Robust Parametrization and Computation of the Trifocal Tensor

P H S Torr, A Zisserman
Robotics Research Group
Department of Engineering Science
Oxford University, Parks Road, Oxford, OX1 3PJ, UK.
Email: *phst,az@robots.oxford.ac.uk*

Abstract

This paper presents an algorithm for computing a Maximum Likelihood Estimate (MLE) of the trifocal tensor. The input to the algorithm is three images of the same scene, and the output is the estimated tensor and corner and line feature matches across the three images that are consistent with this estimate.

Particular novelties of the algorithm are the computation of a trifocal tensor from six point correspondences, and a parametrization of the trifocal tensor which enforces the constraints between the tensor elements. The algorithm uses techniques from robust statistics and is fully automatic.

Results are presented for synthetic and real image triplets. The proposed parametrization is compared to other existing methods.

Keywords: Trifocal tensor, robust estimation, matching, maximum likelihood estimation.

1 Introduction

The trifocal tensor plays a similar role for three views to that played by the fundamental matrix for two. It encapsulates all the (projective) geometric constraints between three views that are independent of scene structure. The tensor only depends on the motion between views and the internal parameters of the cameras, but it can be computed from image correspondences alone without requiring knowledge of the motion or calibration. It is the culmination of developments by a number of researchers including [7, 12, 21, 22, 30, 31, 33, 34]. Recently the trifocal tensor has been used for applications in structure from motion including tracking [2], camera calibration [1], and motion segmentation [29].

Given correspondences for points in two images, the trifocal tensor determines the position of the point in the third (this is known as *transfer*). Similarly, given correspondences for

lines in two images, the same tensor determines the position of the line in the third image. Unlike point epipolar transfer [8, 36], transfer based on the trifocal tensor does not fail for 3D points lying on, or close to, the trifocal plane (the plane defined by the three optical centres of the cameras), or when the three optical centres are collinear.

This paper describes an algorithm for computing a Maximum Likelihood Estimate (MLE) of the trifocal tensor, together with corner and line matches over three images which are consistent with this estimate. The paper extends the state of the art in three ways. First, a robust estimator is developed for the trifocal tensor based on six point correspondences over three views. Six is the minimum number of point correspondences from which the tensor can be computed for 3D points in general position. Second, an error measure is defined such that the computed trifocal tensor is the maximum likelihood estimate, and an efficient first order approximation to this error is given. Third, a simple parametrization is described which enforces the algebraic constraints existing between the elements of the tensor.

The structure of the paper is as follows. In Section 2 there is a review of the trifocal tensor, including a subsection concerning the number degrees of freedom of the tensor, which sets the scene for the six point computation and the parametrization issues. The algorithm to compute the trifocal tensor from three images proceeds in three stages. First, corners and line segments are detected in each image and putative correspondences determined over the three images. The second stage uses RANSAC and minimal sets to remove mismatches (outliers) from this initial set of putative correspondences. Both of these stages are described in Section 3, together with a discussion of why using the minimal six point correspondences is important. The output of the second stage is an estimate of the trifocal tensor based on six points which has maximum support amongst both the putative corner and line correspondences.

The third stage, described in Section 4, is the maximum likelihood estimate of the tensor. This is achieved by optimising a cost function (error measure). The numerical optimisation requires a parametrization of the tensor which enforces the relations between the tensor elements. Section 5 describes this parametrization which is based on the minimal point set provided by RANSAC. The relation to other parametrizations proposed for the tensor by [13, 17, 29] is discussed. The algorithm is summarised and implementation details are given in Section 6. Section 7 compares and assesses the various parametrizations, giving results on both real and synthetic data.

Notation A 3D scene point projects to $\mathbf{x}$, $\mathbf{x}^2$ and $\mathbf{x}^3$ in the first, second and third images respectively, and similarly the image of a line is $\mathbf{l}$, $\mathbf{l}^2$ and $\mathbf{l}^3$. Where $\mathbf{x} = (x_1, x_2, x_3)^\top$ and $\mathbf{l} = (l_1, l_2, l_3)^\top$ are homogeneous three vectors. The correspondence $\mathbf{x} \leftrightarrow \mathbf{x}^2 \leftrightarrow \mathbf{x}^3$ will also be represented by the shorthand notation $\mathbf{x}^{1,2,3}$. The 3×4 camera projection matrix which relates a 3D point $\mathbf{w}$ to its image $\mathbf{x}$ is denoted as $\mathbf{P}$, i.e. $\mathbf{x} = \mathbf{P}\mathbf{w}$, where $\mathbf{w}$ is an homogeneous four vector.

2 Trifocal tensor - review

Corresponding points in three images, and corresponding lines in three images, satisfy trilinear relations which are encapsulated in the trifocal tensor.

The trifocal tensor, $\mathbf{T}$, is a $3 \times 3 \times 3$ homogeneous tensor. Using the tensor a point can be transferred to a third image from correspondences in the first and second:

$$x_l^3 = x_i^2 \sum_{k=1}^{k=3} x_k T_{kjl} - x_j^2 \sum_{k=1}^{k=3} x_k T_{kil} \quad , \tag{1}$$

for all $i, j = 1 \ldots 3$—nine expressions called trilinearities, four of which are *linearly* independent. Similarly, a line can be transferred as

$$l_i = \sum_{j=1}^{j=3} \sum_{k=1}^{k=3} l_j^2 l_k^3 T_{ijk} \tag{2}$$

there are three such equations relating the line correspondences, two of which are independent. It can be seen that the same tensor can be used to transfer both points and lines.

If the first camera matrix is chosen as $\mathbf{P} = [\mathbf{I}|\mathbf{0}]$ (where $\mathbf{I}$ is the 3×3 identity matrix, and $\mathbf{0}$ is a null three-vector), then the trifocal tensor can be computed from:

$$T_{ijk} = P_{ji}^2 P_{k4}^3 - P_{j4}^2 P_{ki}^3 \ , \tag{3}$$

where P_{ij}^w is the ijth element of the camera matrix of the wth image.

2.1 Degrees of Freedom

The trifocal tensor has 27 elements, but only their ratios are significant, leaving 26 coefficients that must be specified. Each triplet of point correspondences provides four independent linear equations for the elements of the tensor, and each triplet of line correspondences provides two linear equations. Therefore provided that $2n_l + 4n_p \geq 26$ (where n_l is the number of lines, and n_p is the number of points), the tensor can be determined (up to scale) using a linear algorithm. The tensor can be computed linearly from a minimum of 7 points or 13 lines or a combination of the two. As in the case of the fundamental matrix [5, 25], the tensor can be estimated from more than the minimum number of correspondences by minimising a cost function e.g. a solution may be computed by eigenvector methods finding the eigenvector with least eigenvalue of a 27×27 matrix.

However, the tensor has only 18 independent elements, and consequently only 18 independent degrees of freedom. This can be seen by considering three 3×4 projection matrices, less 15 projective degrees of freedom, i.e. $3 \times 11 - 15 = 18$. Consequently the 27 elements of the tensor satisfy a number of (polynomial) constraints. This is similar to the situation with the fundamental matrix—there are 9 elements in the 3×3 matrix, but only 7 degrees of freedom (again this follows from $2 \times 11 - 15 = 7$). In the case of the fundamental matrix only the ratio of the elements is significant and there is one cubic constraint that the determinant is zero. For the trifocal tensor the 8 constraints have been investigated [7, 17] but are not as yet thoroughly understood.

In the case of the fundamental matrix if the constraint is not imposed then the epipolar lines do not all intersect in a single epipole [18]. Similarly, if the constraints are not imposed

on the trifocal tensor elements, not only may the epipoles not be defined but the various linear methods of transfer in equation (1) will give different results for the transferred point position [17].

It might be thought that since the tensor has only 18 independent elements it could be computed from five point correspondences, together with the polynomial constraints discussed above, since each point correspondence contributes four linearly independent equations and $5 \times 4 > 18$. This is not the case. As is shown in Appendix A, six point correspondences are required for 3D points in general position.

3 Robust Computation of the tri-focal tensor

This section describes the first and second stage of the algorithm where first a set of putative corner and line correspondences are assembled, and then a robust method is used to remove mismatches and provide an initial estimate of the tensor. This is equivalent to the automatic algorithms developed for computing the fundamental matrix from two images in Torr and Murray [26] and Zhang *et. al.* [35]. The third stage of the algorithm, described in Section 4 uses this initial robust estimate of the tensor as a starting point for the MLE.

