Outlier Detection and Motion Segmentation

P H S Torr and D W Murray

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

We present a new method for solving the problem of motion segmentation, identifying the objects within an image moving independently of the background. We utilise the fact that two views of a static $3D$ point set are linked by a 3×3 *Fundamental Matrix* [**F**]. The Fundamental Matrix contains all the information on structure and motion from a given set of point correspondences and is derived by a least squares method under the assumption that the majority of the image is undergoing a rigid motion. Least squares is the most commonly used method of parameter estimation in computer vision algorithms. However the estimated parameters from a least squares fit can be corrupted beyond recognition in the presence of gross errors or *outliers* which plague any data from real imagery. Features with a motion independent of the background are those statistically inconsistent from the calculated value of [**F**]. Well founded methods for detecting these outlying points are described.

1 Introduction

The introduction of the computer into the field of model fitting has opened up new lines of research that were previously denied to us by reason of their computationally intensive nature. Within this paper we examine methods for the detection of outliers to a least squares fit that would have been previously computationally infeasible. The fitting of linear regression models by least squares is undoubtedly the most widely used modelling procedure. A major drawback, however, is that outliers which are inevitably included in the initial fit can so distort the fitting process that the resulting fit can be arbitrary. A common practice is to search for outliers using the raw residuals. However, the use of these on their own can be misleading.

Much work has been done already on detecting outliers given the case of non-orthogonal regression (reviewed in [2]), where we choose to regress against a given, dependent, variable. Unfortunately there is in this method a tacit assumption that all the error is concentrated in the dependent variable and furthermore that the dependent variable has a non-zero coefficient. In many engineering situations we cannot guarantee these conditions and we must resort to orthogonal regression, where we minimize the sum of squares of the perpendicular distances (i.e. the residuals) between each point and the fitted hyperplane. Little work has been done on outlier detection for orthogonal regression, with the exception of [15]. In this paper we outline two methodologies for outlier detection. In sections 2–5 we describe an extension of previous outlier diagnostics to the realm of orthogonal regression. The method works by assessing the amount of influence that the deletion of each point would have on the final solution. In sections 6–7 we then apply the theory we have developed to the calculation of the Fundamental matrix—a necessary first step in many structure and motion algorithms. Finally in section 8 we outline an alternative approach to outlier detection and show in section 9 how this can be applied to the problem of motion segmentation. We try to fit, with a robust estimator, a single set of motion parameters (encapsulated within the Fundamental matrix) and exclude data with motion inconsistent with this as outliers.

2 Previous work on outlier detection

The basic concept underlying influence is quite simple, we introduce small perturbations into some aspect of the problem formulation and then assess how this changes the outcome of the analysis. The important issues then are the determination of the type of perturbation scheme, the particular aspect of the analysis to monitor and the method of assessment. The general procedure for assessing the influence of a given datum in a regression analysis is to determine the changes that occur when the datum is omitted. Several measures of influence exist in the literature. 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. A diagnostic that is typical of those currently in use is Cook's D [3], which measures the effect on the solution of the deletion of a given datum. Unfortunately, this method was

derived for standard regression when there is error only in the measurement of the dependent variable and so is not applicable to many categories of problem. When there can be errors in all the variables or any variable could have coefficient zero, we solve the minimization by orthogonal regression, minimizing the sum of the squares of the perpendicular distances (i.e. the residuals) between each point and the fitted hyperplane We now outline a method of outlier detection applicable to orthogonal regression.

3 Extending Cook's D to the case of orthogonal regression

We shall now extend Cook's outlier method to orthogonal regression. We propose to calculate the regression, delete the point with maximal influence as outlying, repeating this procedure until we have determined that there are no more outliers. In this section we derive a formula for the influence of a point on the regression. Consider the set of points lieing on a hyperplane $\vec{\mathbf{f}}$ and let their measured values (perturbed by Gaussian noise with standard deviation σ) be: $\vec{\mathbf{z}}_i$, $i = 1 \ldots n$. If $[\mathbf{Z}]$ is the matrix whose rows are $\vec{\mathbf{z}}_i^T$ then the least squares estimate $\vec{\mathbf{f}}^*$ of $\vec{\mathbf{f}}$ is given by the eigenvector of the moment matrix $[\mathbf{M}] = [\mathbf{Z}]^T[\mathbf{Z}]$ corresponding to the minimum eigenvalue. Analogous to Cook's D we choose to observe the effect of deleting an observation on our estimated parameters:$\vec{\mathbf{f}}^*$. We could choose to examine the effects of perturbation on some other aspect of the model, e.g. the covariance matrices or the minimum eigenvalue, but as we are most interested in $\vec{\mathbf{f}}^*$ we shall monitor eigenvector perturbations. We shall calculate a first order approximation to this effect using eigenvector perturbation theory as set out in [7]. Let $[\mathbf{M}]$ be the p-dimensional symmetric moment matrix, having eigenvalues, in increasing order, $\lambda_1 \ldots \lambda_p$ with $\vec{\mathbf{u}}_1 \ldots \vec{\mathbf{u}}_p$ the corresponding eigenvectors forming an orthonormal system. If matrix $[\mathbf{M}]$ is perturbed into

$$[\mathbf{M}]' = [\mathbf{M}] + \delta[\mathbf{M}] \quad , \tag{1}$$

then, if the multiplicity of λ_j is 1, eigenvector $\vec{\mathbf{u}}_j$ is perturbed into

$$\vec{\mathbf{u}}_j' = \vec{\mathbf{u}}_j + \sum_{k \neq j} \frac{\vec{\mathbf{u}}_k^T \delta[\mathbf{M}] \vec{\mathbf{u}}_j}{\lambda_j - \lambda_k} \vec{\mathbf{u}}_k + O(\delta[\mathbf{M}])^2 \quad . \tag{2}$$

