MLESAC: A new robust estimator with application to estimating image geometry

P. H. S. Torr[†] and A. Zisserman[‡]

[†] *Microsoft Research Ltd,*
St George House, 1 Guildhall St,
Cambridge CB2 3NH, UK
philtorr@microsoft.com
and [‡] *Robotics Research Group, Department of Engineering Science*
Oxford University, OX1 3PJ., UK
az@robots.ox.ac.uk

Received March 31, 1995; accepted July 15, 1996

A new method is presented for robustly estimating multiple view relations from point correspondences. The method comprises two parts, the first is a new robust estimator MLESAC which is a generalization of the RANSAC estimator. It adopts the same sampling strategy as RANSAC to generate putative solutions, but chooses the solution to maximize the likelihood rather than just the number of inliers. The second part to the algorithm is a general purpose method for automatically parametrizing these relations, using the output of MLESAC. A difficulty with multi view image relations is that there are often non-linear constraints between the parameters, making optimization a difficult task. The parametrization method overcomes the difficulty of non-linear constraints and conducts a constrained optimization. The method is general and its use is illustrated for the estimation of fundamental matrices, image-image homographies and quadratic transformations. Results are given for both synthetic and real images. It is demonstrated that the method gives results equal or superior to previous approaches.

1. INTRODUCTION

This paper describes a new robust estimator MLESAC which can be used in a wide variety of estimation tasks. In particular MLESAC is well suited to estimating complex surfaces or more general manifolds from point data. It is applied here to the estimation of several of the multiple view relations that exist between images related by rigid motions. These are relations between corresponding image points in two or more views and include for example, epipolar geometry, projectivities etc. These image relations are used for several purposes: (a) matching, (b) recovery of structure [1, 8, 11, 27, 40] (if this is possible), (c) motion segmentation [31, 36], (d) motion model selection [14, 37, 35].

The paper is organized as follows: In Section 2 the matrix representation of the two view relations are given, including the constraints that the matrix elements must satisfy. For example, there is a cubic polynomial constraint on the matrix elements for the fundamental

matrix. It will be seen that any parametrization must enforce this constraint to accurately capture the two view geometry.

Due to the frequent occurrence of mismatches, a RANSAC [4] like robust estimator is used to estimate the two view relation. The RANSAC algorithm is a hypothesis and verify algorithm. It proceeds by repeatedly generating solutions estimated from minimal set of correspondences gathered from the data, and then tests each solution for support from the complete set of putative correspondences. RANSAC is described in Section 4. In RANSAC the support is the *number* of correspondences with error below a given threshold. We propose a new estimator that takes as support the log likelihood of the solution (taking into account the distribution of outliers) and uses random sampling to maximize this. This log likelihood for each relation is derived in Section 3. The new robust random sampling method (dubbed MLESAC—Maximum Likelihood Estimation SAmple Consensus) is adumbrated in Section 5.

Having obtained a robust estimate using MLESAC, the minimum point set basis can be used to parametrize the constraint as described in Section 6. The MLE error is then minimized using this parametrization and a suitable non-linear minimizer. The optimization is constrained because the matrix elements of many of the two view relations must satisfy certain constraints. Note that relations computed from this minimal set always satisfy these constraints. Thus the new contribution is three fold: (a) to improve RANSAC by use of a better cost function; (b) to develop this cost function in terms of the likelihood of inliers and outliers (thus making it robust); and (c) to obtain a consistent parametrization in terms of a minimal point basis.

Results are presented on synthetic and real images in Section 7. RANSAC is compared to MLESAC, and the new point based parametrization is compared to other parametrizations that have been proposed which also enforce the constraints on the matrix elements.

Notation. The image of a 3D scene point $\mathbf{X}$ is $\mathbf{x}$ in the first view and $\mathbf{x}'$ in the second, where $\mathbf{x}$ and $\mathbf{x}'$ are homogeneous three vectors, $\mathbf{x} = (x, y, 1)^\top$. The correspondence $\mathbf{x} \leftrightarrow \mathbf{x}'$ will also be denoted as $\mathbf{x}^1 \leftrightarrow \mathbf{x}^2$. Throughout, underlining a symbol $\underline{x}$ indicates the perfect or noise-free quantity, distinguishing it from $x = \underline{x} + \Delta x$, which is the measured value corrupted by noise.

2. THE TWO VIEW RELATIONS

Within this section the possible relations on the motion of points between two views are summarized, three examples are considered in detail: (a) the Fundamental matrix [3, 10], (b) the planar projective transformation (projectivity), (c) the quadratic transformation. All these two view relations are estimable from image correspondences alone.

The epipolar constraint is represented by the Fundamental matrix [3, 10]. This relation applies for general motion and structure with uncalibrated cameras. 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, with the positions related by

$$\underline{\mathbf{x}}_i'^\top \mathbf{F} \underline{\mathbf{x}}_i = 0 \tag{1}$$

where $\underline{\mathbf{x}} = (\underline{x}, \underline{y}, 1)^\top$ is a homogeneous image coordinate and $\mathbf{F}$ is the Fundamental Matrix.

Should all the observed points lie on a plane, or the camera rotate about its optic axis and not translate, then all the correspondences lie on a projectivity:

$$\underline{\mathbf{x}}' = \mathbf{H}\underline{\mathbf{x}} \ . \tag{2}$$

Should all the points be consistent with two (or more) $\mathbf{F}$ then

$$\underline{\mathbf{x}}^{'\top}\mathbf{F}_1\underline{\mathbf{x}} = 0 \quad \text{and} \quad \underline{\mathbf{x}}^{'\top}\mathbf{F}_2\underline{\mathbf{x}} = 0 \tag{3}$$

thus

$$\underline{\mathbf{x}}^{'} = \mathbf{F}_1\underline{\mathbf{x}} \times \mathbf{F}_2\underline{\mathbf{x}} \tag{4}$$

hence they conform to a quadratic transformation. The quadratic transformation is a generalization of the homography. It is caused by a combination of a camera motion and scene structure, as all the scene points and the camera optic centres lie on a *critical surface* [19], which is a ruled quadric surface. Although the existence of the critical surface is well known, little research has been put into effectively estimating quadratic transformations.

2.1. Degrees of Freedom within Two View Parametrizations

The fundamental matrix has 9 elements, but only 7 degrees of freedom. Thus if the fundamental matrix is parametrized by the elements of the 3×3 matrix $\mathbf{F}$ it is over parametrized. This is because the matrix elements are not independent, being related by a cubic polynomial in the matrix elements, such that $\det[\mathbf{F}] = 0$. If this constraint is not imposed then the epipolar lines do not all intersect in a single epipole [16]. Hence it is essential that this constraint is imposed.

The projectivity has 9 elements and 8 degrees of freedom as these elements are only defined up to a scale. The quadratic transformation has 18 elements and 14 degrees of freedom [18]. Here if the constraints between the parameters are not enforced the estimation process becomes very unstable, and good results cannot be obtained [18], whereas our method has been able to accurately estimate the constraint.