In the first stage corners and line segments are extracted independently in each image. Corners are detected to sub-pixel accuracy using the Harris corner detector [11]. Lines are detected by the standard procedure of: Canny edge detection [3]; edge linking; segmentation of the chain at high curvature points; and finally, straight line fitting to the resulting chain segments. The straight line fitting is by orthogonal regression, with a tight tolerance to ensure that only actual line segments are extracted, i.e. that curves are not piecewise linear approximated.

Putative correspondences are then obtained over the three images by a combination of similarity of features and proximity. For example, for a corner at position (x_1, x_2) in the first image, the search for a match considers all corners within a region centred on (x_1, x_2) in the second image with a threshold on maximum disparity. The strength of candidate matches is measured by cross-correlation. The threshold for match acceptance is deliberately conservative at this stage to minimise incorrect matches. Similarly cross correlation is used to generate initial matches between the second and third images. Full details of this matching, and the similar approach for line segments, are given in [2].

The second stage of the algorithm obtains an estimate of the trifocal tensor based on these correspondences. 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 tensor even if some of the data are mismatches (outliers). An early example of a robust algorithm is the random sample consensus paradigm (RANSAC) of Fischler and Bolles [9]. Given that a large proportion the data may be outlying, the approach is the opposite to conventional smoothing techniques. Rather than using as much data as is possible to obtain an initial solution and then attempting to identify outliers, as small a subset of the data as is feasible to estimate the solution is used (e.g. two point subsets for a line, seven correspondences for a fundamental matrix) here six point correspondences for a trifocal

Repeat for $m = 500$ samplings:

1. Select a random sample of the minimum number of six corner correspondences to estimate the trifocal tensor T. This provides 1 or 3 solutions.
2. For each of these solutions:
   (a) Calculate the error d_i for the ith point correspondence.
   (b) Calculate the error e_j for the jth line correspondence.
   (c) Calculate the total number of inliers over the two sets (i.e errors below a user specified threshold).
3. Select the best solution over all the samples i.e. that with the highest number of inliers.

Table 1: *A brief summary of the second stage of the algorithm, the random sampling. The errors d_i and e_j are defined in Section 4*

tensor, and this process is repeated enough times on different subsets to ensure that there is a 95% chance that one of the subsets will contain only good data points. The best solution is that which maximizes the number of feature (corner and line) correspondences whose error is below a threshold. This second stage of the algorithm is summarized in Table 1.

The method for finding the trifocal tensor from six points uses the theory of Quan [20] for computing an invariant of six points from 3 views, and is described in Appendix A. The method involves the solution of a cubic, and correspondingly provides one or three real solutions for the trifocal tensor. The fact that three solutions might arise from six points does not impose a difficulty to the algorithm, since the best solution of the three is selected when measuring support for each solution from the full set of putative matches.

For random sampling methods, such as RANSAC it is important that the minimum number of correspondences are used, so as to reduce the probability of a mismatch being included in the random sample of correspondences. The number of samples required, and hence the computation required, rises exponentially with the number of data points in the sample [9]. Thus the novel six point solution used here is markedly faster than the seven point method presented in Torr *et al* [29] to achieve the same level of accuracy. Furthermore for computational efficiency we are discouraged from using lines to initialize an estimate since many more lines are needed than points to minimally estimate the tensor. A second reason why the 6 point method is better than the 7 point method is that the constraints are imposed between the elements of the tensor; whereas this is not true for the linear 7 point estimation technique. The issue of constraints is explored further in Section 5.

For the RANSAC algorithm a decision needs to be made as to how many samples should be taken. As pointed out by Fischler and Bolles [9] this should be dependent on the expected number of outliers within the data, and they provide a suitable formula for deciding the number of samples. Assuming no more than 50% outliers then 500 random samples is more

than sufficient.

4 Maximum Likelihood Estimation

Image features (points and lines) correspond over the three images because they are the image of the same 3D entity. The *true* or noise free image features will exactly satisfy the appropriate constraint (1) and (2). However, the observed (measured) image features which arise from these true image features, will not exactly satisfy these relations. The question then arises as to how to best estimate the trifocal tensor given the measured correspondences. In the following subsection we make the assumption that the observed features have been perturbed by a Gaussian noise process. We are then able to develop a maximum likelihood solution for both the tensor and the correspondences. This solution involves minimising a cost function which is a summation over all correspondences of two error measures, one for points and one for line data.

Notation is used to distinguish the three point cases. First there is the *true* point correspondence $\underline{\mathbf{x}} \leftrightarrow \underline{\mathbf{x}}^2 \leftrightarrow \underline{\mathbf{x}}^3$, where noise free data are indicated by underlining; second, the *measured* point correspondence $\mathbf{x} \leftrightarrow \mathbf{x}^2 \leftrightarrow \mathbf{x}^3$; and third the *estimated* point correspondence $\hat{\mathbf{x}} \leftrightarrow \hat{\mathbf{x}}^2 \leftrightarrow \hat{\mathbf{x}}^3$; which is the maximum likelihood of the *true* correspondence given the *measured*. A similar notation is used for lines.

There have been several previous algorithms to estimate the rigidity constraint over three views each proposing a different error metric. Spetsakis and Aloimonos [22], working with a calibrated equivalent of the trifocal tensor, propose a noise model on the 3D features; minimizing distances in 3D. This is not appropriate here for two reason, firstly the reconstruction is projective hence distance in 3D is meaningless. Secondly the noise introduced into the location of the features is an artifact of the feature detectors and hence is essentially image based. Hartley [13] used an overdetermined linear fit for the trifocal tensor, using eigenvector methods without minimising a statistically meaningful quantity. Neither of these approaches enforce the constraints between the tensor elements, a point we return to in Section 5.

However, both Hartley [14] when estimating the rigidity constraint (fundamental matrix) from point correspondences over two images, and Weng *et. al.* [34] when estimating the rigidity constraint from line correspondences over three images consider a cost function which is equivalent to MLE given Gaussian noise, but the formulation is not made explicit.

4.1 MLE for Point Data

In the following for simplicity it is assumed and without loss of generality that the noise in all three images is Gaussian on each image coordinate with zero mean and uniform standard deviation σ. Thus given a true correspondence $\underline{\mathbf{x}}_i \leftrightarrow \underline{\mathbf{x}}_i^2 \leftrightarrow \underline{\mathbf{x}}_i^3$ the probability density function of the noise perturbed data is

$$\Pr(\mathbf{x}_i^{1,2,3}|\mathbf{T}) = \prod_j \frac{1}{\sqrt{2\pi}\sigma} e^{-\left(\underline{x}_i^j - x_i^j\right)^2/(2\sigma^2)} \quad . \tag{4}$$

The log likelihood of all the correspondences $\mathbf{x}_i^{1,2,3}$, $i = 1..n$ where n is the number of correspondences is:

$$\sum_i \log(\Pr(\mathbf{x}_i^{1,2,3}|\mathbf{T})) = \sum_{ij} \left(\underline{x}_i^j - x_i^j\right)^2 + \text{constant} \ .$$

Then the *maximum likelihood estimation* (MLE) of the trifocal tensor $\mathbf{T}$ and true correspondences $\underline{\mathbf{x}}^{1,2,3}$ minimizes this log likelihood, i.e. minimizes

$$D_M = \sum_{ij} \left(\hat{x}_i^j - x_i^j\right)^2 . \tag{5}$$

where $\hat{\mathbf{x}}^{1,2,3}$ are the MLE of the true correspondences, and $\hat{\mathbf{x}}^{1,2,3}$ satisfies (1).

Each point correspondence contributes a term

$$d_i^2 = \sum_j \left(\hat{x}_i^j - x_i^j\right)^2 \tag{6}$$

to the cost function $D_M = \sum_i d_i^2$. Note d^2 is a χ^2 variable with three degrees of freedom. Appendix B describes a computationally efficient first order Taylor approximation which is used during the numerical minimisation of the cost function: Given the correspondence $\mathbf{x}^{1,2,3}$ and $\mathbf{T}$ this approximation provides both $\hat{\mathbf{x}}^{1,2,3}$ and d.

The derivation here is actually for a minimum variance estimator, and so the noise model is slightly more general than just a Gaussian error. It also applicable to any class of error whose log likelihood takes the form given in Equation (6).

4.2 Error for Line Data

It is possible to perform a similar maximum likelihood analysis for lines to that performed for points in the previous section, and this is presented in Appendix C. However at the current time this has not been fully tested and the following heuristic measure has been used for the results presented in this paper: Given the trifocal constraint and the location of a line in two images its location in the third image may be predicted. From empirical tests the best results have been obtained by the minimization of the root mean square of the distances of the end points of the observed line to the predicted line divided by the length of the line, over all three images. This error measure, e is similar to the heuristic measures used in [13, 34]. The standard deviation σ_l of e may be estimated at run time.