If we are examining the effects of deleting the ith observation, on the parameter estimate, then $\delta[\mathbf{M}] = -\vec{\mathbf{z}}_i \vec{\mathbf{z}}_i^T$. This allows us to calculate the change in $\vec{\mathbf{f}}^*$ to the first approximation as

$$\vec{\mathbf{f}}_{(i)}^* = \vec{\mathbf{z}}_i^T \vec{\mathbf{u}}_1 \sum_{k \neq 1} \frac{\vec{\mathbf{u}}_k^T \vec{\mathbf{z}}_i}{\lambda_1 - \lambda_k} \vec{\mathbf{u}}_k + \vec{\mathbf{u}}_1 \quad , \tag{3}$$

where $\vec{\mathbf{f}}_{(i)}^*$ is the estimate of $\vec{\mathbf{f}}^*$ with the ith element deleted.

To be most useful influence should be measured as a scalar quantity. It is necessary, therefore, to use a norm to characterize influence; this norm will map the p vector, $\delta\vec{\mathbf{f}}_{(i)}$, to a scalar. We define the norm in terms of a symmetric, positive definite $p \times p$ matrix $[\mathbf{L}]$ and a scalar c, to give an influence measure $T_i([\mathbf{L}], c)$:

$$\begin{aligned} T_i([\mathbf{L}], c) &\stackrel{\text{def}}{=} \frac{(\vec{\mathbf{f}}_{(i)}^* - \vec{\mathbf{f}}^*)^T [\mathbf{L}] (\vec{\mathbf{f}}_{(i)}^* - \vec{\mathbf{f}}^*)}{c} \\ &= \frac{e_i^2}{c} \Big(\sum_{k \neq 1} \frac{\vec{\mathbf{u}}_k^T \vec{\mathbf{z}}_i}{\lambda_1 - \lambda_k} \vec{\mathbf{u}}_k\Big)^T [\mathbf{L}] \Big(\sum_{l \neq 1} \frac{\vec{\mathbf{u}}_l^T \vec{\mathbf{z}}_i}{\lambda_1 - \lambda_l} \vec{\mathbf{u}}_l\Big) \end{aligned} \tag{4}$$

noting that $\vec{\mathbf{z}}_i^T \vec{\mathbf{u}}_1 = e_i$, the residual for the ith observation. Contours of constant $T_i([\mathbf{L}], c)$ are ellipsoids of dimension equal to the rank of $[\mathbf{L}]$, centred at $\vec{\mathbf{f}}^*$ or equivalently at $\vec{\mathbf{f}}_{(i)}^*$. Clearly the character of $T_i([\mathbf{L}], c)$ is determined by $[\mathbf{L}]$ and c, which may be chosen to reflect specific concerns. We would like to choose a norm to make T_i both scaleless and *invariant* to non-singular transformation of the data. We shall choose $[\mathbf{L}] = [\mathbf{M}]\sigma^{-2}$, in [18] it is shown that $[\mathbf{M}]\sigma^{-2}$ is equivalent to the pseudo inverse of $[\mathbf{\Gamma}]_f$, the covariance matrix of the parameter estimate ($[\mathbf{\Gamma}]_f = \sigma^2 \sum_{k=2}^{p} \frac{\vec{\mathbf{u}}_k \vec{\mathbf{u}}_k^T}{\lambda_k}$). Thus, from equation (4) we can see that:

$$T_i([\mathbf{M}], c) = \frac{(\vec{\mathbf{z}}_i^T \vec{\mathbf{u}}_1 \sum_{k \neq 1} \frac{\vec{\mathbf{u}}_k^T \vec{\mathbf{z}}_i}{\lambda_1 - \lambda_k} \vec{\mathbf{u}}_k)^T [\mathbf{Z}]^T [\mathbf{Z}] (\vec{\mathbf{z}}_i^T \vec{\mathbf{u}}_1 \sum_{l \neq 1} \frac{\vec{\mathbf{u}}_l^T \vec{\mathbf{z}}_i}{\lambda_1 - \lambda_l} \vec{\mathbf{u}}_l)}{c\sigma^2} \tag{5}$$

if we note that $\vec{\mathbf{u}}_j^T [\mathbf{Z}]^T [\mathbf{Z}] \vec{\mathbf{u}}_k = \lambda_k \vec{\mathbf{u}}_j^T \vec{\mathbf{u}}_k = \lambda_k \delta_{jk}$, since $\vec{\mathbf{u}}_j^T \vec{\mathbf{u}}_k = \delta_{jk}$ (the Kronecker delta[1]) then we get

$$T_i = \frac{e_i^2}{c\sigma^2} \sum_{k \neq 1} \left(\frac{\vec{\mathbf{u}}_k^T \vec{\mathbf{z}}_i}{\lambda_1 - \lambda_k} \sqrt{\lambda}_k \right)^2 \quad . \tag{6}$$

