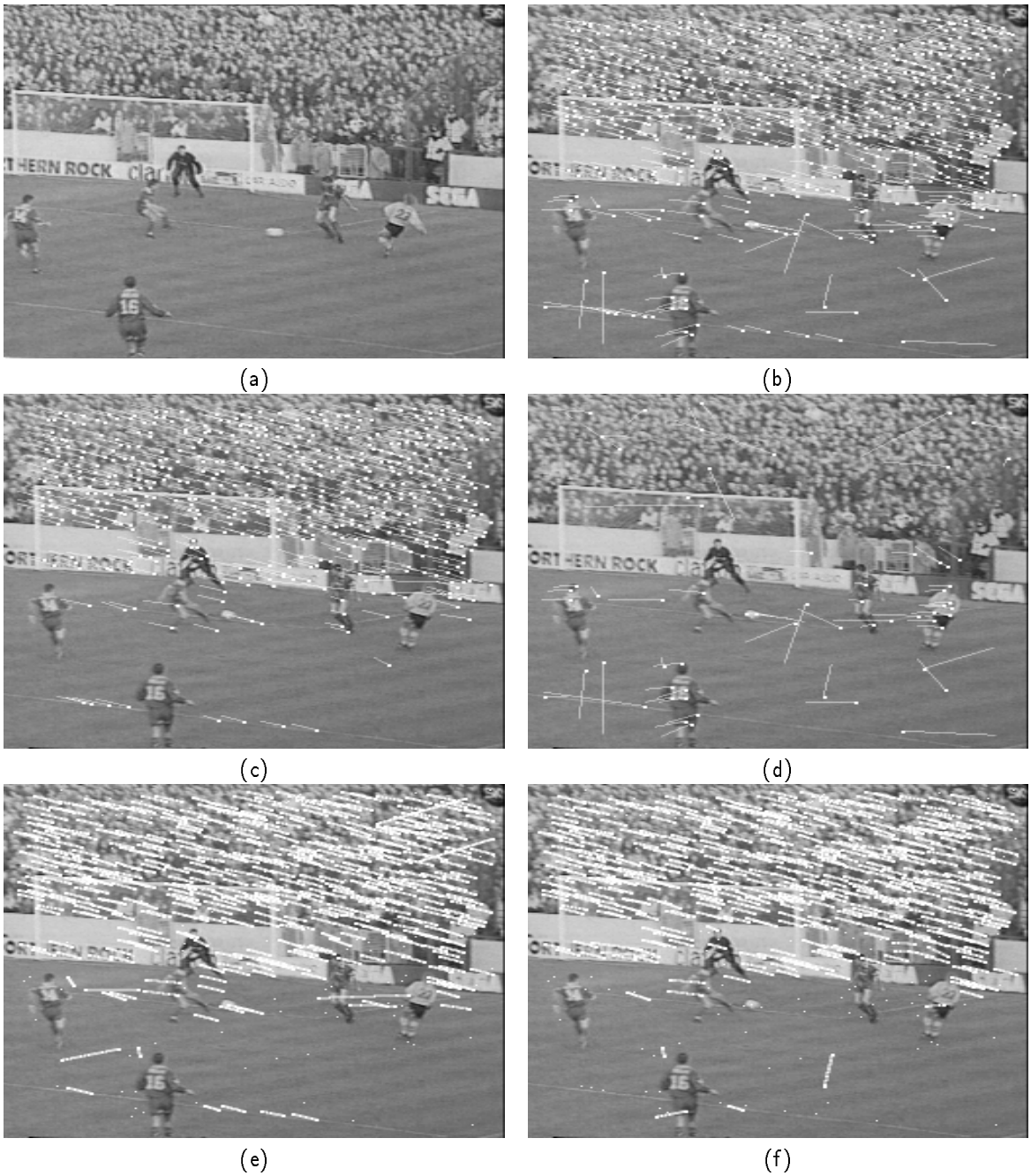

Fig. 15. In (a) (b) two consecutive images of a football match where the camera is panning. The inliers, mainly on the crowd "texture" are shown in (c), and the outliers, many of which are attached to the independently moving players, are given in (d). (e) inliers from M-estimator. (f) inliers from case deletion diagnostic.