4.3 The Total Cost Function

The point and line errors may be combined by summing the squared errors each divided by its standard deviations for points and lines:

$$D_T = \sum_i \frac{d_i^2}{\sigma^2} + \sum_k \frac{{e_k}^2}{\sigma_l^2} \tag{7}$$

providing that there are no outliers. For simplicity this assumes homogeneous variance amongst the coordinates of the point and line; the extension to non-homogeneous variance is given in Appendix C.

However, to take account of outliers, a robust version of this cost function is employed

$$D = \sum_i \gamma\left(\frac{d_i}{\sigma}\right) + \sum_k \gamma\left(\frac{e_k}{\sigma_l}\right) \tag{8}$$

where d and e are point and line errors and $\gamma(x)$ is a robust Huber function [15]:

$$\gamma(x) = \begin{cases} x^2 & x < 1.96 \\ 1.96^2 & x \geq 1.96 \ . \end{cases} \tag{9}$$

The value 1.96 corresponds to the 95% confidence level. This means that an inlier will only be incorrectly rejected (a Type II error) 5% of the time.

The advantage of this Huber cost function (9) it that it allows the minimization to be conducted over all correspondences whether they are outliers or inliers. It might be thought that this could be achieved by simply thresholding the error and not including correspondences with an error above this threshold. But thresholding alone would result in only outlier correspondences being included because they would incur no cost. Hence the Huber cost function applies a fixed cost for each outlier (equivalent in the maximum likelihood formulation to considering the outliers to be from a uniform distribution).

The starting point for the minimisation is the estimate of the trifocal tensor provided by RANSAC. In the following section the parametrization of the tensor used in the minimisation is described.

5 Parameterization of the trifocal tensor

The minimisation of the cost function (8) is a constrained optimisation because a solution for $\mathbf{T}$ is sought which enforces the relations between the elements of the tensor. 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 (degrees of freedom) as the number of independent elements of the tensor, i.e. 18. The advantages and disadvantages of such minimal parametrizations will be discussed.

In Section 3 six points were associated with the best solution provided by RANSAC for the fundamental matrix. The coordinates of these six points provide a natural parametrization which has the advantage that it is always consistent i.e. any choice of these coordinates generates one or three trifocal tensors whose elements satisfy the necessary constraints. Since six point correspondences have 36 degrees of freedom (two for each point in each of three images) this is not a minimal parametrization. However, it can provide a minimal parametrization by partitioning the point coordinates into a fixed set and a set of free

variables. For example, if x^1, y^1, x^2, say, are fixed for each of the 6 correspondences and y^2, x^3, y^3 are free, then we have a minimal and consistent parametrization.

The optimisation of the MLE cost function then proceeds by varying the free coordinates, using a gradient descent scheme described in Section 6, to obtain the minimum. Although, the six point correspondences for the parametrization initially correspond to real point correspondences, this is not necessary, and after the minimisation the free variables will not correspond in general to their initial (real) positions. It is best to think of the points as *virtual* points, parametrizing the tensor.

5.1 Variations on the six-point parametrization

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 parametrization is consistent, but may not be minimal. Although a non-minimal parametrization over parametrizes the tensor, 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.

There are some disadvantages with the minimal parametrization described above where the free variables are y^2, x^3, y^3 (hereafter referred to as method **M0**). Suppose that the estimated epipolar lines in the second image were parallel to the y-axis, then fixing x^2 and altering y^2 for each basis point would not change the estimated relative motion between image 1 and 2; i.e. there would be a whole range of trifocal tensors that could never be reached by the search of the parameter space. To overcome this problem four variations on the approach are considered. Figure 1 illustrates two of these variations for a two dimensional example.

The first variation is to fix x^1 and have the other 30 as free. This method will be referred to as **M1**. The second focusses on whether improvements can be obtained by fixing other combinations of point coordinates. For example, instead of fixing x^1, y^1, x^2 and varying y^2, x^3, y^3, fixing x^1, x^2, y^2 or other combinations differing over the points. Such a scheme was tried by first minimizing over x^1, y^1, x^2 with y^2, x^3, y^3 fixed and then minimizing over x^1, y^1, y^2 with x^2, x^3, y^3 fixed. This method will be referred to as **M2**. This method only ever has 18 DOF at any one time but all the coordinates are given a shake at some point. The third, **M3** fixes the coordinates in the third image and perturbs those in the other two, giving a 24 degree of freedom parametrization.

Finally method **M4** moves coordinates in $\mathcal{R}^6$ in a direction orthogonal to the constraint surface (variety) defined by the trifocal tensor. This variety of dimension 3 in the $\mathcal{R}^6$ space of coordinates is discussed in Appendix B. The direction of motion is illustrated in Figure 1d for a two dimensional case. In the trifocal case there is a 3-dimensional space of directions perpendicular to the variety. Perturbing each point in this space then has three degrees of freedom, so the parametrization has 18 dof in total, i.e. it is minimal.

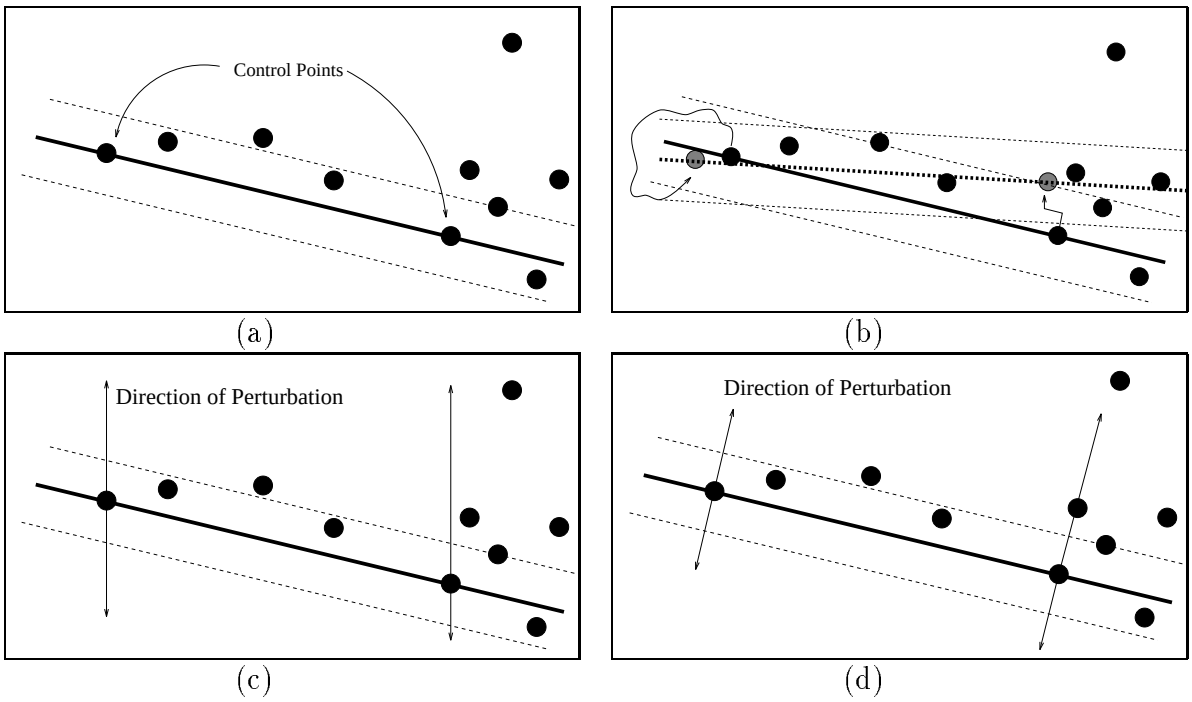

Figure 1: *A 2D example of the parametrization: a line fit is parametrized by two points. (a) The outcome of RANSAC fitting a line to a set of points, the bold line is that fitted, the thin dashed lines the thresholds for determining whether points are inlying. Initially there are five inliers and five outliers. (b) The result after the gradient descent stage, moving the two control points to optimise the fit. Finally there are seven inliers and three outliers. The location of the control points are shown in grey. (c) Perturbation in a fixed direction, this is equivalent to fixing some coordinates and then varying others, as in method* **M1**. *(d) Perturbation in a direction orthogonal to the line, comparable to method* **M4**.

5.2 Previous Parameterizations

Now some previous parametrizations of the trifocal tensor are discussed, and in Section 7 they are compared with the 6 point parametrizations for estimating the trifocal tensor. Existing consistent parametrizations generally proceed by initially computing the three 3×4 projection matrices, and computing the tensor from these [17]. In the following section the 6 point parametrization, with varying numbers of free variables, is compared to:

1. Method **M5**: A 24 DOF parametrization
   based on $\mathbf{P}^1 = [\mathbf{I}|\ \mathbf{0}]$ for the first projection matrix, and two 3×4 projection matrices, $\mathbf{P}^2, \mathbf{P}^3$ for the other two images, each with 12 elements. This parametrization is consistent but not minimal.