We can use the results of the singular value decomposition to derive a computationally simple form for equation (6). Let the singular value decomposition [16, 19] of $[\mathbf{Z}]$, be given by

$$[\mathbf{Z}] = [\mathbf{V}][\mathbf{\Lambda}][\mathbf{U}]^T \quad , \tag{7}$$

where $[\mathbf{V}]$ is a $n \times p$ matrix whose columns are the left hand singular vectors of $[\mathbf{D}]$, $[\mathbf{U}]$ is a $p \times p$ matrix whose columns are the right hand singular vectors of $[\mathbf{D}]$ and $[\mathbf{\Lambda}]$ is the diagonal matrix of the corresponding singular values of $[\mathbf{D}]$: $[\mathbf{\Lambda}] = \mathrm{diag}(\lambda_1^{\frac{1}{2}}, \lambda_2^{\frac{1}{2}}, \ldots, \lambda_p^{\frac{1}{2}})$ in ascending order such that $\lambda_1^{\frac{1}{2}}$ is the minimum singular value. Then it, if $[\mathbf{V}](i,k)$ is the ikth element of $[\mathbf{V}]$, is easy to see that

$$\vec{\mathbf{u}}_k^T \vec{\mathbf{z}}_i = [\mathbf{V}](i,k) \lambda_k^{\frac{1}{2}} \quad , \tag{8}$$

and so

$$T_i = \frac{e_i^2}{c} \sum_{k \neq 1} \left(\frac{[\mathbf{V}](i,k)\lambda_k}{\lambda_1 - \lambda_k} \right)^2 \quad . \tag{9}$$

Let us define

$$l_i \stackrel{\mathrm{def}}{=} \sum_{k \neq 1} \left(\frac{[\mathbf{V}](i,k)\lambda_k}{\lambda_1 - \lambda_k} \right)^2 \tag{10}$$

it can be shown that [18]:

$$\mathrm{VAR}(e_i) = (1 - l_i)\sigma^2 \quad , \tag{11}$$

thus we can internally Studentize the residual by dividing it by an estimate of its own standard deviation, setting $c = 1 - l_i$ leads to

$$T_i = \frac{e_i^2}{\sigma^2} \left(\frac{l_i}{1 - l_i} \right) \tag{12}$$

which is both a revealing and computationally efficient form. Thus $T_i([\mathbf{M}], c)$ is proportional to the residual multiplied by a leverage factor. This leverage factor will be large only if the orthogonal projection of the observation $\vec{\mathbf{z}}_i$ onto $\vec{\mathbf{u}}_k$, such that $k \neq 1$, is large and the corresponding eigenvalue: λ_k is small. If λ_1 is much smaller than all the other eigenvalues

$$\left(\frac{\lambda_k}{\lambda_1 - \lambda_k} \right) \approx 1 \tag{13}$$

thus $l_i \approx h_{zii}$ where h_{zii} is the ith diagonal element of the hat matrix for $[\mathbf{Z}]$, defined by $[\mathbf{Z}]([\mathbf{Z}]^T[\mathbf{Z}])^{-1}[\mathbf{Z}]^T = [\mathbf{V}][\mathbf{V}]^T$. It can be shown that h_{zii} is invariant to non-singular affine transformations of the coordinate system [18] and thus should produce more consistent results in situations were the modelling has to be repeated and the coordinate system might undergo such transformations.

$$\begin{aligned} T_i &\approx \frac{e_i^2}{c} \sum_{k=1} [\mathbf{V}](i,k)^2 \\ &= \frac{e_i^2}{c} h_{zii} \\ &\approx \frac{e_i^2}{\sigma^2} \left(\frac{h_{zii}}{1 - h_{zii}} \right) \quad . \end{aligned}$$

In our experimentation we consider the two measures:

$$T_{1i} = \frac{e_i^2}{\sigma^2} \left(\frac{l_i}{1 - l_i} \right) \tag{14}$$

[1] $\delta_{jk} = 0$ if $j \neq k$, $\delta_{jk} = 1$ if $j = k$.

and

$$T_{2i} = \frac{e_i^2}{\sigma^2}\left(\frac{h_{zii}}{1-h_{zii}}\right) \tag{15}$$

and compare the two. To remove outliers we propose to delete the point with maximal influence and recompute the regression, repeating this procedure until the maximum influence lies below a threshold, as defined below. A brief outline of the algorithm is given in figure 1.

4 A Termination Criterion

We need some method of deciding on which iteration to stop the algorithm, which should at least be after we have eliminated all the outliers. We propose to do this by examining the point with maximum influence at each iteration and seeing whether this lies below a certain threshold. Thus we propose testing the null hypothesis that the point with maximal influence is an outlier against the alternative hypothesis that it is not using a modified t-test. If the test indicates that there are any outliers then we shall delete the point with maximum influence and recompute the regression. We delete only one point at a time due to the fact that the theory we have provided is valid only for single case deletions. We must, however, exercise care in our choice of threshold, as we shall now show. The outlier diagnostic is composed of two separate terms, the Studentized residual $t_i = \frac{e_i}{\sigma_i}$ and the leverage,

$$T_{2i} = t_i^2 l_i \quad . \tag{16}$$

We shall derive a threshold by considering what would be suitable thresholds for the two separate terms, and then multiply them together. As $l_i \approx h_{ii}$ we shall set the threshold at $2p/n$. The choice of threshold for the Studentized residual requires some care. For instance, suppose that we had $n = 64$ and $p = 4$ and that we decided to test whether the residual with largest t_i was an outlier. The probability that a t statistic with 60 degrees of freedom exceeds 2.0 in absolute value is 0.05; however, the probability that the largest of 64 independent t-tests exceeds 2.0 is $1 - 0.95^{64} = 0.964$, suggesting quite clearly the need for a different critical value. The technique we have adopted is to use critical values based on the *Bonferroni inequality*[2], which states that for n tests each of size α, the probability of labelling at least one point an outlier in no greater than $n\alpha$. The procedure is conservative, since the Bonferroni inequality would specify only that the probability of the maximum of 64 tests exceeding 2.0 is no greater than 64(0.05), which is larger than 1. However, choosing the critical value to be $\frac{\alpha}{n} \times 100\%$ point of t will give a significance level of no more than $\frac{\alpha}{n} n = \alpha$. Thus in our example we could choose a level of $0.05/65 = 0.00077$ for each test to give an overall level of no more than $65(0.00077) = 0.05$. We store a look up table of different critical values. Typically for $p = 8$ and $n > 40$ the critical value tends to 3.55, thus leading to a joint threshold for T_i of $25p/n$. Thus it can be seen that the generalized distance measure T_i may also provide an improved termination criterion over previous residual based methods.