2.2. Concatenated or Joint Image Space

Each pair of corresponding points $\underline{\mathbf{x}}$, $\underline{\mathbf{x}}^{'}$ defines a single point in a measurement space $\mathcal{R}^4$, formed by considering the coordinates in each image. This space is the 'joint image space' [38] or the 'concatenated image space' [24]. It might be considered somewhat eldritch to join the coordinates of the two images into the same space, but this makes sense if we assume that the data are perturbed by the same noise model (discussed in the next subsection) in each image, implying that the same distance measure for minimization may be used in each image. The image correspondences $\{\underline{\mathbf{x}}_i\} \leftrightarrow \{\underline{\mathbf{x}}_i^{'}\}, i = 1, \ldots n$, induced by a rigid motion have an associated algebraic variety V in $\mathcal{R}^4$. Fundamental matrices define a three dimensional variety in $\mathcal{R}^4$, whereas projectivities and quadratic transformations are only two dimensional.

Given a set of correspondences the (unbiased) minimum variance solution for $\mathbf{F}$ is that which minimizes the sum of squares of distances orthogonal to the variety from each point $(x, y, x^{'}, y^{'})$ in $\mathcal{R}^4$ [12, 14, 15, 21, 23, 26, 35]. This is directly equivalent to the reprojection error of the back projected 3D projective point.

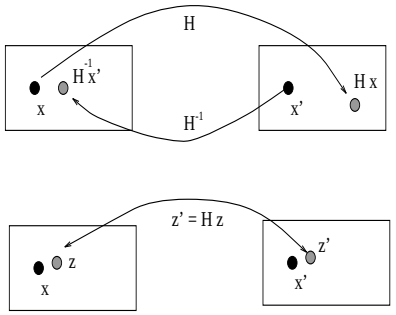


FIGURE 1

In previous work such as [17] the transfer error has often been used as the error function e.g. for fitting $\mathbf{H}$ this is

$$d^2\left(\mathbf{x}, \mathbf{H}^{-1}\mathbf{x}^{'}\right) + d^2\left(\mathbf{x}^{'}, \mathbf{H}\mathbf{x}\right) \tag{5}$$

where $d()$ is the Euclidean image distance between the points. The transfer distance is different from the orthogonal distance as shown in Figure 1. This is discussed further in relation to the maximum likelihood solution derived in Section 3.

3. MAXIMUM LIKELIHOOD ESTIMATION IN THE PRESENCE OF OUTLIERS

Within this section the maximum likelihood formulation is given for computing any of the multiple view relations. In the following we make the assumption, without loss of generality, that the noise in the two images is Gaussian on each image coordinate with zero mean and uniform standard deviation σ. Thus given a true correspondence the probability density function of the noise perturbed data is

$$\Pr(\mathbf{D}|\mathbf{M}) = \prod_{i=1\ldots n}\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n e^{-\left(\sum_{j=1,2}(\underline{x}_i^j - x_i^j)^2 + (\underline{y}_i^j - y_i^j)^2\right)/(2\underline{\sigma}^2)}, \tag{6}$$

where n is the number of correspondences and $\mathbf{M}$ is the appropriate 2 view relation, e.g. the fundamental matrix or projectivity, and $\mathbf{D}$ is the set of matches. The negative log likelihood of all the correspondences $\mathbf{x}_i^{1,2}$, $i = 1..n$:

$$-\sum_{i=1\ldots n}\log(\Pr(\mathbf{x}_i^{1,2}|\mathbf{M},\sigma)) = \sum_{i=1\ldots n}\sum_{j=1,2}\left((\underline{x}_i^j - x_i^j)^2 + (\underline{y}_i^j - y_i^j)^2\right), \tag{7}$$

discounting the constant term. Observing the data, we infer that the true relation $\mathbf{M}$ minimizes this log likelihood. This inference is called "Maximum Likelihood Estimation".

Given two views with associated relation for each correspondence $\mathbf{x}^{1,2}$ the task becomes that of finding the maximum likelihood estimate, $\hat{\mathbf{x}}^{1,2}$ of the true position $\underline{\mathbf{x}}^{1,2}$, such that $\hat{\mathbf{x}}^{1,2}$ satisfies the relation and minimizes $\sum_{j=1,2}\left(\hat{x}_i^j - x_i^j\right)^2 + \left(\hat{y}_i^j - y_i^j\right)^2$. The MLE

error e_i for the ith point is then

$$e_i^2 = \sum_{j=1,2} \left(\hat{x}_i^j - x_i^j\right)^2 + \left(\hat{y}_i^j - y_i^j\right)^2 \tag{8}$$

Thus $\sum_{i=1\ldots n} e_i^2$ provides the error function for the point data, and $\mathbf{M}$ for which $\sum_i e_i^2$ is a minimum is the maximum likelihood estimate of the relation (fundamental matrix, or projectivity). Hartley and Sturm [12] show how e, $\hat{\mathbf{x}}$ and $\hat{\mathbf{x}}'$ may be found as the solution of a degree 6 polynomial. A computationally efficient first order approximation to these is given in Torr *et al.* [32, 34, 35].

The above derivation assumes that the errors are Gaussian, often however features are mismatched and the error on $\mathbf{m}$ is not Gaussian. Thus the error is modeled as a mixture model of Gaussian and uniform distribution:-

$$\Pr(e) = \left(\gamma \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{e^2}{2\sigma^2}) + (1-\gamma)\frac{1}{v}\right) \tag{9}$$

where γ is the mixing parameter and v is just a constant (the diameter of the search window), σ is the standard deviation of the error on each coordinate. To correctly determine γ and v entails some knowledge of the outlier distribution; here it is assumed that the outlier distribution is uniform, with $-\frac{v}{2}..+\frac{v}{2}$ being the pixel range within which outliers are expected to fall (for feature matching this is dictated by the size of the search window for matches). Therefore the error minimized is the negative log likelihood:

$$-L = -\sum_i \log\left(\gamma\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\left(\sum_{j=1,2}(\underline{x}_i^j - x_i^j)^2 + (\underline{y}_i^j - y_i^j)^2\right)/(2\sigma^2) + (1-\gamma)\frac{1}{v}\right)\right) . \tag{10}$$

Given a suitable initial estimate there are several ways to estimate the parameters of the mixture model, most prominent being the EM algorithm [2, 20], but gradient descent methods could also be used. Because of the presence of outliers in the data the standard method of least squares estimation is often not suitable as an initial estimate, and it is better to use a robust estimate such as RANSAC which is described in the next section.

4. RANSAC

The aim is to be able to compute all these relations from image correspondences over two views. This computation requires initial matching of points (corners) over the image pairs. Corners are detected to sub-pixel accuracy using the Harris corner detector [9]. Given a corner at position (x, y) in the first image, the search for a match considers all corners within a region centred on (x, y) in the second image with a threshold on maximum disparity. The strength of candidate matches is measured by sum of squared differences in intensity. The threshold for match acceptance is deliberately conservative at this stage to minimise incorrect matches. Because the matching process is only based on proximity and similarity, mismatches will often occur. These are sufficient to render standard least squares estimators useless. Consequently robust methods must be adopted, which can provide a good estimate of the solution even if some of the data are mismatches (outliers).