2. Method **M6**: A 26 DOF parametrization
   using the 26 coefficients of the trifocal tensor as the parameters with one element to unity (the largest element is chosen given a prior estimate from RANSAC). This parametrization is *not* consistent and not minimal.

Finally as a bench mark to all the methods Hartley's [12] linear method will be tested on the data sets and dubbed **M7**.

6 Algorithm Summary and Implementation Details

A brief summary of the algorithm is as follows:

1. Putative matching of points and lines over the three images using proximity and similarity measures.

2. Robust random sampling methods based on minimal (six) point sets provide an initial classification into inliers and outliers (mismatches). This also provides a six point basis to determine the trifocal tensor, used in the next stage.

3. A non linear gradient descent minimization of the Huber cost function, using the six point basis provided by the last step as the initial solution and parametrization.

The non-linear minimization is conducted using the method described in Gill and Murray [10], which is a modification of the Gauss-Newton method. 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 are over parametrized the algorithm has an effective strategy for discarding redundant combinations of the variables. This makes it ideal for comparing minimizations conducted with differing amounts of over parametrization (see the next section).

All the points and lines are included in the minimization, but the effect of outliers are removed as the Huber function places a ceiling on the value of their errors, (thus they do not effect 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, which has the advantage of reducing the number of false classifications, which might arise by classifying the correspondences at too early a stage. Typically, as the minimization progresses borderline outliers are redesignated inliers.

Additionally, if it is discovered that the feature (point or line) 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 trifocal tensor, in the iterative processes described above, features that are flagged as outliers are re-matched to their most likely candidate that minimizes d (for points) or e (for lines).

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 trifocal tensor cannot be uniquely estimated from the six points and the algorithm given in Appendix A will arbitrarily select a solution. To avoid this problem the rank of the matrix given by (16) can be examined. If this rank is less than 3, 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 the tensor.

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.* [28].

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

$$\sigma_p = \left(\sum_{ij} \frac{\left(\hat{\underline{x}}_i^j - \underline{x}_i^j \right)^2}{n} \right)^{\frac{1}{2}} \tag{10}$$

for the set of inliers, where $\hat{\underline{x}}_i^j$ is the point closest to the noise free datum $\underline{x}_i^j$ which satisfies the trilinearities (1) for the *estimated* tensor. This provides a measure of how far the estimated tensor is from the true data. In the case of real data the accuracy is assessed from the the standard deviation of the inliers

$$\sigma_r = \left(\sum_{i} \frac{d_i^2}{n} \right)^{\frac{1}{2}} . \tag{11}$$

The first test compared the 7 point and 6 point methods using RANSAC to generate the trifocal tensor, i.e. just stage 1 and 2 of the algorithm. The 7 point method finds the

Method	DOF	Evaluations	σ_p
M0 Six point fix x^1, y^1, x^2	18	593	0.25
M1 Six point fix x^3	30	4800	0.34
M2 Six point fix x^1, y^1, x^2 then x^1, y^1, y^2	18	994	0.24
M3 Six point fix x^1, y^1	24	4763	0.62
M4 Six point orthogonal perturbation	18	703	0.28
M5 $\mathbf{P}^2$ $\mathbf{P}^3$	24	2765	0.44
M6 Elements of $\mathbf{T}$	26	848	0.32
M7 Linear	27	1	1.41

Table 2: *The DOF, average number of evaluations of the total cost function D (given in Equation (8)) in the gradient descent algorithm, and the standard deviation σ_p (10) for the perfect synthetic point data.*

trifocal tensor from 7 points linearly as an eigenvector of a 27×27 matrix [29], but it is not consistent. Synthetic data were randomly generated in three space; 100 sets of 100 corresponding triples were generated. The image data were perturbed by Gaussian noise, standard deviation 1.0, and then quantized to the nearest 0.1 pixel. Mismatched features were then introduced to make a given percentage of the total, between 10 and 50 percent. Only the 50-90 percent of inlying data were used to assess the result the ground truth error for each estimator (as there is no ground truth defined for the outliers). It was found that the 6 point method performed better than the 7 point giving a standard deviation around $10 - 20\%$ lower. This is for two reasons, the first being that that the six point algorithm requires fewer correspondences to estimate and so has less chance of including an outlier; the second and perhaps more important is that the six point algorithm exactly encodes the constraints on the parameters of the trifocal constraint. The seven point algorithm on the other hand has too many degrees of freedoms, 26 when there should only be 18. Generally about one third of the 6 point samples produce one solution, the rest three solutions.

The six point algorithm is also considerably faster, in the case of the seven point algorithm the eigenvector of a 27×27 matrix must be found, which is slower than the solution of the cubic described in Appendix A. Furthermore far fewer six point samples need to be taken to get a given degree of confidence in the result, given a certain proportion of outliers. After stage 2 the σ (10) of the data was 2.21.

The second test compared the **M0-7** parametrizations. Synthetic test were conducted with noise on the pixel at 0.5 for 100 outlier free correspondences, the results are shown in Table 2. It can be seen that all of the methods dramatically reduce the error from that provided by the initial RANSAC estimate. It is interesting to observe that the linear method is not enormously worse than the non-linear. The parametrizations in terms of the 6 point bases outperform **M5**, the **P** parametrization. The difference between the 6 point methods is less extreme. As expected minimal parametrizations require far fewer function evaluations. The following test demonstrates the results of the algorithm **M2** on two typical real examples.

Chapel Sequence Figures 2 (a)-(c) shows three views of an outdoor chapel, the camera moves around the chapel rotating to keep it in view; (g)-(i) show the corner features detected in each image. Figures 3 shows the results for the point matches and Figure 4 for the line

matches. As the minimization progresses the basis points move only a few hundredths of a pixel each. Generally the matching process worked quite well, but line matching is still the poor cousin of point matching, and requires improvement. It is hoped that implementation of the general maximum likelihood formulation for points and lines given in Appendix C will improve matters.

Model Sequence Figures 2 (d)-(f) shows three views of an indoor model house, and (j)-(l) all the corners detected. Figures 5 (a)-(c) show the matched corners superimposed on each image. It can be seen that the corners stick well to 3D scene features such as the house chimney etc. Again coordinates of the point basis move on average only a few hundredths of a pixel. Figures 5 (d)-(f) show the matches, inliers and outliers provided by the whole algorithm. The rotation of the house can be clearly seen. All the detected lines are shown in Figure 6 (a)-(c) and from these matched lines are shown in Figure 6 (d)-(f).

8 Conclusions and Future Work

Why does the point parametrization work so well? One reason is that the 6 points initially selected by RANSAC are known to provide a good estimate of the trifocal tensor (because there is a lot of support for this solution). Hence the initial estimate of the six point basis provided by RANSAC 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 trifocal tensors that might actually arise are searched for.

The Hartley linear algorithm currently takes about as much time to run as one Huber cost function evaluation. Thus the MLE algorithm has a computational cost two orders of magnitude higher than the linear, but only reduces the error by a factor of 5 or 6. However, the news is not so bleak because firstly, the linear algorithm requires a robust addition if it is to be used for data contaminated by mis-matches, and secondly the linear method does not return a tensor whose elements satisfy the constraints. Further computational cost is required to coerce the tensor into this form, and methods to achieve this are still a research issue. Although the MLE cost function is certainly a standard to be measured against, it may well be that cost functions exist which are cheaper to evaluate and are almost as veridical, or that the MLE cost function can be implemented more efficiently.

The general method (of minimal parametrization in terms of basis points found from RANSAC) could be used for any other estimation problem in vision, for instance estimating the fundamental matrix, projectivities, camera matrices etc. Such extensions are explored in Torr and Zisserman [27]. Rather than using RANSAC to estimate the starting parameters the random sample that minimises the Huber cost function could be selected. This has the advantage that the robust estimator would then be minimizing the MLE cost function that we have derived in this paper as the optimal cost function. This method has been implemented and is reported in [27].

In conclusion, the methodology is general and could be used outside of vision in any problem where minimal parametrizations are not immediately obvious, and the constraints may be determined from some minimal number of points.