Fig. 16. Two images from a sequence showing the movement of a person towards the camera as the camera moves to keep him in view, with the resulting correspondences from corner matching in (b). The combined RANSAC/Huber M-estimator segments the set of correspondences into (c) inliers and (d) outliers. (e) inliers from M-estimator. (f) inliers from case deletion diagnostic.

11.2. Conclusion

In this paper we have surveyed the range of categories of robust estimators, and have applied them to the computation of the fundamental matrix. We have extended several results obtained in the statistical literature for ordinary regression to orthogonal regression, where the computation of the fundamental Matrix is a linear problem and a hyperplane is being fitted, and thence to use of the geometric distance, where computing the fundamental matrix is a non-linear (actually bilinear) problem, and a hyper-surface is fitted.

Intra-category comparisons were made using large synthetic data sets and the best in each category compared in an inter-category competition. Methods were evaluated on relative efficiency and breakdown point. Random sampling techniques were shown to provide the best solution, in tests LMS generally gave a better fit than RANSAC and we advocate its use, but RANSAC is superior in the following circumstances: Firstly when there are more than 50% outliers LMS cannot provide a good solution, this might occur in cases of independent motion. Secondly if half the data are well fitted by multiple solutions for $\mathbf{F}$ (i.e. half the data are on a plane), the LMS fit will be very unstable. Thus RANSAC is a more generally robust algorithm than LMS, and should be applied when the input might fall into the above categories.

M-estimators, which are more satisfying from a statistical standpoint, were shown to suffer if the initial estimate was poor, as when initialization is performed using non-robust least squares. However, if the M-estimator method was initialized using robust random sampling, the combination provided better results than random sampling alone. Even small improvements have a marked effect on the resulting epipolar geometry.

For the M-estimation itself an iterative least squares scheme was developed. Of itself it can not enforce the constraint that the determinant of the fundamental matrix must be zero, and so in a final step, a non-linear minimization replaces the iterative least squares, using a parameterization that enforces $|\mathbf{F}| = 0$. This ensures unbiased recovery of the epipole, as described by Luong *et al.* (1993). We have found that this tripartite approach gives very satisfactory results. The experiments on real imagery showed that overall estimation could be further improved by using the robust estimation to provide epipolar constraint to the matcher.

The above method functions even with a high degree of outlier contamination, but at the expense of computation time. If there are few outliers and cost is an issue then case deletion diagnostics provide an efficient way of judging the relative merit of correspondences. Case deletion methods work well on smaller data sets.

We have not considered here structural constraints here, in particular the visibility constraint. Some outliers, although consistent with the epipolar geometry, might appear behind the camera, allowing them to be identified as outlying. Nor have we considered the natural extension of the estimation process to motion segmentation: this is explored in (Torr & Murray 1993,1994) and (Torr *et al.* 1995).

There have been some other notable comparisons of robust estimators in computer vision, we shall now compare our findings with those of the other studies. Meer *et al.* (1991) compared M-estimators and LMS for data smoothing and found that LMS more than halved the error, giving clearly superior results. Kumar and Hanson (1994) compared M-estimation and random sampling using both the LMS and RANSAC error criteria for the problem of pose determination. The conclusion they reach is equivocal, and they quote from Li (1985):: "No one robust regression technique has proven superior to all others in all situations, partly because of handling many forms of influential observations".