5 Results

5.1 Synthetic data tests

We now give the results of many thousands of tests done on synthetic data. There are two types of error possible: we could classify an outlier as a non-outlier (Type I error) or a non-outlier as an outlier (Type II error). A higher termination threshold will increase the number of Type I errors whilst decreasing the number of Type II errors. A lower threshold will increase the number of Type II errors whilst decreasing the number of Type I errors, it is hoped that by using the Bonferonni inequality to derive a threshold a happy trade off can be made, this is borne out empirically [18]. We ran the experiment on 8 dimensional data randomly generated so as to lie on a random hyperplane within a hypercube centred at the origin, side 100. The data was perturbed by Gaussian noise standard deviation 1.0. Outliers were introduced into the data to make a given percentage of the total between 5 and 50 percent, (in increments of 5 percent) these outliers could be totally at random within the region, thus they might accidently lie on the plane with a small probability. The experiment was repeated 200 times for each given level of contamination to generate the graphs shown in figure 2. The different case deletion diagnostics were compared with the removing the maximal residual heuristic (e.g. computing the regression, removing the point with maximal

[2] Tabulated in [21].

Figure 1: *Indicating a brief outline of the outlier detection algorithm.*

residual, and repeating this process until the maximum residual divided by its variance lies below our threshold). From these graphs it can be seen that the T_1 diagnostic gives the best discriminatory power, for this combination of n and p. The problem of detecting outliers becomes more difficult as the dimension p of our data increases and as the number n of points decreases. Consider that any three points can be fitted by a plane and given three points and one outlier it is impossible to determine which point is indeed the outlier. Thus the ratio $\frac{n}{p}$ must be quite high otherwise the algorithm will be doomed.

6 Application Finding the Fundamental Matrix

To show the algorithm in use in a real application we turn to the problem of finding the Fundamental matrix from point correspondences. Firstly we shall introduce the Fundamental matrix. It is a remarkable fact that, even when the motion and intrinsic calibration of a camera is not known, there exists a set of linear equations linking the images of world points over two views, providing they have all undergone a rigid transformation. This can expressed as follows: let the set of homogeneous image points $\{\vec{\mathbf{m}}_i\}, i = 1, \ldots n$, be transformed to the set $\{\vec{\mathbf{m}}'_i\}$ on the image plane, when the camera has undergone an unknown rotation about the origin followed by a translation. Then there exists a 3×3 matrix $[\mathbf{F}]$ such that

$$(x'_i, y'_i, 1)^T [\mathbf{F}](x_i, y_i, 1) = 0 \tag{17}$$

where $\vec{\mathbf{m}}_i = (x_i, y_i, 1)$ are the homogeneous coordinates of the ith feature in the first frame and $\vec{\mathbf{m}}'_i = (x'_i, y'_i, 1)$ are the inhomogeneous coordinates of the ith feature in the second frame, for all i. Each pair of homogeneous image points satisfies:

$$f_1 x'_i x_i + f_2 x'_i y_i + f_3 x'_i + f_4 y'_i x_i + f_5 y'_i y_i + f_6 y'_i + f_7 x_i + f_8 y_i + f_9 = 0 \tag{18}$$

where

$$[\mathbf{F}] = \begin{bmatrix} f_1 & f_2 & f_3 \\ f_4 & f_5 & f_6 \\ f_7 & f_8 & f_9 \end{bmatrix} \quad , \tag{19}$$

with constraint $|[\mathbf{F}]| = 0$. Points that are imaged from $3D$ objects with different motion parameters (modulo a collineation) will give rise to $2D$ points with distinct Fundamental matrices. Alternatively the Fundamental matrix may be thought of as a quadric surface in the space defined by (x, y, x', y'), thus we can apply a modification of Sampson's [13] iterative conic fitting method to find $[\mathbf{F}]$ and remove outliers.

Finally we note that it is possible to group points as belonging to the same object if they are consistent with a single Fundamental matrix (this forms the basis of our motion segmentation methods).

7 Results of deletion formulae on [F]

A brief outline of our algorithm is as follows:

1. A feature detector and matcher is applied to two images in order to get a set of correspondences. The methods used are described in [4].
2. A least sum of squares estimation is made of the Fundamental matrix [**F**] outliers are removed and the parameters re-estimated on the remaining data set, with each point reweighted by an estimate of its own standard deviation (which it can be shown is equal to the grad at that point: $\nabla[\mathbf{F}]$ as in Sampson's method for conic fitting). This step may be repeated until there are no more outliers to the fit.
3. Once we have eliminated outliers use the derived solution as the starting point for a non-linear minimization technique as in [10] that enforces $|[\mathbf{F}]| = 0$.

Results are shown in figure 3 on a sequence of images taken from a moving car on the PROMETHEUS project, kindly supplied by Pilkington. The images are 128×100 taken from an infra-red camera. The top image shows the corners detected by the corner detector described in [4], the matching procedure worked by using cross correlation in a square window. The middle image shows the data consistent with the fit with $\sigma = 0.5$, the bottom image shows the rejected outliers.