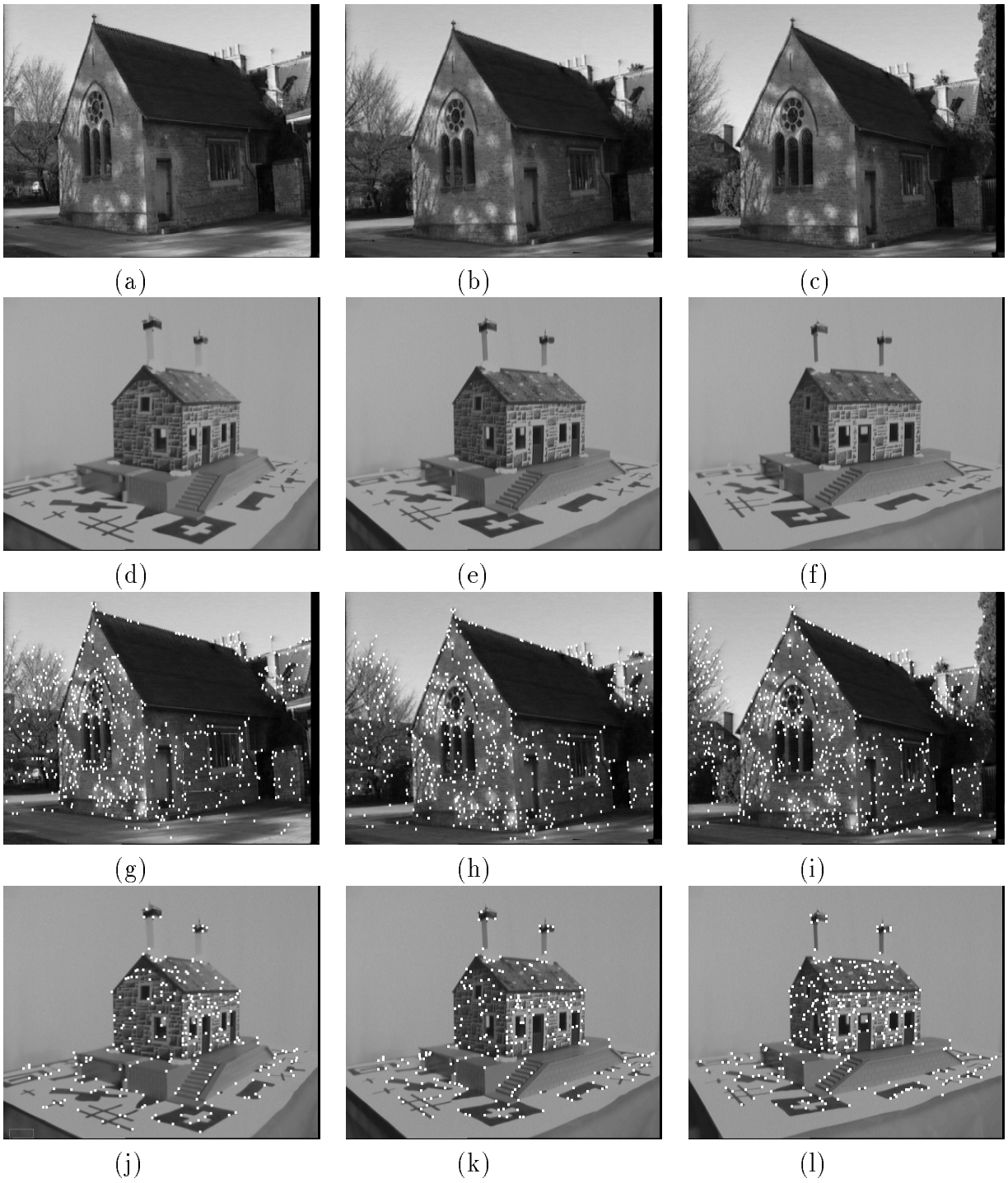

Figure 2: *The two triplets of test images: (a) (b) (c) Three images from a sequence of a chapel, acquired by a hand-held camcorder. Camera motion is lateral and a few centimetres between frames. The image size is* 760×550 *pixels. (d) (e) (f) Three images of a model house, the house rotates. (g) (h) (i) The initial set of detected corners on the chapel sequence, and (j) (k) (l) on the model.*

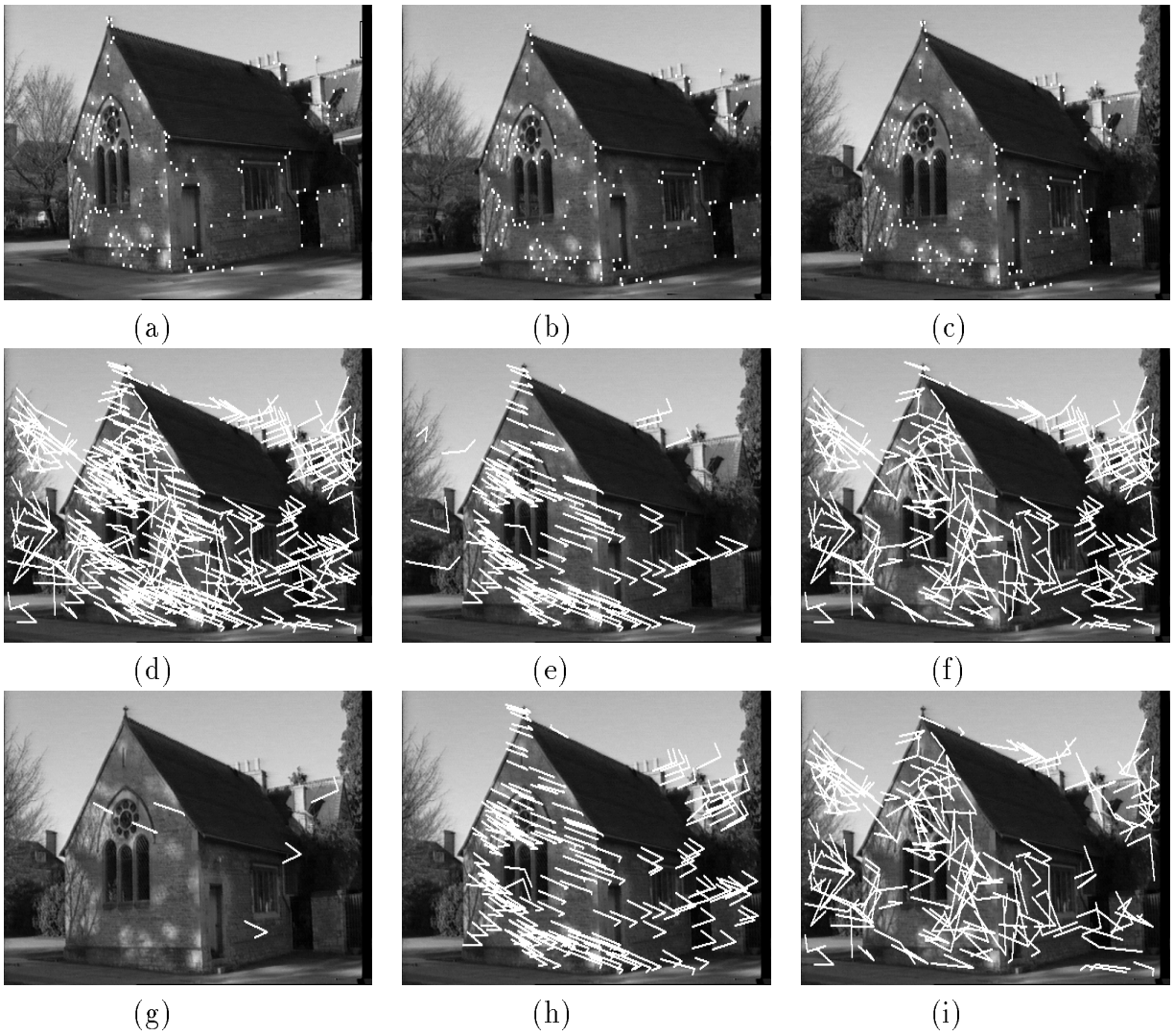

Figure 3: **(a) (b) (c)** *The matched corners shown superimposed on each image.* **(d) (e) (f)** *Matches are shown for all three images, with the features joined together and superimposed on the third image. There are 341 corners tracked over all three images (from about 500 tracked pairwise).* **(d)** *all the matches,* **(e)** *the inliers after RANSAC and* **(f)** *outliers. There are 146 inliers. After RANSAC the variance of d (which is a χ^2 variable with 3 degrees of freedom) is 1.621.* **(g) (h) (i)** *Showing the results after the non-linear minimization.* **(g)** *the six point basis selected by RANSAC.* **(h)** *the inliers after the non linear scheme,* **(i)** *outliers. After the non-linear scheme the number of inliers has increased to 184. The variance of d for the inliers has decreased to 0.7538. Note all the outliers on the tree to the left have been removed.*