Our results force us to disagree with this pessimistic view, and we now explore the reasons for this difference in opinion. They reject random sampling in certain cases due to the following rationale. "Given the observations for the inliers are noisy, it is conceivable that the pose returned by the consensus algorithm explains a significant set of observation with "low leverage" quite well and makes an inlier with "high leverage" an outlier. This indeed might occur if 50 percent of the data are consistent with multiple solution for LMS, in which case the median will be near zero for wildly different solutions to the data. But our suggested approach of RANSAC combined with an M-estimator will typically overcome this prob-

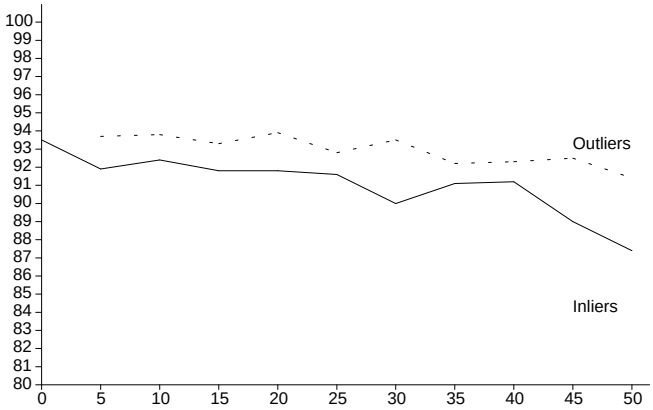

Fig. 17. The percentage of outliers and inliers correctly discovered for given percentages of contamination using the best robust estimator combining random sampling and Huber.

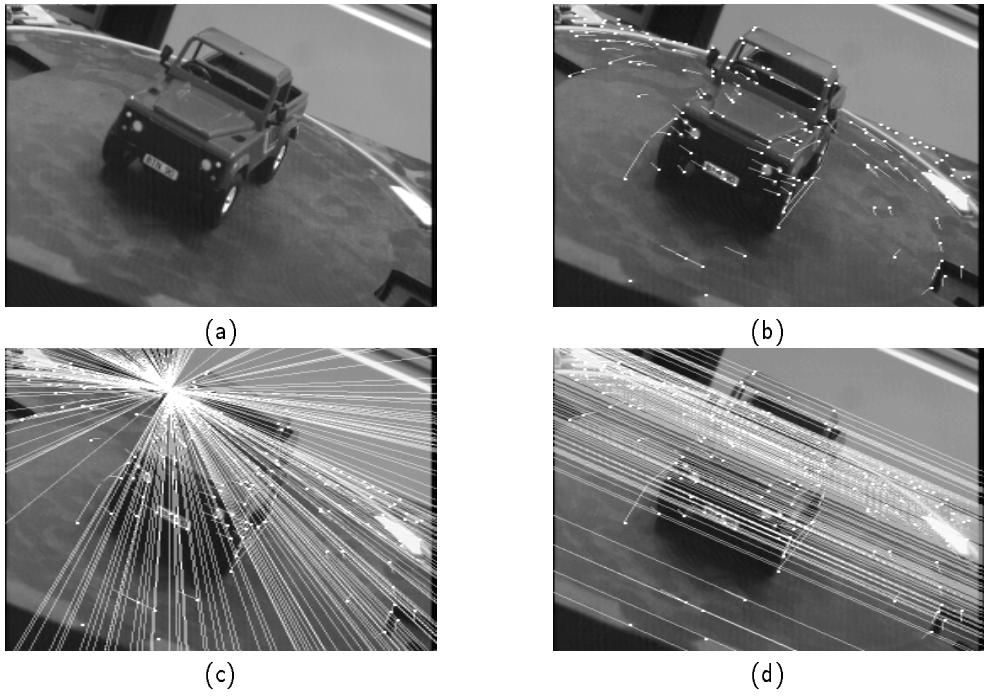

Fig. 18. In (a) (b) two consecutive images of a buggy rotating on a turntable. (b) has 167 matches superimposed on the second image. (c) (d) show two epipolar geometries generated by two distinct fundamental matrices, 139 correspondences are consistent with the fundamental matrix in (a), 131 are consistent with the fundamental matrix in (b) yet the two epipolar geometries obviously differ.

lem. We suggest using RANSAC together with some first reasonable guess at the standard deviation σ (possibly even obtained from LMS), and as noted the final result is reasonably insensitive to the first guess at σ. The problem of degeneracy (multiple solutions fitting the data) is discussed further in the next section.