Potentially there are a significant number of mismatches amongst the initial matches. Correct matches will obey the epipolar geometry. The aim then is to obtain a set of "inliers"

consistent with the epipolar geometry using a robust technique. In this case "outliers" are putative 'matches' inconsistent with the epipolar geometry. Robust estimation by random sampling (such as RANSAC) has proven the most successful [4, 30, 39].

First we describe the application of RANSAC to the estimation of the fundamental matrix. Putative fundamental matrices (up to three real solutions) are computed from random sets of seven corner correspondences (the minimum number required to compute a fundamental matrix). The fundamental matrices may be estimated from seven points by forming the data matrix:

$$\mathbf{Z} = \begin{bmatrix} x'_1x_1 & x'_1y_1 & x'_1 & y'_1x_1 & y'_1y_1 & y'_1 & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \\ x'_7x_7 & x'_7y_7 & x'_7 & y'_7x_7 & y'_7y_7 & y'_7 & x_7 & y_7 & 1 \end{bmatrix} . \tag{11}$$

The solution for $\mathbf{F}$ can be obtained from the two dimensional nullspace of $\mathbf{Z}$. Let $\mathbf{f}_1$ and $\mathbf{f}_2$ be obtained from the two right hand singular vectors of $\mathbf{Z}$ with singular values of zero, thus they form an orthogonal basis for the null space. Let $\mathbf{U}_1$ and $\mathbf{U}_2$ be the 3×3 matrices corresponding to $\mathbf{f}_1$ and $\mathbf{f}_2$. Then the three fundamental matrices $\mathbf{F}_l$, $l = 1, 2, 3$ consistent with $\mathbf{Z}$ can be obtained from $\mathbf{F}_l = \alpha\mathbf{U}_1 + (1-\alpha)\mathbf{U}_2$, subject to a scaling and the constraint $\det[\mathbf{F}] = 0$ (which gives a cubic in α from which 1 or 3 real solutions are obtained). The support for this fundamental matrix is determined by the number of correspondences in the initial match set with error e (given in (8)) below a threshold T. The error used is the negative log likelihood which is derived in the last section. If there are three solutions, then each is tested for support. This is repeated for many random sets, and the fundamental matrix with the largest support is accepted. The output is a set of corner correspondences consistent with the fundamental matrix, and a set of mismatches (outliers).

For projectivities each correspondence provides two constraints on the parameters:

$$\mathbf{v}_1^\top \mathbf{h} = 0 \quad \text{and} \quad \mathbf{v}_2^\top \mathbf{h} = 0 \tag{12}$$

where

$$\begin{aligned} \mathbf{v}_1 &= \begin{pmatrix} x & y & 1 & 0 & 0 & 0 & -xx' & -yx' & -x' \end{pmatrix} \\ \mathbf{v}_2 &= \begin{pmatrix} 0 & 0 & 0 & x & y & 1 & -xy' & -yy' & -y' \end{pmatrix} \end{aligned}$$

and $\mathbf{h}$ is the corresponding vector of the elements of $\mathbf{H}$. Thus four points may be used to find an exact solution. RANSAC proceeds in much the same manner, with minimal sets of four correspondences being randomly selected, and each set generating a putative projectivity. The support for each set is measured by calculating the negative log likelihood for all the points in the initial match set, and counting the number of correspondences with error below a certain threshold determined by consideration of the inlier and outlier distributions.

To estimate a quadratic transformation from seven correspondences the method used for generating fundamental matrices is modified. A critical surface is a ruled quadric passing through both camera centres. Seven correspondences define a quadric through the camera centres. If it is ruled then there will be three real fundamental matrices $\mathbf{F}_1$, $\mathbf{F}_2$ and $\mathbf{F}_3$ formed from the design matrix $\mathbf{Z}$ given in (11) of the seven points. These matrices can be used to generate the critical surface. In this case, any two of the fundamental matrices may

be combined to give the quadratic transformation by using Equation (4) (it does not matter which two as any pair gives the same result as any other pair). If only one solution is real then another sample can be taken.

How many samples should be used?. Ideally every possible subsample of the data would be considered, but this is usually computationally infeasible, so an important question is how many subsamples of the data set are required for statistical significance. Fischler and Bolles [4] and Rousseeuw and Leroy [22] proposed slightly different means of calculation, but each proposition gives 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

$$\Upsilon = 1 - (1 - (1 - \epsilon)^p)^m, \tag{13}$$

where ϵ is the fraction of contaminated data, and p the number of features in each sample. 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 allowing the algorithm to "jump out" of RANSAC 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%. Generally, assuming no more than 50% outliers then 500 random samples is more than sufficient.

5. THE ROBUST ESTIMATOR: MLESAC

The RANSAC algorithm has proven very successful for robust estimation, but having defined the robust negative log likelihood function $-L$ as the quantity to be minimized it becomes apparent that RANSAC can be improved on.

One of the problems with RANSAC is that if the threshold T for considering inliers is set too high then the robust estimate can be very poor. Consideration of RANSAC shows that in effect it finds the minimum of a cost function defined as

$$C = \sum_i \rho\left(e_i^2\right) \tag{14}$$

where $\rho()$ is

$$\rho(e^2) = \begin{cases} 0 & e^2 < T^2 \\ \text{constant} & e^2 \geq T^2 \end{cases}. \tag{15}$$

In other words inliers score nothing and each outlier scores a constant penalty. Thus the higher T^2 is the more solutions with equal values of C tending to poor estimation e.g. if T were sufficiently large then all solutions would have the same cost as all the matches would be inliers. In Torr and Zisserman [34] it was shown that at no extra cost this undesirable situation can be remedied. Rather than minimizing C a new cost function can be minimized

$$C_2 = \sum_i \rho_2\left(e_i^2\right) \tag{16}$$

where the robust error term ρ_2 is

$$\rho_2(e^2) = \begin{cases} e^2 & e^2 < T^2 \\ T^2 & e^2 \geq T^2 \end{cases}. \tag{17}$$

This is a simple, redescending M-estimator [13]. It can be seen that outliers are still given a fixed penalty but now inliers are scored on how well they fit the data. We set $T = 1.96\sigma$ so that Gaussian inliers are only incorrectly rejected five percent of the time. The implementation of this new method (dubbed MSAC *m-estimator sample consensus*) yields a modest to hefty benefit to all robust estimations with absolutely no additional computational burden. *Once this is understood there is no reason to use* RANSAC *in preference to this method.* Similar schemes for robust estimation using random sampling and M-estimators were also proposed in [29] and [25].

The definition of the maximum likelihood error allows us to suggest a further improvement over MSAC. As the aim is to minimise the negative log likelihood of the mixture $-L$ then it makes sense to use this as the score for each of the random samples. The problem is that the mixing parameter γ is not directly observed. But given any putative solution for the parameters of the model it is possible to recover γ that provides the minimum $-L$, as this is a one dimensional search it provides little computational overhead

To estimate γ, using Expectation Maximization (EM), a set of indicator variables needs to be introduced: η_i, $i = 1 \ldots n$, where $\eta_i = 1$ if the ith correspondence is an inlier, and $\eta_i = 0$ if the ith correspondence is an outlier. The EM algorithm proceeds as follows treating the η_i as missing data [5]: (1) generate a guess for γ, (2) estimate the expectation of the η_i from the current estimate of γ, (3) make a new estimate of γ from the current estimate of η_i and go to step (2). This procedure is repeated until convergence and typically requires only two or three iterations.