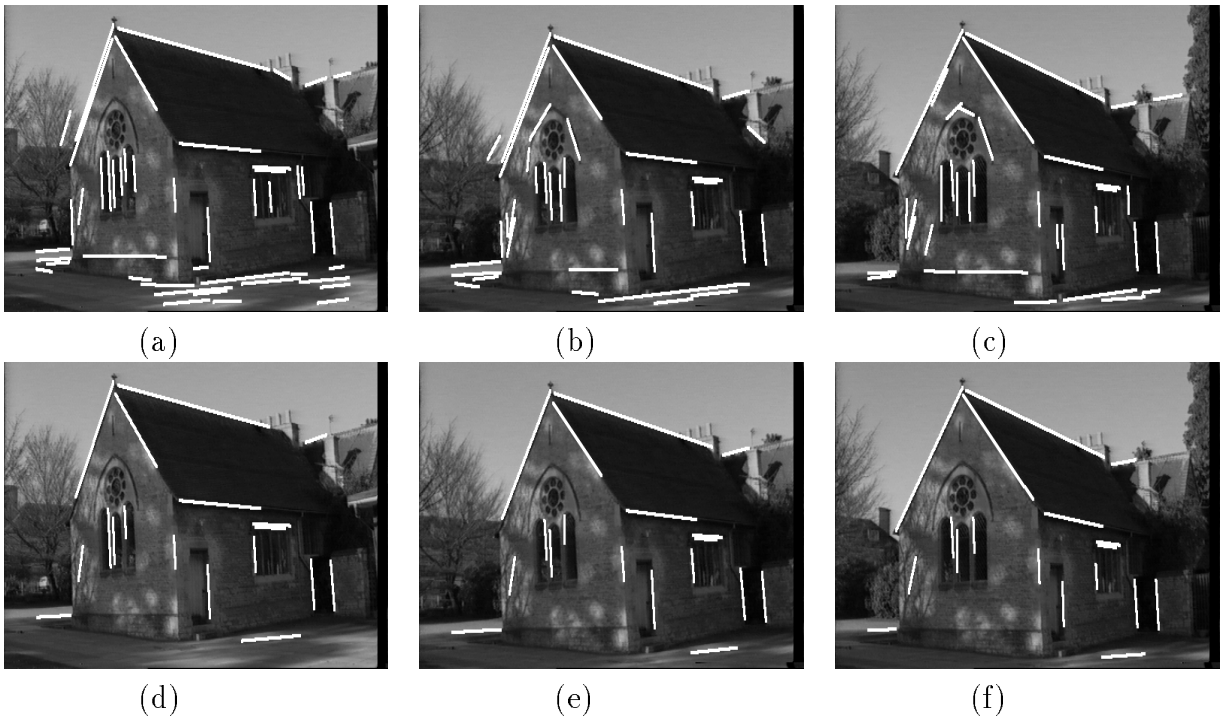

Figure 4: **(a) (b) (c)** *All the lines that are detected over the three images are shown superimposed on views 1,2 and 3 respectively of the chapel.* **(d) (e) (f)** *The inlying lines matched shown in views 1,2 and 3 respectively.*

Acknowledgements We thank Stéphane Laveau for discussions and comments on the original manuscript, Ian Reid, Long Quan and Olivier Faugeras for comments on the original manuscript, and Paul Beardsley for line detection software. Financial support was provided by ACTS Project Vanguard.

A Computation of the trifocal tensor from six point correspondences

In this appendix the method for finding the trifocal tensor from six points is detailed, the method is inspired by the method of Quan [20] the derivation follows that given in Weinshall *et al.* [32].

The algorithm requires six space points in general position, otherwise the trifocal tensor cannot be uniquely determined. The six points can be assigned canonical projective coordinates as follows: $(1,0,0,0)^\top$, $(0,1,0,0)^\top$, $(0,0,1,0)^\top$, $(0,0,0,1)^\top$, $(1,1,1,1)^\top$ and $(X,Y,Z,W)^\top$ where X,Y,Z,W are unknown. Similarly, and without loss of generality the image coordinates of the first four points in each image are assigned to the projective basis in each image, i.e. the coordinates of the six image points are $(1,0,0)^\top$, $(0,1,0)^\top$, $(0,0,1)^\top$, $(1,1,1)^\top$, $(x_5^{(i)},y_5^{(i)},w_5^{(i)})^\top$, and $(x_6^{(i)},y_6^{(i)},w_6^{(i)})^\top$; and $\mathbf{B}^{(i)}$ is the 3×3 projectivity that takes the image coordinates into this canonical frame. It is a simple matter to calculate $\mathbf{B}^{(i)}$

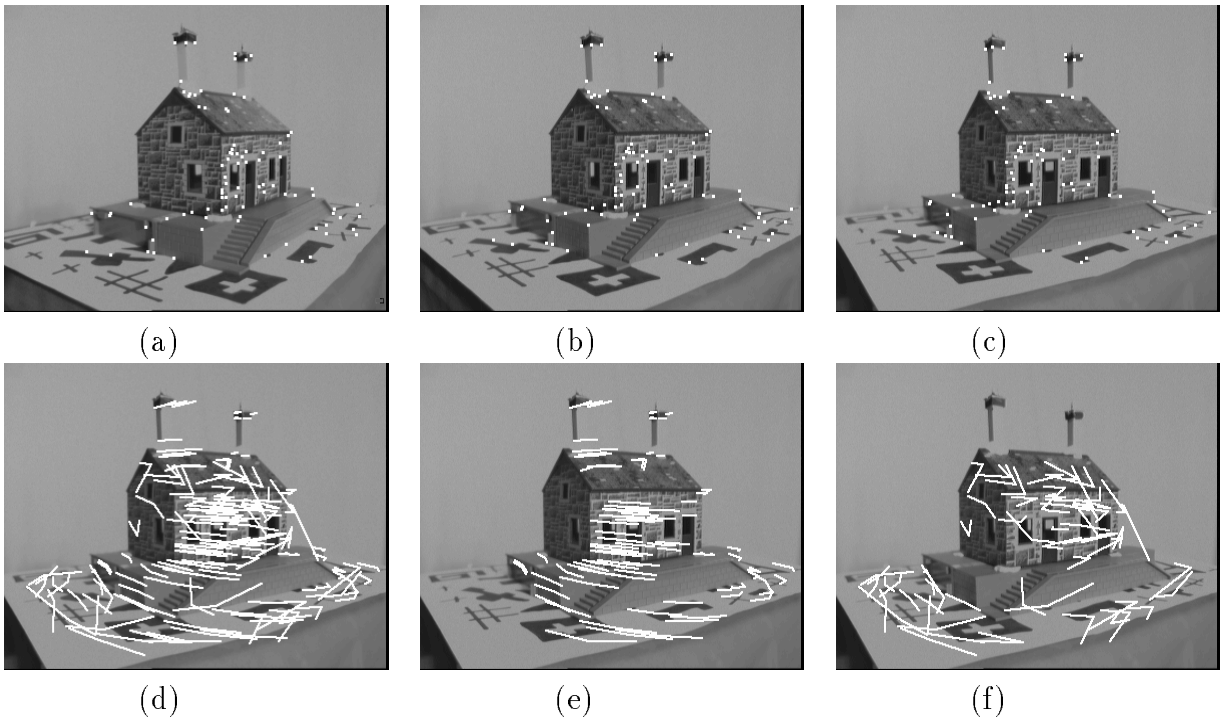

Figure 5: **(a) (b) (c)** *Corner features matched over the three images are shown superimposed on images 1, 2 and 3 respectively of the model house. Examination of the chimney and other salient features shows that the corners have become fixed to scene features.* **(d) (e) (f)** *matched points, inliers and outliers3 , the corner from all three images are shown on the third, joined by lines from the corner in image 1 to 2, and 2 to 3. Observing (e) it is possible to see the rotation of the model house.*