The work of Zhang *et al* (1994) was developed independently of ours. They too explored the use of robust estimators to estimate the fundamental matrix. Although there is little comparative work between estimators in their paper, they suggest using LMS to estimate $\mathbf{F}$. Generally we prefer RANSAC followed by an M-estimator, unless the input is controlled to exclude independent mo-

tion etc, as noted previously. Our method requires somewhat less sampling in the random sampling phase, using only 7 points rather than 8 as they require. This leads to a speedier convergence. Furthermore, any $\mathbf{F}$ obtained from 7 points automatically has $\det \mathbf{F} = 0$ whereas that obtained from 8 points will not. This is useful if the final result of RANSAC is fed directly into a non-linear minimizer. Another difference is that Zhang *et al.* use the epipolar distance e, which we have rejected in favour of the Sampson/Weng measure, which is theoretically more satisfying as it is a closer approximation to the maximum likelihood estimator. Their paper points out that linear methods carry bias. In this paper, we have discussed a robust non-parametric method for removing this bias. Our work has carried the analysis further by assessing the fit to the *true data* set when the ground truth is known. We also note that our test databases are larger than in previous studies, giving a greater indication of the reliability of the result.

11.3. A postscript on degeneracy

It has already been noted that when using random sampling there might be degeneracy in any seven point samples. In fact there is a broader problem here, one that has been long neglected in the statistics literature. It is how are the possible solutions to be determined if the true inlying data as a whole are degenerate, but degeneracy is broken by a handful of rogue outliers? Figure 18 gives an example. Parts (a) and (b) show two frames and the resulting point correspondences from a sequence where a toy truck is rotated on a turntable. The majority of 167 data are degenerate, and Figures 18 (c) and (d) shows two epipolar geometries consistent with the veridical inlying data. Running random sampling can generate approximations to either the first has 139 inliers, the second 131. When random sampling is performed on this set, either solution can be found.

The question arises as to how to determine which solution is valid, or whether the data are degenerate. This question becoming more problematic the more outliers there are in the data. An account of when degeneracy might arise, how to detect it and how to conduct estimation in the presence of degeneracy is given in (Torr et al. 1995a) and will be the subject of a sister paper.

Acknowledgements

This work was supported by a Grant GR/J65372 and a Research Studentship to PHST from the UK Engineering and Physical Science Research Council. The authors would like to thank Paul Beardsley, Andrew Zisserman, Steven Maybank, Larry Shapiro, Mike Brady and the reviewers for their comments.

Appendix

M-estimator algorithm

1. Initialize the weights $w_i = 1$, $\gamma_i = 1$ for each correspondence.
2. For a number of iterations (we used 5):
   - 2.1 Weight the ith constraint by multiplying it by $w_i\gamma_i$.
   - 2.2 Calculate $\mathbf{F}$ by orthogonal regression.
   - 2.3 Project the estimated $\mathbf{F}$ onto the nearest rank 2 matrix using the singular value decomposition.
   - 2.4 Calculate the algebraic residuals r_i.
   - 2.5 For each correspondence (dropping the subscript i) calculate weighting

$$w_S = \left(\frac{1}{r_x^2 + r_y^2 + r_{x'}^2 + r_{y'}^2} \right)^{1/2}$$

   - 2.6 Calculate the distance $d_i = w_{Si} r_i$.
   - 2.7 Calculate γ_i, e.g. for Huber:

$$\gamma_i = \begin{cases} 1 & d_i < \sigma \\ \sigma/|d_i| & \sigma < d_i < 3\sigma \\ 0 & d_i > 3\sigma. \end{cases}$$

Case deletion algorithm

1. Set weights $w_i = 1$, for each correspondence.
2. Until $\sum_i d_i^2 < \chi^2$ where $i \in \{\text{inliers}\}$, do:
   - 2.1 Weight the ith constraint by multiplying it by w_i.
   - 2.2 Calculate $\mathbf{F}$ by orthogonal regression using all correspondences that are still inlying.