In more detail for stage (1) the initial estimate of γ is $\frac{1}{2}$. For stage (2) denote the expected value of η_i by z_i then it follows that $\Pr(\eta_i = 1|\gamma) = z_i$. Given an estimate of γ this can be estimated as:

$$\Pr(\eta_i = 1|\gamma) = \frac{p_i}{p_i + p_o} \tag{18}$$

and $\Pr(\eta_i = 0|\gamma) = 1 - z_i$. Here p_i is the likelihood of a datum given that it is an inlier:

$$p_i = \gamma \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^k \exp\left(-\left(\sum_{j=1,2} (\underline{x}_i^j - x_i^j)^2 + (\underline{y}_i^j - y_i^j)^2\right) / (2\sigma^2)\right) \tag{19}$$

and p_o is the likelihood of a datum given that it is an outlier:

$$p_o = (1 - \gamma)\frac{1}{v}. \tag{20}$$

For stage (3)

$$\gamma = \frac{1}{n}\sum_i z_i \ . \tag{21}$$

This method is dubbed MLESAC (maximum likelihood consensus). For real systems it is sometimes helpful to put a prior on γ, the expected proportion of inliers, this depends

1. Detect corner features using the Harris corner detector [9].

2. Putative matching of corners over the two images using proximity and cross correlation.

3. Repeat until 500 samples have been taken or "jump out" occurs as described in Section 4.

   (i) Select a random sample of the minimum number of correspondences $S_m = \{\mathbf{x}_i^{1,2}\}$.

   (ii) Estimate the image relation $\mathbf{M}$ consistent with this minimal set using the methods described in Section 4.

   (iii) Calculate the error e_i for each datum.

   (iv) For MSAC calculate C_2, or for, MLESAC calculate γ and hence $-L$ for the relation, as described in Section 5.

4. Select the best solution over all the samples i.e. that with lowest $-L$ or C_2. Store the set of correspondences S_m that gave this solution.

5. Minimize robust cost function over all correspondences, using the point basis provided by the last step as the parametrization, as described in Section 6.

]*A brief summary of all the stages of estimation*

TABLE 1

[

on the application and is not pursued further here. The two algorithms are summarized in Table 1. The output of MLESAC (as with RANSAC) is an initial estimate of the relation, together with a likelihood that each correspondences is consistent with the relation. The next step is to improve the estimate of the relation using a gradient descent method.

6. NON-LINEAR MINIMIZATION

The maximization of the likelihood is a constrained optimisation because a solution for $\mathbf{F}$, $\mathbf{Q}$ or $\mathbf{H}$ is sought that enforces the relations between the elements of the constraint. If a parametrization enforces these constraints it will be termed *consistent*. In the following we introduce a consistent parametrization and describe variations which result in a *minimal* parametrization. A minimal parametrization has the same number of parameters as the number of independent elements (degrees of freedom) of the constraint. The advantages and disadvantages of such minimal parametrizations will be discussed.

The key idea is to use the point basis provided by the robust estimator as the parametrization. For the simplest case, the projectivity, a four point basis is provided. By fixing x, y and varying x', y' for each correspondence, elements of the projectivity may be parametrized in terms of the 4 correspondences and a standard gradient descent algorithm [7] can be conducted with x', y' as parameters. Note this parametrization has exactly 8 DOF (2 variables for each of the 4 correspondences). Another approach is to alter all the 16 coordinates, the non-linear minimization conducted in this higher dimensional parameter space will discard extraneous parameters automatically. This approach has the disadvantage that it requires an increased number of function evaluations as there are more parameters than degrees of freedom. Similarly, 7 points may be used to encode the fundamental matrix and the parameters so encoded are guaranteed to be consistent, i.e. their elements satisfy the necessary constraints (Sometimes the 7 points may provide three solutions, in which case the one with lowest error is used). This method of parametrization in *term of points* was first proposed in Torr and Zisserman [33].

A number of variations on the free/fixed partition will now be discussed, as well as constraints on the direction of movement during the minimisation. In all cases the parametriza-

tion is consistent, but may not be minimal. Although a non-minimal parametrization over parametrizes the image constraint, the main detrimental effects is likely to be the cost of the numerical solution and poor convergence properties. The former is one of the measures used to compare the parametrizations in Section 7.

First the parametrizations for $\mathbf{F}$ are described. Given the minimal number of correspondences that can encode one of the image relations three coordinates can be fixed and one varied e.g. we could encode $\mathbf{F}$ by seven correspondences $(x_i, y_i) \leftrightarrow (x_i^{'}, y_i^{'})$ $i = 1 \ldots 7$, by fixing the $x, y, x^{'}$ coordinates of these correspondences the space of $\mathbf{F}$ is parametrized by the seven $y^{'}$ coordinates. This is referred to as parametrization **P1**. The parametrization **P1** for $\mathbf{H}$ and $\mathbf{Q}$ fixes x, y for the minimal basis set and varies $x^{'}$, $y^{'}$. This parametrization is both minimal and consistent, but the disadvantage for $\mathbf{F}$ of this is that should the epipolar lines in image 2 be parallel to the y axis then the movement of these points will not change $\mathbf{F}$. In order to overcome this disadvantage method **P2** moves coordinates in $\mathcal{R}^4$ in a direction orthogonal to the constraint surface (variety) defined by the image relation. The direction of motion is illustrated in Figure 2 for a two dimensional case. Here and for $\mathbf{F}$ there is one orthogonal direction to the manifold. In the cases of $\mathbf{H}$ and $\mathbf{Q}$, method **P2** moves each coordinate in $x, y, x^{'}, y^{'}$ in two directions orthogonal to the manifold. Perturbing each point in this space then has two degrees of freedom, so the parametrization has 8 dof in total (for $\mathbf{H}$), i.e. it is minimal. In fact, both **P1** and **P2** are minimal and consistent having 7 DOF for $\mathbf{F}$, 8 DOF for $\mathbf{H}$ and 14 DOF for $\mathbf{Q}$. A third method **P3** is now defined that uses all the coordinates of the correspondences encoding the constraint as parameters. Giving a 28 DOF parametrization for $\mathbf{F}$ and $\mathbf{Q}$, and 16 for $\mathbf{H}$. Note that **P3** is over parametrized having more parameters than degrees of freedom.

By way of comparison method **P4** is the linear method for each constraint and is used as a benchmark. Furthermore each constraint is estimated using standard parametrizations as follows. The projectivity and quadratic transformations are estimated by fixing one of the elements (the largest) of the matrix. For the projectivity this is minimal whereas it is not for the quadratic transform. The non-linear parametrization fixing the largest element is dubbed **P5**. **P6** is Luong's parametrization for the fundamental matrix. This is a 7 DOF parametrization in terms of the epipoles and epipolar homography designed by Luong *et al* [16], this is both minimal and consistent.