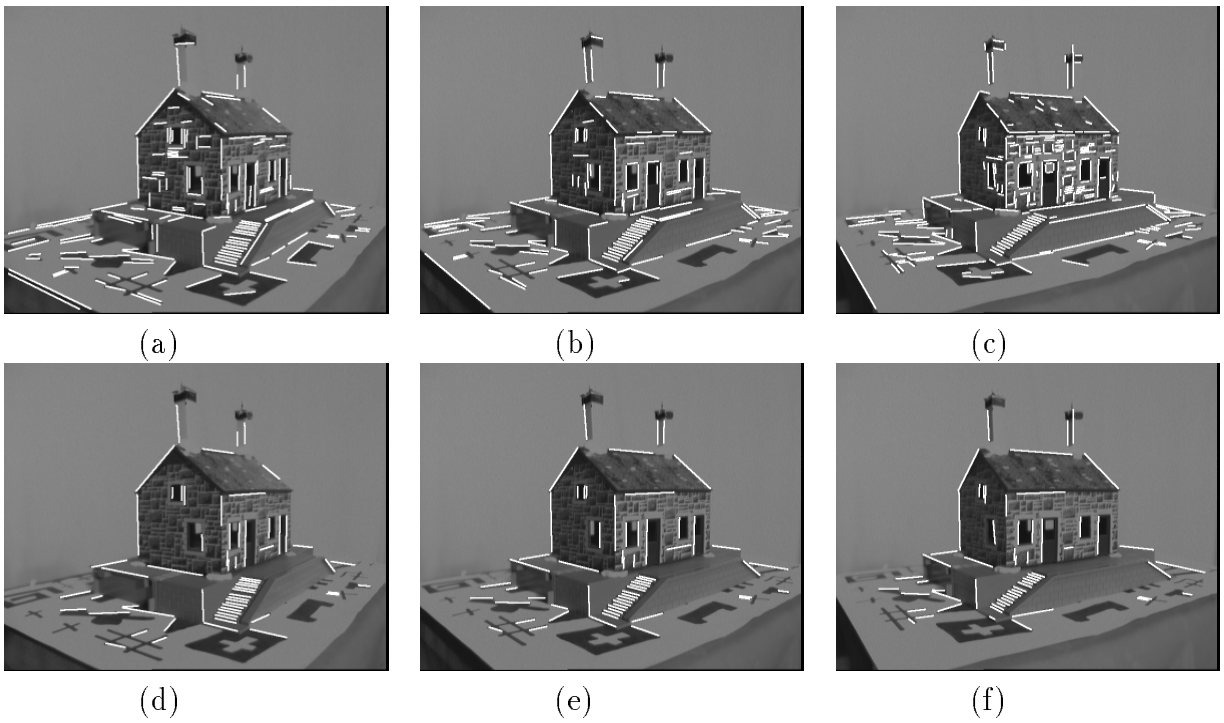

Figure 6: **(a) (b) (c)** *All the detected lines on the model house (i.e. raw input to the algorithm).* **(d) (e) (f)** *inlying matched lines provided by the non-linear estimator, lines, shown superimposed on each view. Note that the lines on the stairs of the house have been successfully matched which indicates the high degree of accuracy of our recovered constraint.*

efficiently: if $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$ and $\mathbf{x}_4$ are to be transformed to a canonical frame then

$$\mathbf{B}^{(i)} = \begin{bmatrix} \lambda_1\mathbf{x}_1 & \lambda_2\mathbf{x}_2 & \lambda_3\mathbf{x}_3 \end{bmatrix}^{-1} \tag{12}$$

where

$$\begin{pmatrix} \lambda_1 & \lambda_2 & \lambda_3 \end{pmatrix}^\top = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \mathbf{x}_3 \end{bmatrix}^{-1} \mathbf{x}_4 \tag{13}$$

Once the canonical system is set up,

$$\begin{bmatrix} 1 & 0 & 0 & 1 & x_5^{(i)} & x_6^{(i)} \\ 0 & 1 & 0 & 1 & y_5^{(i)} & y_6^{(i)} \\ 0 & 0 & 1 & 1 & w_5^{(i)} & w_6^{(i)} \end{bmatrix} = \begin{bmatrix} \alpha^{(i)} & 0 & 0 & \delta^{(i)} \\ 0 & \beta^{(i)} & 0 & \delta^{(i)} \\ 0 & 0 & \gamma^{(i)} & \delta^{(i)} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 1 & X \\ 0 & 1 & 0 & 0 & 1 & Y \\ 0 & 0 & 1 & 0 & 1 & Z \\ 0 & 0 & 0 & 1 & 1 & W \end{bmatrix} \tag{14}$$

for each image $i = 1, 2, 3$. Thus recovery of the trifocal tensor is equivalent to recovering the coordinates of the sixth point $(X, Y, Z, W)^\top$ and $(\alpha^{(i)}, \beta^{(i)}, \gamma^{(i)}, \delta^{(i)})$ for each camera.

From (14) the values of the sixth space point and camera parameters may be obtained in terms of the fifth and sixth image coordinate as follows:

$$\frac{x_5^{(i)}}{w_5^{(i)}} = \frac{\alpha^{(i)} + \delta^{(i)}}{\gamma^{(i)} + \delta^{(i)}}$$

$$\begin{aligned}
\frac{y_5^{(i)}}{w_5^{(i)}} &= \frac{\beta^{(i)} + \delta^{(i)}}{\gamma^{(i)} + \delta^{(i)}} \\
\frac{x_6^{(i)}}{w_6^{(i)}} &= \frac{\alpha^{(i)} X + \delta^{(i)} W}{\gamma^{(i)} Z + \delta^{(i)} W} \\
\frac{y_6^{(i)}}{w_6^{(i)}} &= \frac{\beta^{(i)} Y + \delta^{(i)} W}{\gamma^{(i)} Z + \delta^{(i)} W}
\end{aligned}$$

or

$$\begin{bmatrix} w_5^{(i)} & 0 & -x_5^{(i)} & w_5^{(i)} - x_5^{(i)} \\ 0 & w_5^{(i)} & -y_5^{(i)} & w_5^{(i)} - y_5^{(i)} \\ w_6^{(i)} X & 0 & -x_6^{(i)} Z & w_6^{(i)} W - x_6^{(i)} W \\ 0 & w_6^{(i)} Y & -y_6^{(i)} Z & w_6^{(i)} W - y_6^{(i)} W \end{bmatrix} \begin{pmatrix} \alpha^{(i)} \\ \beta^{(i)} \\ \gamma^{(i)} \\ \delta^{(i)} \end{pmatrix} = 0 \ . \tag{15}$$

The 4×4 matrix on the left is has rank 3 [20] and is given purely in terms of the image coordinates and the sixth space point. As it is of rank 3 the determinant must be zero:

$$\begin{aligned}
(-x_5^{(i)} y_6^{(i)} + x_5^{(i)} w_6^{(i)})(WX - YZ) + (x_6^{(i)} y_5^{(i)} - y_5^{(i)} w_6^{(i)})(WY - YZ) + \\
(-x_6^{(i)} w_5^{(i)} + y_6^{(i)} w_5^{(i)})(WZ - YZ) + (-x_5^{(i)} w_6^{(i)} + y_5^{(i)} w_6^{(i)})(XY - YZ) + \\
(x_5^{(i)} y_6^{(i)} - y_6^{(i)} w_5^{(i)})(XZ - YZ) &= 0 \ .
\end{aligned}$$

This is true in each of the three images:

$$\begin{bmatrix} (-x_5^{(1)} y_6^{(1)} + x_5^{(1)} w_6^{(1)}) & (x_6^{(1)} y_5^{(1)} - y_5^{(1)} w_6^{(1)}) & (-x_6^{(1)} w_5^{(1)} + y_6^{(1)} w_5^{(1)}) \\ (-x_5^{(2)} y_6^{(2)} + x_5^{(2)} w_6^{(2)}) & (x_6^{(2)} y_5^{(2)} - y_5^{(2)} w_6^{(2)}) & (-x_6^{(2)} w_5^{(2)} + y_6^{(2)} w_5^{(2)}) \\ (-x_5^{(3)} y_6^{(3)} + x_5^{(3)} w_6^{(3)}) & (x_6^{(3)} y_5^{(3)} - y_5^{(3)} w_6^{(3)}) & (-x_6^{(3)} w_5^{(3)} + y_6^{(3)} w_5^{(3)}) \end{bmatrix}$$

$$\begin{bmatrix} (-x_5^{(1)} w_6^{(1)} + y_5^{(1)} w_6^{(1)}) & (x_5^{(1)} y_6^{(1)} - y_6^{(1)} w_5^{(1)}) \\ (-x_5^{(2)} w_6^{(2)} + y_5^{(2)} w_6^{(2)}) & (x_5^{(2)} y_6^{(2)} - y_6^{(2)} w_5^{(2)}) \\ (-x_5^{(3)} w_6^{(3)} + y_5^{(3)} w_6^{(3)}) & (x_5^{(3)} y_6^{(3)} - y_6^{(3)} w_5^{(3)}) \end{bmatrix} \quad \mathbf{t} = \quad 0 \tag{16}$$