8 Robust Statistics

One approach to outlier identification are the regression diagnostics we have already described. These diagnostics are certain quantities computed from the data with the purpose of pinpointing influential points, after which the outliers can be removed or corrected, followed by a least squares analysis on the remaining cases. When there is only a single outlier or small number, the method we have presented works quite well by looking at the effect of deleting one point at a time. However, it is much more difficult to diagnose outliers when there are several of them, or when the outliers themselves possess structure. As will be the case when we solve for the Fundamental matrix and we do not have a dominant motion in the image but instead have two or more distinct motions each giving rise to a large number of data points. When there are many outliers then it is better to use a *robust regression* method, which tries to devise estimators that are not so strongly affected by outliers. There is no exact definition of the word robust, but in general is an attempt to reduce the sensitivity of that estimator to departures from the underlying parametric model e.g. non-normality for the case of least squares. We shall consider two similar algorithms where up to half of the data may be outlying and show how they can be applied to the problem of motion segmentation.

8.1 Random Sampling in LMS and RANSAC

Given that a large proportion of our data may be useless 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, we use as small a subset of the data as is feasible to estimate the parameters (e.g. two point subsets for a line). We repeat this process enough times to ensure that there is a 95% chance that one of the subsets will contain only good data points. We will identify the best solution as that which optimizes some criterion—minimizes the median squared residual in the case of LMS and maximizes the number of points whose residual is below a threshold in the case of RANSAC. Once we have removed outliers the set of points identified as non-outliers may be combined to give a final solution.

We now calculate how many subsamples we require, [6, 11] propose slightly different means of calculation but the number of samples is roughly the same and we give here the [11] method of calculation is used. Ideally we would like to consider every possible subsample, but this is usually computationally infeasible, thus we wish to choose m, the number of samples, sufficiently high to give a probability in excess of say 95% that we have included a good subsample within our selection. The expression for this probability Υ is

$$\Upsilon = 1 - (1 - (1 - \epsilon)^p)^m \tag{20}$$

where ϵ is the fraction of contaminated data. Table 4 gives some sample values of the number m of subsamples required to ensure $\Upsilon \geq 0.95$ for given p and ϵ. We shall now discuss the different methods of determining the best solution in LMS and RANSAC.

8.2 RANSAC

An early example of a robust algorithm is the random sample consensus paradigm (RANSAC) [6]. The algorithm randomly selects the minimum number of data to instantiate a model and estimate a set of parameters. The number of data points in the whole set that is within a threshold of the model is determined. If this number of points is greater than some other threshold the model is accepted as valid, or alternatively we repeat the algorithm for the number of times prescribed in table 4 and select the solution that gives the maximum amount of consistent data.

Given that we are calculating our parameter estimates by eigenvector methods we can calculate the variance of each residual (assuming that the solution is correct) as $\sigma_i^2 = \sigma^2(\nabla\vec{\mathbf{f}}_i)^2$ where $\nabla\vec{\mathbf{f}}_i$ is the gradient of $\vec{\mathbf{f}}$ with respect to $\vec{\mathbf{z}}_i$. We can then perform a Bonferonni test on each residual to see whether each point is outlying as shown in equation (21)

$$z_i = \begin{cases} \text{non outlier} & |\frac{e_i}{\sigma_i}| \leq \text{t-test} \\ \text{outlier} & \text{otherwise} \end{cases} . \tag{21}$$

8.3 The Least Median of Squares Estimator

Rousseeuw proposes taking as our parameter estimate the vector of parameters that minimizes the median squared residual over all our samples. Thus given a set $\{\vec{\boldsymbol{z}}\}$ of n, p-dimensional data points we take subsamples J of p points and use a singular value decomposition to recover the parameter estimate $\vec{\mathbf{f}}_J^*$. For each $\vec{\mathbf{f}}_J^*$ one also determines the corresponding median squared residual with respect to the whole set. Once a solution has been determined we can proceed in a similar manner to RANSAC to find the outliers.

8.4 Outline of the Robust Algorithms

A brief summary of the two random sampling algorithms is as follows:

1. Take a random sample of the minimum number of data points to make a parameter estimate $\vec{\mathbf{f}}^*$.
2. (a) In the case of the LMS estimator calculate the median residual over all the data.
   (b) In the case of the RANSAC estimator calculate the size of the data set consistent with $\vec{\mathbf{f}}^*$.
3. Repeat steps 1 and 2 for a set number of iterations, or until the solution is deemed sufficiently good.
4. Select the best solution i.e. that with the lowest median residual or the biggest consistent data set.
5. Re-estimate the parameters using all the data that has been identified as consistent, a different more computationally expensive estimator may be used at this point e.g. a non-linear minimization on the Euclidean distance.

Results of the algorithms can be seen in figure 5, the robust algorithms are compared to the T_1 case deletion diagnostic. It can be seen that the LMS algorithm marginally outperforms the RANSAC algorithm up to about 40% outliers after which it gives many more Type II errors. The LMS will fail totally when the number of outliers exceeds 50% as it is based on a median estimator. The T_1 diagnostic gradually declines in performance, after 15% outliers it gives many more Type I and II errors, but has the advantage that if there are only a few outliers then it is much more computationally efficient.