After applying MLESAC, the non-linear minimization is conducted using the method described in Gill and Murray [6], which is a modification of the Gauss-Newton method. All the points are included in the minimization, but the effect of outliers are removed as the robust function places a ceiling on the value of their errors, (thus they do not affect the Jacobian of the parameters), unless the parameters move during the iterated search to a value where that correspondence might be reclassified as an inlier. This scheme allows outliers to be re-classed as inliers during the minimization itself without incurring additional computational complexity. This has the advantage of reducing the number of false classifications, which might arise by classifying the correspondences at too early a stage.

An advantage of the method of Gill and Murray is that is does not require the calculation of any second order derivatives or Hessians. Furthermore if the data is over parametrized the algorithm has an effective strategy for discarding redundant combinations of the variables, and choosing efficient subsets of direction to search in parameter space. This makes it ideal for comparing minimizations conducted with different amounts of over parametrization.

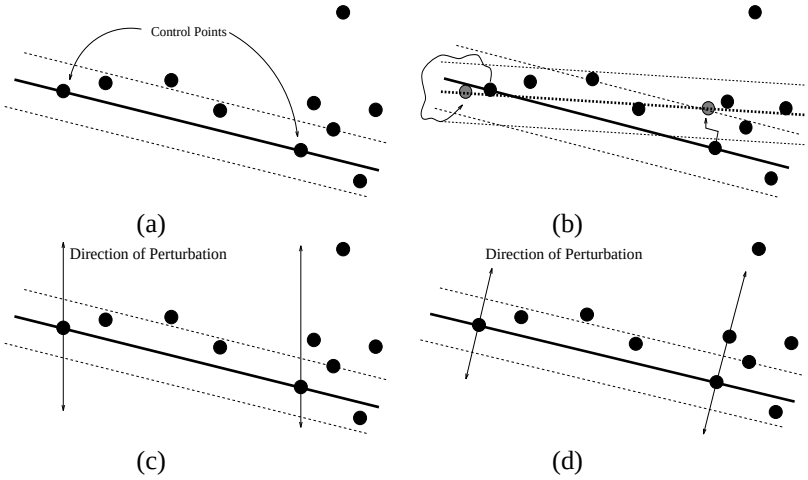


FIGURE 2

As a match may be incorrect, it is desirable that, if in the course of the estimation process we discover that the corner is mismatched, we are able to alter this match. In order to achieve this we store for each feature not only its match, but all its candidate matches that have a similarity score over a user defined threshold. After each estimation of the relation, in the iterative processes described above, features that are flagged as outliers are re-matched to their most likely candidate that minimizes the negative log likelihood.

Convergence problems might arise if either the chosen basis set is *exactly* degenerate, or the data as a whole are degenerate. In the first case the image relation $\mathbf{M}$ cannot be uniquely estimated from the basis set. To avoid this problem the rank of the design matrix $\mathbf{Z}$ matrix given by (11) can be examined. If the null space is greater than 2 (1 for $\mathbf{H}$), which it surely will be given degenerate data, then that particular basis can be discarded. Provided the basis points do not become *exactly* degenerate then any basis set is suitable for parametrizing $\mathbf{M}$.

In the second case, should the data as a whole be degenerate then the algorithm will fail to converge to a suitable result, the discussion of degeneracy is beyond the scope of this paper and is considered further in Torr *et al.* [35].

7. RESULTS

We have rigorously tested the various parametrizations on real and synthetic data. Two measures are compared: The first assesses the accuracy of the solution. The second measure is the number of cost function evaluations made i.e. the number of times D is evaluated. In the case of synthetic data the first measure is

$$\sigma_p = \left(\sum_{ij} \frac{d^2\left(\hat{\mathbf{x}}_i^j, \underline{\mathbf{x}}_i^j\right)}{n} \right)^{\frac{1}{2}} \tag{22}$$

for the set of inliers, where $\hat{\mathbf{x}}_i^j$ is the point closest to the noise free datum $\underline{\mathbf{x}}_i^j$ which satisfies the image relation, $\mathbf{x}_i^j$ is the ith point in the j image, and $d()$ is the Euclidean image distance

between the points. This provides a measure of how far the estimated relation is from the true data i.e. we test the fit of our computed relation from the noisy data against the *known* ground truth. In the case of real data the accuracy is assessed from the standard deviation of the inliers

$$\sigma_r = \left(\sum_i \frac{e_i^2}{n} \right)^{\frac{1}{2}} . \tag{23}$$

First experiments were made on synthetic data randomly generated in three space; 100 sets of 100 3D points were generated. The points were generated in the field of view 10-20 focal lengths from the camera. The image data was perturbed by Gaussian noise, standard deviation 1.0, and then quantized to the nearest 0.1 pixel. We then introduced mismatched features to make a given percentage of the total, between 10 and 50 percent. With synthetic data the estimate can be compared with the ground truth as follows: The standard deviation of the error of the *actual* noise free projections of the synthetic correspondences to the fitted relation is measured. This gives a good measure of the validity of each method in terms of the ground truth.

A comparison was made between the robust estimators looking at the standard deviation of the ground truth error before applying the gradient descent stage, for various percentages of outliers. The results were found to be dramatically improved: a reduction of variance from 1.43 to 0.64 when estimating a projectivity, suggesting that MLESAC should be adopted instead of RANSAC. After the non-linear stage the standard deviation of the ground truth error σ_p drops to 0.22. Figure 3 shows that estimated error on synthetic data (conforming to random fundamental matrices) for four random sampling style robust estimators: RANSAC [4], LMS [22, 39], MSAC and MLESAC. It can be seen that MSAC and MLESAC outperform the other two estimators, providing a $5 - 10\%$ improvement. This is because the first two have a more accurate assessment of fit, whereas LMS uses only the median, and RANSAC counts only the number of inliers. For this example the performance of MSAC and MLESAC are very close, MLESAC gives slightly better results but at the expense of more computation (the estimation of the mixing parameter γ for each putative solution). Thus the choice of MLESAC or MSAC for a particular applications depends on whether speed or accuracy is more important.

The initial estimate of the seven (for a fundamental matrix or critical surface) or four (for an image-image homography/projectivity) point basis provided by stage 2 is quite close to the true solution and consequently stage 3 typically avoids local minima. In general the non-linear minimisation requires far more function evaluations than the random sampling stage. However, the number required varies with parametrization, and is an additional measure (over variance) on which to assess the parametrization.