2.3 Project the estimated $\mathbf{F}$ onto the nearest rank 2 matrix using the singular value decomposition.

2.4 Calculate the algebraic residuals r_i.

2.5 Calculate the influence of each correspondence T_i.

2.6 Cast out correspondence with largest T_i.

2.7 For each correspondence (dropping the subscript i) calculate the weighting

$$w_S = \left(\frac{1}{r_x^2 + r_y^2 + r_{x'}^2 + r_{y'}^2} \right)^{1/2} .$$

2.8 Calculate the distance $d_i = w_{Si} r_i$.

The Random Sampling Algorithm

1. Repeat for m samplings as determined in Table 3:

   1.1 Select a random sample of the minimum number of data points to make a parameter estimate $\mathbf{F}$.

   1.2 Calculate the distance measure d_i for each feature given $\mathbf{F}$.

   1.3 If using RANSAC, calculate the number of inliers consistent with $\mathbf{F}$, using the method prescribed in Section 8
   Else if using the LMS estimator calculate the median error.

2. Select the best solution — i.e. the biggest consistent data set. In the case of ties select the solution which has the lowest standard deviation of inlying residuals.
3. Re-estimate the parameters using all the data that has been identified as consistent. A more effective, and possible computationally expensive estimator such as Powell's method (Teukolsky *et al.* 1988; NAG 1988) may be used at this point.

Notes

1. If the points have unequal variance each element may be weighted by its standard deviation.
2. Kanatani (1994) page 322 provides an interesting discussion about this assertion.
3. The robber *Prokroustes* was fabled to fit victims to his bed by stretching or lopping. Hartley (1995) has suggested a preconditioning that should be used before the fundamental matrix is replaced by its nearest rank 2 equivalent.
4. In the case when different axes have different variances we transform the data by scaling all the coordinates (e.g. each column of $\mathbf{Z}$) by their standard deviation, in order to obtain uniform variance.

References

Ballard, D.H. and Brown, C.M., 1982. *Computer Vision.* Prentice–Hall, New Jersey.

Bar-Shalom, Y. and Fortmann, T.E., 1988. *Tracking and Data Association.* Academic Press.

Beardsley, P.A., Torr, P.H.S and Zisserman, A.P., 1996. 3d model acquisition from extended image sequence. OUEL Report 2089/96, Department of Engineering Science, University of Oxford.

Bookstein, F., 1979. Fitting conic sections to scattered data. *Computer Vision Graphics and Image Processing*, 9:56–71.

Chaterjee, S. and Hadi, A.S., 1988. *Sensitivity Analysis in Linear Regression.* John Wiley, New York.

Cook, R.D. and Weisberg, S., 1980. Characterisations of an empirical influence function for detecting influential cases in regression. *Technometrics*, 22:337–344.

Critchley, F., 1985. Influence in principal component analysis. *Biometrika*, 72:627–636.

Dempster, A.P, Laird, N.M. and Rubin, D.B., 1997. Maximum likelihood from incomplete data via the em algorithm. *J. R. Statist. Soc.*, 39 B:1–38.

Devlin S.J., Gnanadesikan, R., and Kettering, J.R., 1981. Robust estimation of dispersion matrices and principal components. *J. Amer. Stat. Assoc.*, 76:354–362.

Faugeras, O.D., 1992. What can be seen in three dimensions with an uncalibrated stereo rig? In G. Sandini, editor, *Proc. 2nd European Conference on Computer Vision , LNCS 588*, pages 563–578.

Fischler, M.A. and Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography. *Commun. Assoc. Comp. Mach.*, vol. 24:381–95.

Gill, P.E. and Murray, W., 1978. Algorithms for the solution of the nonlinear least-squares problem. *SIAM J Num Anal*, 15(5):977–992.

Golub, G.H., 1973. Some modified eigenvalue problems. *SIAM Review*, 15(2):318–335.