9 The Problem of Motion Segmentation

We now address the problem of how to detect a set of moving objects in the two dimensional projection of an otherwise rigid scene, given that the camera is moving in an arbitrary and unpredetermined manner—a problem as of yet unsolved in the field of robotic vision. Motion is a powerful cue for image and scene segmentation in the human visual system as evidenced by the ease with which we see otherwise perfectly camouflaged creatures as soon as they move, and by the strong cohesion perceived when even disparate parts of the image move in a way that could be interpreted in terms of a rigid motion in the scene. In robotic vision, motion segmentation turns out to be a most demanding problem in classification.

One approach to motion segmentation is to describe the motion field locally, recovering first and higher order derivatives of the motion field within small image patches to construct a Taylor series expansion. Adiv [1] and Waxman and Duncan [20] used this approach, making second order approximations to the flow field. Segmentation progressed by merging adjacent patches having similar expansions. Clearly, in order to determine the derivatives locally, such methods require dense and highly accurate determination of motion.

A second type of solution looks for discontinuities within the motion field. Again, however, this is done by comparison of field descriptions computed in small patches in the image: the use of too large patches would result in poor localization of motion boundaries. Examples of this approach can be seen in 2D in the work of Thompson, Mutch and Berzins [17] and Schunck [14] who presented algorithms for motion boundary detection based upon edge detection using the Laplacian of the Gaussian. A slightly different approach was taken by Murray and Williams [9] who exploited 3D scene constraints to detect motion boundaries. They assumed that the scene comprised planar facets and used the fact that the scene reconstruction algorithm was ill-conditioned with respect to noise, monitoring the condition number of the transformation matrix in small regions of the image. If the region straddled a boundary the condition number exhibited a maximum. Again, as both these solutions use small patches, they require dense, high quality motion information. In practice, such information is all but unobtainable from imagery of everyday scenes.

The problems inherent in local solutions suggest a move towards a global solution, where one tries to fit a single set of parameters across the entire image and then to exclude inconsistent data. The assumption that a major part of the image is consistent with a single set of parameters depends of course on the nature of the scene, but it is often reasonable to assume that most of the scene is moving coherently relative to the moving camera. It has been shown that, for any set of points in three space seen from two views, the epipolar geometry is captured by a 3×3 matrix, termed the *Fundamental Matrix* [8, 5, 12], which satisfies the condition, given in equation (17), for all pairs of corresponding homogeneous image points. We shall use the robust algorithms presented above to estimate Fundamental matrices for a scene and flag all outliers as potentially independently moving.

Results of the LMS segmentation are shown in figure 3 on an image taken from a sequence showing a robot translating 4 pixels to the left as the background appears to translate $0-6$ pixels down and $0-1$ pixels right depending on the depth. The image is 128×128. The corners were found by the corner detector described in [4], the matching procedure worked by using cross correlation in a square window. The top right image shows all the matches, the middle right the consistent data and the bottom shows the outliers from a full (i.e non-affine) Fundamental matrix fit with $\sigma = 0.7$, it can be seen the algorithm successfully segments the independently moving robot.

10 Conclusion

This paper has addressed several issues: we have presented two methodologies for outlier detection and shown how these can be used in motion analysis. One outlier detection method is based upon outlier diagnostics and standard least squares assumptions. The other is based upon robust statistics attempting to make a parameter estimate that is unaffected by outliers, rather than making a non-robust least squares estimation and looking for outliers to that. The diagnostics are characterised by some function of the residuals, the leverage and the degrees of freedom. The residuals give an indication of poorness of fit and the leverage gives an indication of remoteness [**Z**] space, thus any point with high leverage (eg far from the other points in the data cloud) may be viewed as outlying, even if it is consistent with our estimator. The interpretation of single case diagnostics is further complicated by the appearance of multiple outliers, which can act together to reinforce or cancel their group influence. In an attempt to compensate for this we remove the outliers detected on a given fit and refit to the remaining data. The main advantage of the diagnostic approach is its inherent speed, when there are few outliers, requiring much less computation time than the robust method. Overall the robust method is computationally more intensive but produces better results.

We have further shown how the robust methods may be used to solve the problem of motion segmentation. Which is the problem of how to detect a set of moving objects in the two dimensional projection of an otherwise rigid scene, given that the camera is moving in an arbitrary and unpredetermined manner. We utilise the fact that a transformation modulo a collineation in three space can be described by a 3×3 *Fundamental Matrix* [**F**] linking the coordinates of the points before and after the transformation in the image plane. Given that it highly unlikely that a scene will undergo a projective transformation we assume Euclidean $3D$ rigidity for the same real world imagery. The Fundamental Matrix is derived by the robust methods described in this paper, assuming at least half of the image is undergoing a coherent projectively rigid motion. Thus points with non-rigid motion are deemed to be those inconsistent from the calculated value of [**F**].

References

[1] G. Adiv. Inherent ambiguities in recovering 3-d motion and structure from a noisy flow field. In *Proceedings, CVPR '85 (IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, June 10–13, 1985)*, IEEE Publ. 85CH2145-1., pages 70–77. IEEE, IEEE, 1985.

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

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

[4] Droid. The droid 3d vision system. Technical Report 72/88/N488U, Plessey Research Roke Manor, 1988.

[5] O.D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In *Proceedings of 3rd European Conference on Computer Vision*, pages 563–578, 1992.

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

[7] G.H. Golub and C.F. Van Loan. *Matrix Computations* 2^{nd} *edition.* The John Hopkins University Press, 1989.

[8] R. Hartley. Estimation of relative camera positions for uncalibrated cameras. In *Proc. ECCV-92*, pages 579–87, 1992.

[9] D.W. Murray and N.S. Williams. Detecting the image boundaries between optical flow fields from several moving planar facets. *Pattern Recognition Letters*, 4:87–92, 1986.