Fundamental matrix-synthetic data.. The five parametrizations for **F** were compared along with the linear method for fitting the Fundamental matrix. The results are summarised in Table 2. Luong's method **P6** produced a standard deviation of 0.32 with an average of 238 function evaluations in the non-linear stage. The 28 DOF parametrization **P3** in terms of points did significantly worse in the trials with an estimated standard deviation of 0.53 and an average of 2787 function evaluations, whereas the 7 DOF **P2** parametrization did significantly better with an estimated standard deviation of 0.22 at an average of 119 function evaluations. It remains yet to discover why the "orthogonal" parametrization provided the best results, but in general it seems that movement perpendicular to the

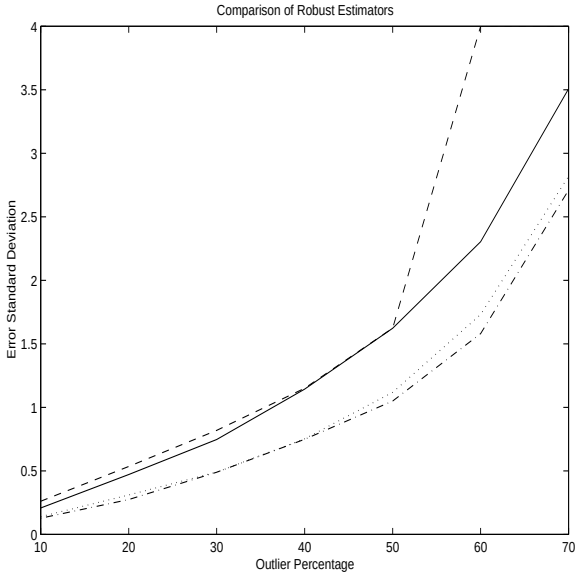


FIGURE 3

TABLE 2

The DOF, average number of evaluations of the total cost function in the gradient descent algorithm, and the standard deviation $\boldsymbol{\sigma_p}$ (22) for the perfect synthetic point data for the fundamental matrix.

Method	DOF	Evaluations	σ_p
P1 Vary y'	7	120	0.34
P2 Orthogonal Perturbation	7	119	0.22
P3 Vary x, y, x', y'	28	2787	0.53
P4 Linear	-	-	0.85
P5 Fix largest element of $\mathbf{F}$	7	260	0.54
P6 Luong	7	238	0.32

constraint of the point produces speedy convergence to a good solution, perhaps because it moves the parameters in the direction that will change the constraint the most.

Fundamental matrix-Chapel Sequence. Figures 4 (a)-(b) show two views of an outdoor chapel, the camera moves around the chapel rotating to keep it in view. The matches from the initial cross correlation are shown in (c) and it can be seen that they contain a fair number of outliers. The basis provided by MLESAC is given in (d). As the minimization progresses the basis points move only a few hundredths of a pixel each, but the solution was much improved, final inliers and outliers are shown in (e) and (f); the standard deviation of the inlying data σ_r have decreased from 0.67 to 0.23.

Projectivity-synthetic data.. When fitting the projectivity the several different parametrizations **P1-P3, P5** were used. In this case, the results of all parametrizations were almost identical, with standard deviations of the error within 0.04 pixels of each other. The number of function evaluations required in the non-linear minimization was on average 124 for the 8 DOF orthogonal parametrization, compared with 457 for the 16 DOF. Hence the 8