Golub, G.H. and van Loan, C.F., 1989. *Matrix Computations.* The John Hopkins University Press.

Numerical Algorithms Group, 1988. *NAG Fortran Library vol 7.*

Gu, M. and Eisenstat, S.C., 1995. Downdating the singular value decomposition. *SIAM J. Matrix Analysis and Applications*, 16:793–810.

Hampel J.P., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A., 1986. *Robust Statistics: An Approach Based on Influence Functions.* Wiley, New York.

Hartley, R.I., 1992. Estimation of relative camera positions for uncalibrated cameras. In *Proc. 2nd European Conference on Computer Vision , LNCS 588, Santa Margherita Ligure*, pages 579–587. Springer-Verlag.

Hartley, R.I., 1995. In defence of the 8-point algorithm. *Proc. 5th Int Conf on Computer Vision, Boston MA*, pp 1064–1070. IEEE Computer Society Press, Los Alamitos CA.

Hartley, R.I. and Sturm, Y., (1994). Triangulation. In *AIUWS*, pages 957–966.

Exploring data tables, trends and shapes. Hoaglin, D.C., Mosteller, F. and Tukey, J.W. editors (1985) *Robust Regression.* John Wiley and Sons.

Huber, P.J., 1981. *Robust Statistics.* John Wiley and Sons.

Kanatani, K., 1992. *Geometric Computation for Machine Vision.* Oxford University Press.

Kanatani, K., 1994. Statistical bias of conic fitting and renormalization. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 16(3):320–326.

Kanatani, K., 1996. *Statistical Optimization for Geometric Computation: Theory and Practice.* Elsevier Science, Amsterdam.

Kendall, M. and Stuart, A., 1983. *The Advanced Theory of Statistics.* Charles Griffin and Company, London.

Kumar, R. and Hanson, A.R., 1994. Robust methods for estimating pose and a sensitivity analysis. *Computer Vision, Graphics and Image Processing*, vol. 60(3):313–342.

Li, G., 1985. Exploring data tables, trends and shapes. In D. C. Hoaglin, F. Mosteller, and J. W. Tukey, editors, *Robust Regression*, pages 281–343. John Wiley & Sons.

Li, H., Lavin, M.A. and LeMaster, R.J., 1986. Fast Hough transforms: a hierarchical approach. *Computer Vision, Graphics and Image Processing*, 36:139–161.

Longuet-Higgins, H.C., 1981. A computer algorithm for reconstructing a scene from two projections. *Nature*, vol.293:133–135.

Luong, Q.T., 1992. *Matrice fondamentale et calibration visuelle sur l'environnement: Vers use plus grande autonomie des systemes robotiques.* PhD Thesis, Paris University.

Luong, Q.T., Deriche, R., Faugeras, O.D. and Papadopoulo, T., 1993. On determining the fundamental matrix: analysis of different methods and experimental results. INRIA Technical Report 1894, INRIA-Sophia Antipolis.

Maronna, R.A., 1976. Robust M-estimators of multivariate location and scatter. *Ann. Stat.*, 4:51–67.

McLauchlan, P.F., 1990. *Describing Textured Surfaces using Stereo Vision.* PhD Thesis, AI Vision Research Unit, University of Sheffield.

Meer, M., Mintz, D. and Rosenfeld, A. (1991). Robust regression methods for computer vision: A review. *International Journal of Computer Vision*, 6:59–70.

Mosteller, F. and Tukey, J.W., 1977. *Data and Analysis and Regression.* Addison-Wesley, Reading MA.

Olsen, S.I., 1992. Epipolar line estimation. In G. Sandini, editor, *Proc. 2nd European Conference on Computer Vision, LNCS 588*, pages 307–311. Springer-Verlag.

Pearson, K., 1901. On lines and planes of closest fit to systems of points in space. *Philos. Mag. Ser. 6*, 2:559.

Pratt, V., 1987. Direct least squares fitting of algebraic surfaces. *Computer Graphics*, 21(4):145–152.