[10] D. W. Murray P. Beardsley, A. Zisserman. Projective structure from image sequences. Technical Report To be published, Robotics Group, University of Oxford, 1993.

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

[12] A. Zisserman S. Demey and P. A. Beardsley. Affine and projective structure from motion. In D. Hogg, editor, *Proc. British Machine Vision Conference*, pages 49–58. Springer-Verlag, Sept 1992. Leeds.

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

[14] B.G. Schunck. Image flow segmentation and estimation by constraint line clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:1010–1027, 1989.

[15] L.S. Shapiro and J.M. Brady. Rejecting outliers and estimating errors in an orthogonal regression framework. Oxford Tech Report OUEL 1974/93, 1993.

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

[17] W.B. Thompson, K.M. Mutch, and V.A. Berzins. Dynamic occlusion analysis in optical flow fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7:374–383, 1985.

[18] P.H.S. Torr and D. W. Murray. Outlier detection and motion segmentation. Technical Report 1987/93, University of Oxford, 1993.

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

[20] A.M. Waxman and J.H. Duncan. Binocular image flows: Steps toward stereo-motion fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:715–729, 1986.

[21] S. Weisberg. *Applied Linear Regression.* Wiley, 1985.

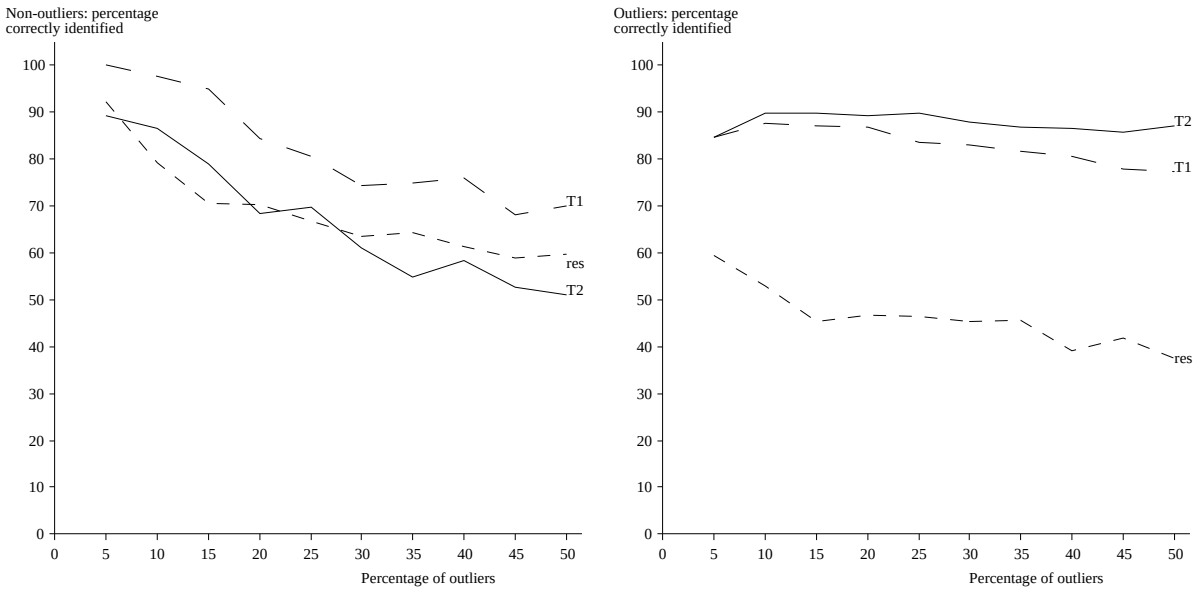

Figure 2: . *Giving a comparison of the percentage of outliers and non-outliers correctly identified for contaminations of the data in the range 0-50 percent. The algorithm was run on a 8 dimensional data set, size 100, the experiments were repeated 200 times. The long dashed line is the result for* T_{1i}*, the solid for* T_{2i}*, the short dashes for maximal residual deletion.*