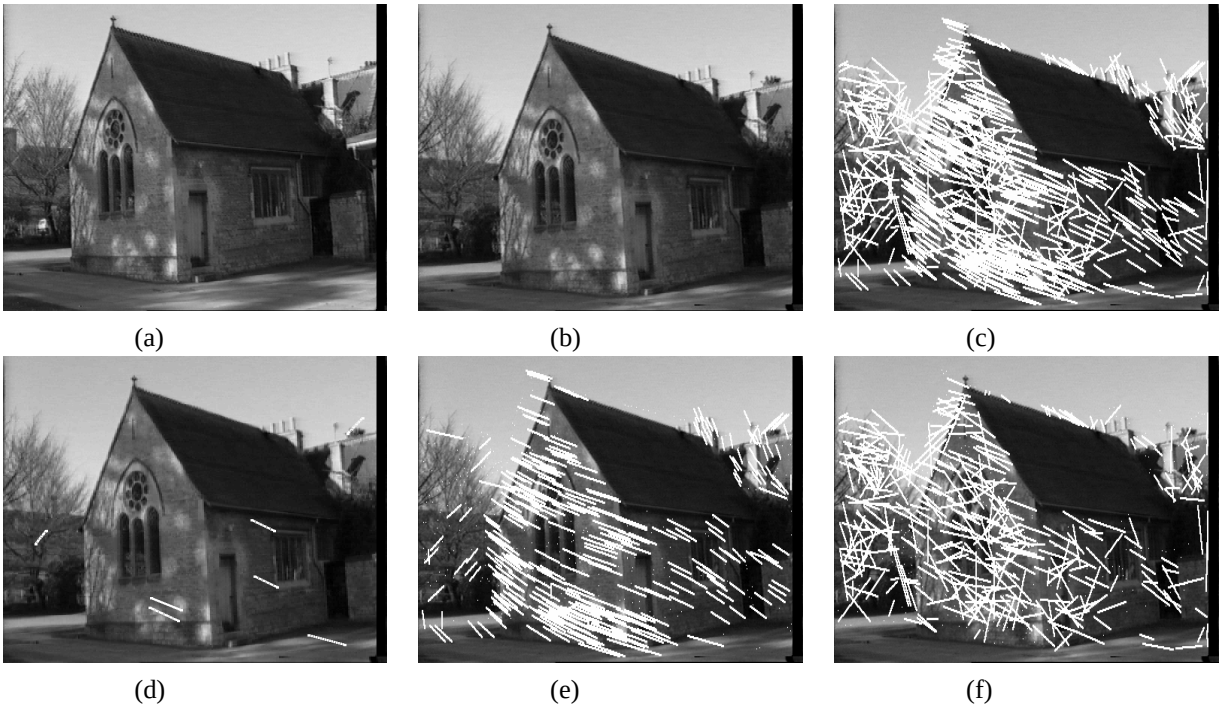


FIGURE 4

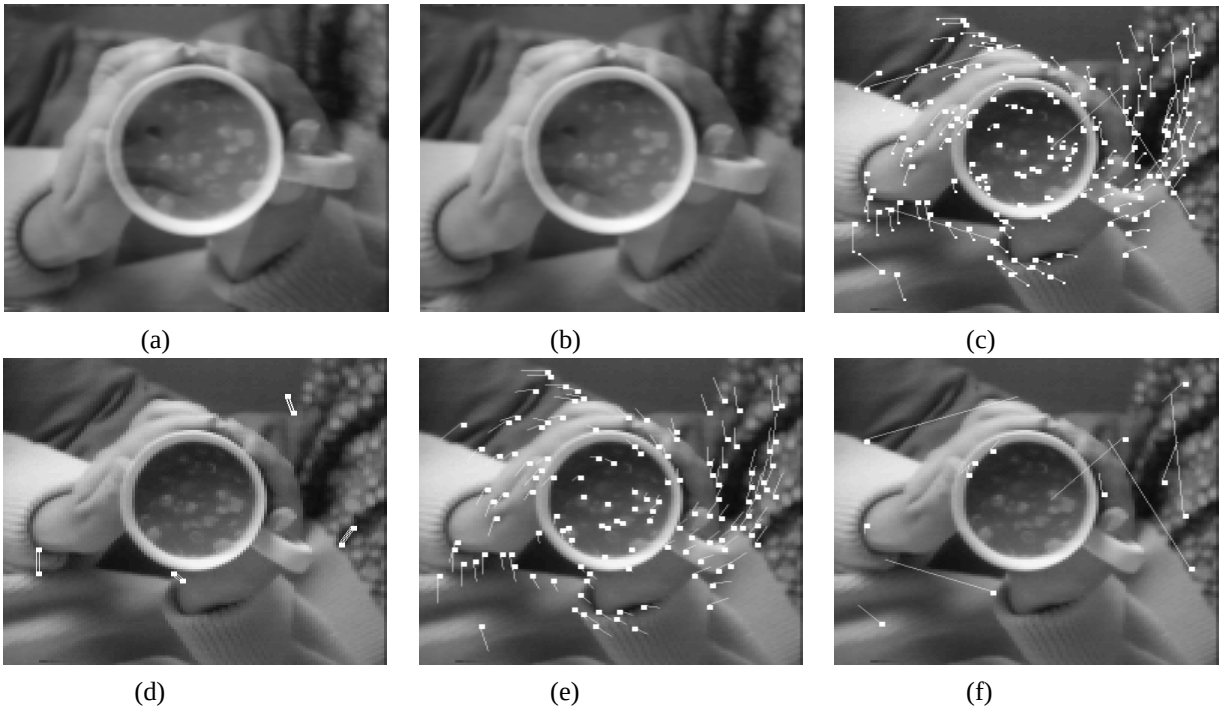
TABLE 3

The DOF, average number of evaluations of the total cost function in the gradient descent algorithm, and the standard deviation $\boldsymbol{\sigma_p}$ (22) for the perfect synthetic point data for the homography matrix.

Method	DOF	Evaluations	σ_p
P1 Vary x' and y'	8	132	0.34
P2 Orthogonal Perturbation	8	124	0.30
P3 Vary x, y, x', y'	16	457	0.31
P4 Linear	-	-	0.42
P5 Fix largest element of $\mathbf{H}$	8	260	0.32

DOF has the slight advantage of being somewhat faster. The lack of difference between the parametrizations might be explained by the lack of complex constraints between the elements of the homography matrix, which is defined up to a scaling by 9 elements.

Projectivity-Cup Data.. Figure 5 (a) shows the first and (b) the second image of a cup viewed from a camera undergoing cyclotorsion about its optic axis combined with an image zoom. The matches are given in (c), basis (d), inliers (e) outliers (f) for this scene when fitting a projectivity. It can be seen that outliers to the cyclotorsion are clearly identified. The non linear step does not produce any new inliers as the MLESAC step has successfully eliminated all mismatches, the error on the inliers is reduced by 1% when the image coordinates in one image are fixed and those in the other varied.



(a) (b) (c)

(d) (e) (f)

FIGURE 5

TABLE 4

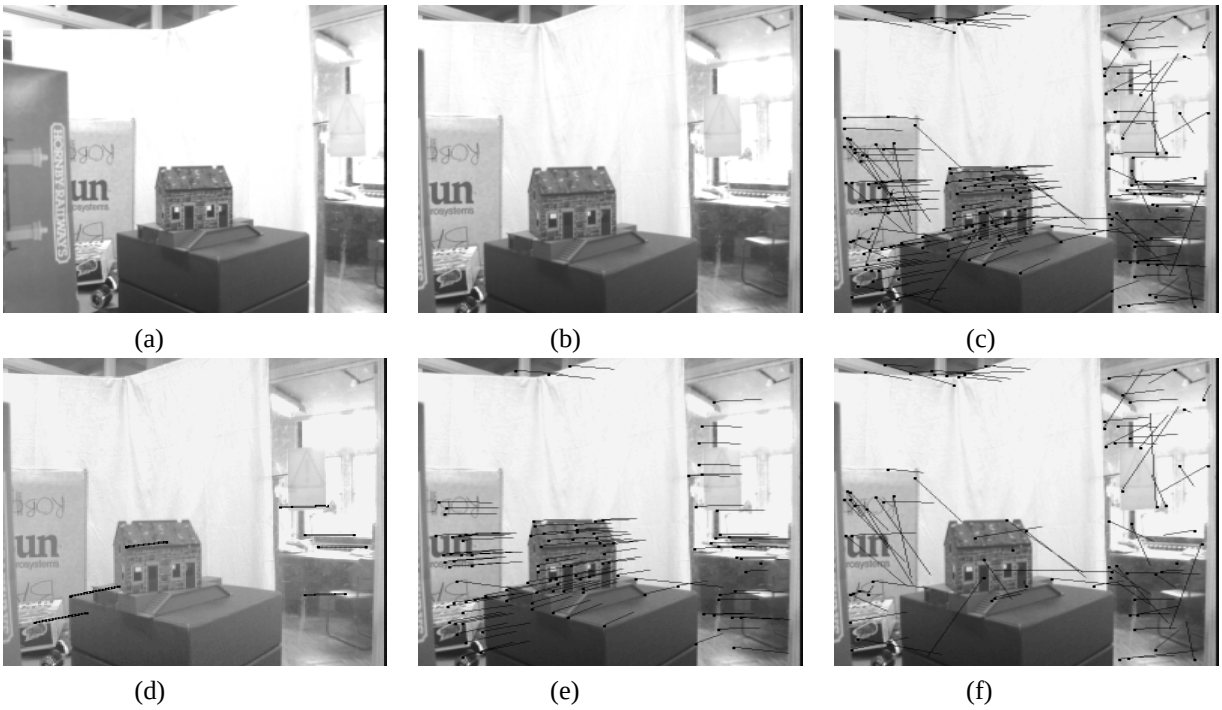
The DOF, average number of evaluations of the total cost function in the gradient descent algorithm, and the standard deviation $\boldsymbol{\sigma_p}$ (22) for the perfect synthetic point data for the quadratic transformation.

Method	DOF	Evaluations	σ_p
P1 Vary $x^{'}$ and $y^{'}$	14	791	0.36
P2 Orthogonal Perturbation	14	693	0.30
P3 Vary x, y, $x^{'}$, $y^{'}$	28	3404	0.57
P4 Linear	-	-	0.64
P5 Fix largest element of $\mathbf{H}$	17	913	0.39

Quadratic Transformation-synthetic data.. The results for the synthetic tests are summarised in Table 4, it can be seen that the orthogonal perturbation parametrization **P2** gives the best results, and outperforms the inconsistent parametrization **P5** as well as the over parametrization **P3**.

Quadratic Transformation-Model house data.. Figure 6 (a) (b) shows a scene in which a camera rotates and translates whilst fixating on a model house. The standard deviation of the inliers improves from 0.39 after the estimation by MLESAC to 0.35 after the non-linear minimization. The important thing about this image is that structure recovery for this image pair proved highly unstable. The reason for this instability is not immediately apparent until the good fit of the quadratic transformation is witnessed, indicating that structure cannot

be well recovered from this scene. In fact the detected corners approximately lie near a quadric surface which also passes through the camera centres. This is shown in Figure 7.



(a) (b) (c)

(d) (e) (f)

FIGURE 6

8. CONCLUSION

Within this paper an improvement over RANSAC: MLESAC has been shown to give better estimates on our test data. A general method for constrained parameter estimation has been demonstrated, and it has been shown to provide equal or superior results to existing methods. In fact few such general purpose methods exist to estimate and parametrize complex quantities such as critical surfaces. The general method (of minimal parametrization in terms of basis points found from MLESAC) could be used for other estimation problems in vision, for instance estimating the Quadrifocal tensor between four images, complex polynomial curves etc. The general methodology could be used outside of vision in any problem where minimal parametrizations are not immediately obvious, and the relations may be determined from some minimal number of points.