Roth, G. and Levine, M.D., 1993. Extracting geometric primitives. *Computer Vision, Graphics, and Image Processing*, 58(1):1–22.

Rousseeuw, P.J., 1987. *Robust Regression and Outlier Detection.* Wiley, New York.

Sampson, P.D., 1982. Fitting conic sections to 'very scattered' data: An iterative refinement of the Bookstein algorithm. *Computer Graphics and Image Processing*, 18:97–108.

Shapiro, L.S. and Brady, J.M., 1993. Rejecting outliers and estimating errors in an orthogonal regression framework. Department of Engineering Science, University of Oxford, Technical Report OUEL 1974/93, and to appear *Philosophical Transactions of the Royal Society.*

Spetsakis, M. and Aloimonos, Y., 1991. A multi-frame approach to visual motion perception. *International Journal of Computer Vision*, 6:245–255.

Sprent, P., 1989. *Applied Nonparametric Statistical Methods.* Chapman and Hall, London.

Stewart, C.V., 1995. MINPRAN, a new robust estimator for computer vision. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.PAMI-17,no.10:925–938.

Taubin, G., 1991. Estimation of planar curves, surfaces, and nonplanar space curves defined by implicit equations with applications to edge and range image segmentation. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.PAMI-13,no.11:1115–1138.

Thisted, R.A., 1988. *Elements of Statistical Computing.* Chapman and Hall, New York.

Torr, P.H.S., 1995. *Outlier Detection and Motion Segmentation.* DPhil Thesis, University of Oxford.

Torr, P.H.S, Maybank, S. and Zisserman, A. (1996) Robust detection of degenerate configurations for the fundamental matrix. OUEL Report 2090/96, Department of Engineering Science, University of Oxford.

Torr, P.H.S. and Murray, D.W., 1992. Statistical detection of non-rigid motion. In D. Hogg, editor, *Proc. 3rd British Machine Vision Conference, Leeds, September 1992*, pages 79–88. Springer-Verlag.

Torr, P.H.S. and Murray, D.W., 1993a. Statistical detection of independent movement from a moving camera. *Image and Vision Computing*, 1(4):180–187.

Torr, P.H.S. and Murray, D.W., 1993b. Outlier detection and motion segmentation. In P. S. Schenker, editor, *Proc. Sensor Fusion VI, Boston MA, June 1993*, pages 432–443. SPIE volume 2059.

Torr, P.H.S. and Murray, D.W., 1994. Stochastic motion segmentation. In J-O. Ecklundh, editor, *Proc. 3rd European Conference on Computer Vision, Stockholm, May 1994*, pages 328–338. Springer–Verlag.

Torr, P.H.S., Zisserman, A. and Maybank, S., 1995a. Robust detection of degeneracy. *Proc. 5th Int Conf on Computer Vision, Boston MA*, pp 1037–1044. IEEE Computer Society Press, Los Alamitos CA.

Torr, P.H.S, Zisserman, A. and Murray, D.W., 1995b. Motion clustering using the trilinear constraint over three views. In R. Mohr and C. Wu, editors, *Europe-China*

Workshop on Geometrical Modelling and Invariants for Computer Vision, pages 118–125. Springer–Verlag.

Tsai, R.Y. and Huang, T.S., 1984. 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.

Teukolsky, S.A., Press, W.H., Flannery, B.P. and Vetterling, W.T., 1988. *Numerical Recipes in C, the art of scientific computing*. Cambridge University Press, Cambridge.

Weng, J., Ahuja, N. and Huang, T.S., 1993. Optimal motion and structure estimation. *IEEE PAMI*, vol.15(9):864–884.

Weng, J., Huang, T.S. and Ahuja, N., 1989. Motion and structure from two perspective views: Algorithms, error analysis, and error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:451–476.

Zhang, Z., Deriche, R., Faugeras, O.D. and Luong, Q.T., 1994. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *AI Journal*, vol.78:87–119.