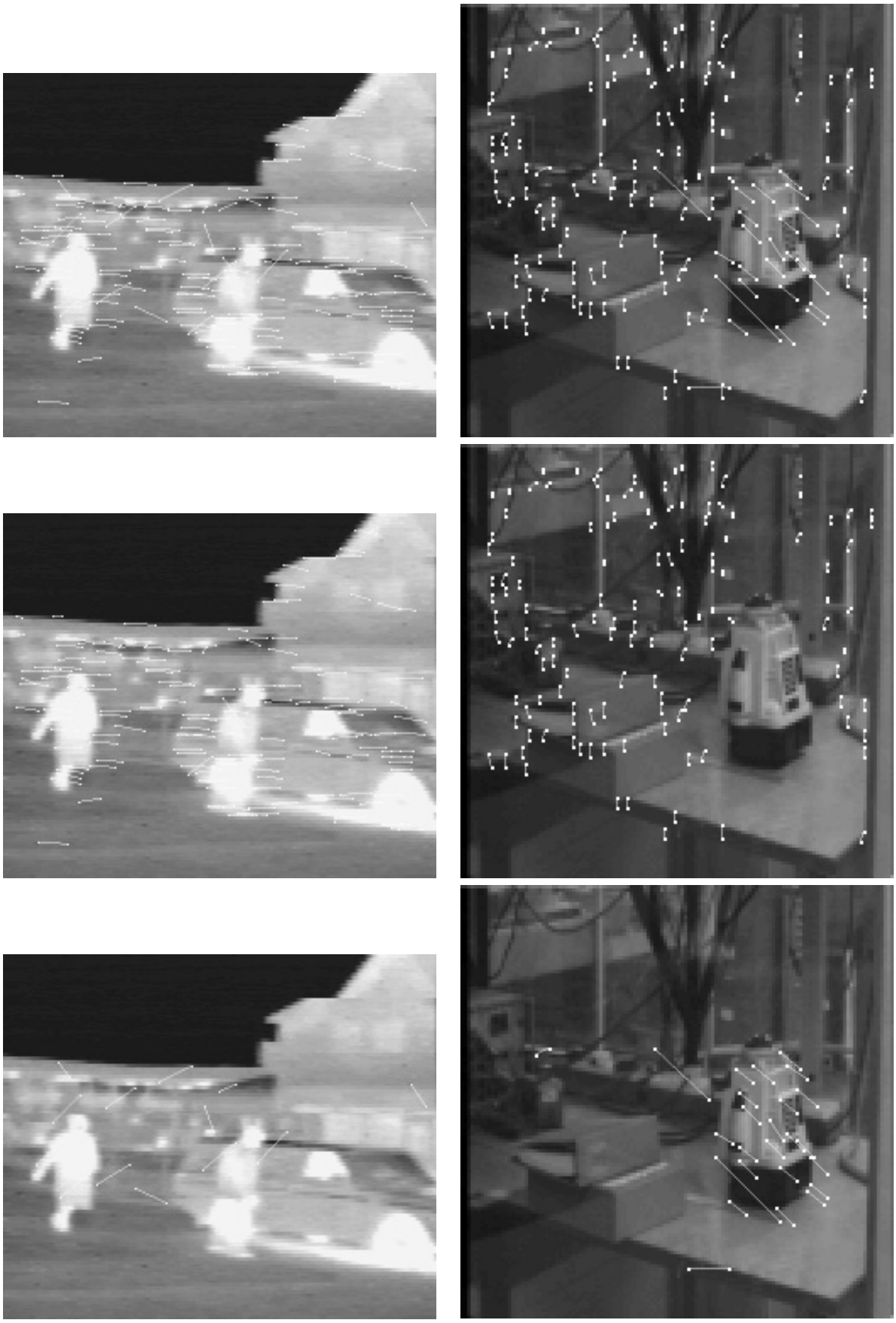

Figure 3: *The left set of images shows the results of the case deletion formulae. The right the results of segmentation using a robust estimator.*

Dimension	Fraction of Contaminated Data						
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

Figure 4: *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 we have selected in one of our subsamples are non-outliers. It can be seen from this that far from be computationally prohibitive, the robust algorithm may require less repetitions than there are outliers, as it is not directly linked to the number of outliers only the proportion of outliers. If the fraction of data that is contaminated is unknown, as it 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.*

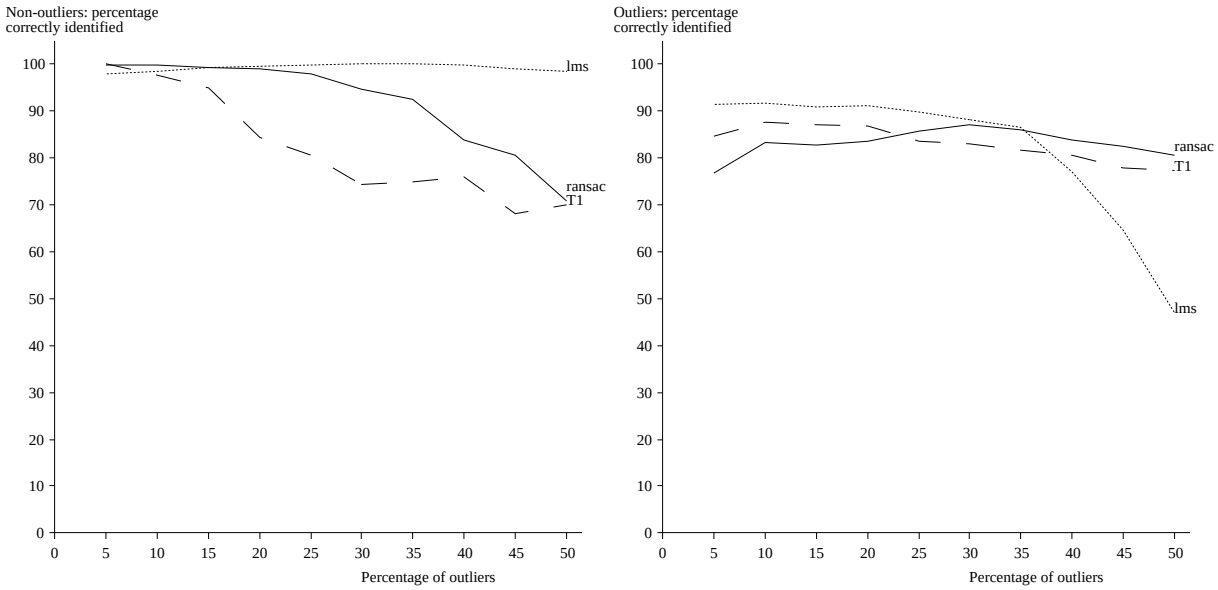

Figure 5: . *Giving a comparison of the percentage of outliers and non-outliers correctly identified for contaminations of the data in the range 0-50 percent. The algorithm was run on a 8 dimensional data set, size 100, the experiments were repeated 200 times. The dotted line is for the LMS algorithm, the solid for RANSAC and the dashed for the T_1 case deletion diagnostic.*