Why does the point parametrization work so well? One reason is that the minimal point set initially selected by MLESAC is known to provide a good estimate of the image relation (because there is a lot of support for this solution). Hence the initial estimate of the point basis provided by MLESAC is quite close to the true solution and consequently the non-linear minimisation typically avoids local minima. Secondly the parametrization is consistent which means that during the gradient descent phase only image relations that might actually arise are searched for.

FIGURE 7

It has been observed that the MLESAC method of robust fitting is good for initializing the parameter estimation when the data are corrupted by outliers. In this case there are just two class to which a datum might belong, inliers or outliers. The MLESAC method may be generalized to the case when the data has arisen from a more general mixture model involving several classes, such as in clustering problems. Preliminary work to illustrate this has been conducted [28].

There are two extensions that are trivial to MLESAC that allow the introduction of prior information[1], although prior information is not used in this paper. The first is to allow a prior $\Pr(\mathbf{M})$ on the parameters of the relation, which is just added to the score function $-L$. The second is to allow a prior on the number of outliers and the mixing parameter γ.

Acknowledgements

Thank you to the reviewers for helpful suggestions, and Richard Hartley for helpful discussions.

REFERENCES

1. P. Beardsley, P. H. S. Torr, and A. Zisserman. 3D model aquisition from extended image sequences. In B. Buxton and Cipolla R., editors, *Proc. 4th European Conference on Computer Vision, LNCS 1065, Cambridge*, pages 683–695. Springer–Verlag, 1996.
2. A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the em algorithm. *J. R. Statist. Soc.*, 39 B:1–38, 1977.
3. O.D. Faugeras. 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, Santa Margherita Ligure*, pages 563–578. Springer–Verlag, 1992.

[1] In this case the name MLESAC is still appropriate as the aim is still to maximize the likelihood, now it is the posterior likelihood

4. M. Fischler and R. 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.
5. A. Gelman, J. Carlin, H. Stern, and D. Rubin. *Bayesian Data Analysis.* Chapman and Hall, 1995.
6. P. E. Gill and W. Murray. Algorithms for the solution of the nonlinear least-squares problem. *SIAM J Num Anal*, 15(5):977–992, 1978.
7. Numerical Algorithms Group. *NAG Fortran Library vol 7*. NAG, 1988.
8. C. Harris. The DROID 3D vision system. Technical Report 72/88/N488U, Plessey Research, Roke Manor, 1988.
9. C. Harris and M. Stephens. A combined corner and edge detector. In *Proc. Alvey Conf.*, pages 189–192, 1987.
10. R. I. Hartley. 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, 1992.
11. R. I. Hartley. Euclidean reconstruction from uncalibrated views. In *Proc. 2nd European-US Workshop on Invariance, Azores*, pages 187–202, 1993.
12. R. I. Hartley and P. Sturm. Triangulation. In *DARPA Image Understanding Workshop, Monterey, CA*, pages 957–966, 1994.
13. P. J. Huber. Projection pursuit. *Annals of Statistics*, 13:433–475, 1985.
14. K. Kanatani. *Statistical Optimization for Geometric Computation: Theory and Practice.* Elsevier Science, Amsterdam, 1996.
15. M. Kendall and A. Stuart. *The Advanced Theory of Statistics.* Charles Griffin and Company, London, 1983.
16. Q. T. Luong, R. Deriche, O. D. Faugeras, and T. Papadopoulo. On determining the fundamental matrix: analysis of different methods and experimental results. Technical Report 1894, INRIA (Sophia Antipolis), 1993.
17. Q. T. Luong and O. D. Faugeras. Determining the fundamental matrix with planes: Instability and new algorithms. *CVPR*, 4:489–494, 1993.
18. Q. T. Luong and O. D. Faugeras. A stability analysis for the fundamental matrix. In J. O. Eklundh, editor, *Proc. 3rd European Conference on Computer Vision, LNCS 800/801, Stockholm*, pages 577–586. Springer–Verlag, 1994.
19. S.J. Maybank. Properties of essential matrices. *Int. J. of Imaging Systems and Technology*, 2:380–384, 1990.
20. G.I. McLachlan and K. Basford. *Mixture models: inference and applications to clustering.* Marcel Dekker. New York, 1988.
21. V. Pratt. Direct least squares fitting of algebraic surfaces. *Computer Graphics*, 21(4):145–152, 1987.
22. P. J. Rousseeuw. *Robust Regression and Outlier Detection.* Wiley, New York, 1987.
23. P.D. Sampson. Fitting conic sections to 'very scattered' data: An iterative refinement of the Bookstein algorithm. *Computer Vision, Graphics, and Image Processing*, 18:97–108, 1982.
24. L. S. Shapiro. *Affine Analysis of Image Sequences.* PhD thesis, Oxford University, 1993.
25. C. V. Stewart. Bias in robust estimation caused by discontinuities and multiple structures. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol.PAMI-19,no.8:818–833, 1997.
26. G. Taubin. 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, 1991.
27. C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorisation approach. *International Journal of Computer Vision*, 9(2):137–154, 1992.
28. P. Torr, R. Szeliski, and P. Anandan. An integrated bayesian approach to layer extraction from image sequences. In *Seventh International Conference on Computer Vision*, volume 2, pages 983–991, 1999.
29. P. H. S. Torr, P. A. Beardsley, and D. W. Murray. Robust vision. In J. Illingworth, editor, *Proc. 5th British Machine Vision Conference, York*, pages 145–155. BMVA Press, 1994.
30. P. H. S. Torr and D. W. Murray. Outlier detection and motion segmentation. In P. S. Schenker, editor, *Sensor Fusion VI*, pages 432–443. SPIE volume 2059, 1993. Boston.
31. P. H. S. Torr and D. W. Murray. Stochastic motion clustering. In J.-O. Eklundh, editor, *Proc. 3rd European Conference on Computer Vision, LNCS 800/801, Stockholm*, pages 328–338. Springer–Verlag, 1994.
32. P. H. S. Torr and D. W. Murray. The development and comparison of robust methods for estimating the fundamental matrix. *Int Journal of Computer Vision*, 24(3):271–300, 1997.

33. P. H. S. Torr and A Zisserman. Robust parameterization and computation of the trifocal tensor. *Image and Vision Computing*, 15:591–607, 1997.

34. P. H. S. Torr and A. Zisserman. Robust computation and parametrization of multiple view relations. In U Desai, editor, *ICCV6*, pages 727–732. Narosa Publishing House, 1998.

35. P. H. S. Torr, A Zisserman, and S. Maybank. Robust detection of degenerate configurations for the fundamental matrix. *CVIU*, 71(3):312–333, 1998.

36. P. H. S. Torr, A. Zisserman, and D. W. Murray. 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, 1995.

37. P.H.S. Torr. An assessment of information criteria for motion model selection. In *CVPR97*, pages 47–53, 1997.

38. W. Triggs. The geometry of projective reconstruction i: Matching constraints and the joint image. In *Proc. 5th Int'l Conf. on Computer Vision, Boston*, pages 338–343, 1995.

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

40. Z. Zhang and O. Faugeras. *3D Dynamic Scene Analysis*. Springer-Verlag, 1992.