therefore the vector $\mathbf{t} = (WX - YZ, WY - YZ, WZ - YZ, XY - YZ, XZ - YZ)$ lies in the null space of the matrix on the left, which, if the points are in general position has rank 3. The two dimensional null space of the matrix in (16) may be recovered by a singular value decomposition. Let $\mathbf{t}_1$ and $\mathbf{t}_2$ be the two vectors spanning this null space. Then up to a scale factor $\mathbf{t} = \mathbf{t}_1 + \alpha \mathbf{t}_2$. There is a cubic constraint on the elements of $\mathbf{t}$: if $\mathbf{t} = (t_1, t_2, t_3, t_4, t_5)$ then $t_1 t_2 t_5 - t_2 t_3 t_5 - t_2 t_4 t_5 = t_1 t_3 t_4 - t_2 t_3 t_4 - t_3 t_4 t_5$ which can be proven by substituting the X, Y, Z, W values of the t_i. Imposing this cubic constraint on $\mathbf{t} = \mathbf{t}_1 + \alpha \mathbf{t}_2$ leads to a cubic equation in α with one or three real solutions for $\mathbf{t}$. Given $\mathbf{t}$ then $(X, Y, Z, W)^\top$ may be recovered as follows:

$$\frac{X}{W} = \frac{t_4 - t_5}{t_2 - t_3} \qquad \frac{Y}{W} = \frac{t_4}{t_1 - t_3} \qquad \frac{Z}{W} = \frac{t_5}{t_1 - t_2} \tag{17}$$

assuming that $W \neq 0$. If $W = 0$ then the sixth point is on the plane at infinity and it is trivial to use an alternative set of equations to recover $(X, Y, Z, W)^\top$. Given $(X, Y, Z, W)^\top$ the parameters of the camera matrices $(\alpha^{(i)}, \beta^{(i)}, \gamma^{(i)}, \delta^{(i)})$ may be recovered in a linear manner from (14).

From the camera matrices the structure may be initialised directly in the original coordinate system. The camera matrices are:

$$\mathbf{P}^{(i)} = \mathbf{B}^{-1} \begin{bmatrix} \alpha^{(i)} & 0 & 0 & \delta^{(i)} \\ 0 & \beta^{(i)} & 0 & \delta^{(i)} \\ 0 & 0 & \gamma^{(i)} & \delta^{(i)} \end{bmatrix} . \tag{18}$$

To recover the trifocal tensor the first camera is set to $[\mathbf{I}|\mathbf{0}]$ (effected by a simple transformation of the coordinates) then the trifocal tensor's coefficients are given by (3). Alternatively, the trifocal tensor elements can be computed directly from the determinants of 4×4 matrices constructed from the rows of the camera matrices [7], without requiring the first camera to have a canonical form.

B A First Order Approximation to Point Error

Given a (measured) point correspondence $\mathbf{x} \leftrightarrow \mathbf{x}^2 \leftrightarrow \mathbf{x}^3$, or $\mathbf{x}^{1,2,3}$, over three images and a trifocal tensor $\mathbf{T}$ we require for the MLE cost function the correspondence $\hat{\mathbf{x}}^{1,2,3}$ which is closest to $\mathbf{x}^{1,2,3}$ whilst exactly satisfying the 9 trilinear relations arising from the trifocal tensor. Both this correspondence $\hat{\mathbf{x}}^{1,2,3}$ and the minimum distance d can be obtained as a first order approximation. This approximation is based on the works of Taubin [23] and Kanatani [16] on parametric surface fitting.

Consider the space $\mathcal{R}^6$ formed by joining together the three images with coordinate system $(x_1^1, x_2^1, x_1^2, x_2^2, x_1^3, x_2^3)$. The trifocal tensor $\mathbf{T}$ can be thought of as a dimension three[1] variety in $\mathcal{R}^6$. The variety can be defined from the nine trilinear point relations (1): Noting these trilinearities as $t^i(\hat{\mathbf{x}}^{1,2,3}) = 0$, $i = 1..9$. Then the variety $\mathbf{T}$ defines a set $Z(T) = \{\hat{\mathbf{x}}^{1,2,3} : t^i(\hat{\mathbf{x}}^{1,2,3}) = 0, i = 1..9\}$. The set of polynomials may be written as a vector equation $\mathbf{t}(\hat{\mathbf{x}}^{1,2,3}) = 0$.

We wish to find a point $\hat{\mathbf{x}}^{1,2,3}$ which satisfies $\mathbf{t}(\hat{\mathbf{x}}^{1,2,3}) = 0$, and minimises $||\hat{\mathbf{x}}^{1,2,3} - \mathbf{x}^{1,2,3}||$. The minimum distance d is the orthogonal distance of the point $\mathbf{x}^{1,2,3}$ in $\mathcal{R}^6$ to the variety.

Consider a correspondence $\mathbf{x}^{1,2,3}$ which is not on $Z(T)$, then

$$\mathbf{t}(\mathbf{x}^{1,2,3}) = \mathbf{r}$$

a non-zero nine-vector of residuals. A Taylor expansion of $\mathbf{t}$ about $\mathbf{x}^{1,2,3}$ gives:

$$\mathbf{t}(\hat{\mathbf{x}}^{1,2,3}) = \mathbf{t}(\mathbf{x}^{1,2,3}) + \mathbf{J}|_{\mathbf{X}^{1,2,3}} \left(\hat{\mathbf{x}}^{1,2,3} - \mathbf{x}^{1,2,3}\right) = \mathbf{r} + \mathbf{J}|_{\mathbf{X}^{1,2,3}} \left(\hat{\mathbf{x}}^{1,2,3} - \mathbf{x}^{1,2,3}\right) \tag{19}$$

[1] The reason that the variety has dimension four can be seen from the fact that each point on the variety also corresponds to a unique point in the world. Thus the (Euclidean) coordinates (X, Y, Z) of world points provide a natural coordinate system for the intrinsic coordinates on the variety.

to first order, where $\mathbf{J}$ is the Jacobian of the set of trilinearities

$$\mathbf{J} = \begin{bmatrix} \frac{\partial t^1(\mathbf{X}^{1,2,3})}{\partial x_1^1} & \cdots & \frac{\partial t^1(\mathbf{X}^{1,2,3})}{\partial x_2^3} \\ \vdots & \vdots & \vdots \\ \frac{\partial t^9(\mathbf{X}^{1,2,3})}{\partial x_1^1} & \cdots & \frac{\partial t^9(\mathbf{X}^{1,2,3})}{\partial x_2^3} \end{bmatrix}_{9\times 6} = \begin{bmatrix} \nabla^\top t^1(\mathbf{x}^{1,2,3}) \\ \vdots \\ \nabla^\top t^9(\mathbf{x}^{1,2,3}) \end{bmatrix} . \tag{20}$$

The Jacobian has rank three for general points on $Z(T)$, which corresponds to the dimension of the tangent plane.

Returning to the Taylor expansion (19). Since $\mathbf{t}(\hat{\mathbf{x}}^{1,2,3}) = 0$ it follows that

$$\mathbf{r} = -\mathbf{J}\left(\hat{\mathbf{x}}^{1,2,3} - \mathbf{x}^{1,2,3}\right) \tag{21}$$

The pseudo-inverse of $\mathbf{J}$ provides a solution for the vector that minimizes $||\hat{\mathbf{x}}^{1,2,3} - \mathbf{x}^{1,2,3}||$. It follows that

$$\hat{\mathbf{x}}^{1,2,3} = \mathbf{x}^{1,2,3} - \mathbf{J}^+\mathbf{r} \tag{22}$$

where $\mathbf{J}^+$ is the pseudo-inverse of $\mathbf{J}$ returned by a singular value decomposition (SVD) thus the square of the distance is given by

$$d^2 = ||\hat{\mathbf{x}}^{1,2,3} - \mathbf{x}^{1,2,3}||^2 = \mathbf{r}^\top(\mathbf{J}\mathbf{J}^\top)^+\mathbf{r}$$

If the set of trilinearities are linear, which is the case under orthographic projection [21, 24] or under affine [19] viewing conditions, then this approximation is exact.

C General Covariances

Within this appendix the maximum likelihood arguments for points and lines are generalized to arbitrary covariance matrices on the data.

If the covariance matrix of $\mathbf{x}^{1,2,3}$ is not the identity, then the MLE requires a Mahalanobis distance for the point error. Let $\Sigma_x^{1,2,3}$ be the 6×6 covariance matrix of $\mathbf{x}^{1,2,3} = (x_1^1, x_2^1, x_1^2, x_2^2, x_1^3, x_2^3)^\top$. Then the point error

$$d^2 = \sum_j \left(\hat{\mathbf{x}}^j - \mathbf{x}^j\right)^2 = \left(\hat{\mathbf{x}}^{1,2,3} - \mathbf{x}^{1,2,3}\right)^2$$

is replaced by the Mahalanobis distance

$$d^2|_{\Sigma_x^{1,2,3}} = \left(\hat{\mathbf{x}}^{1,2,3} - \mathbf{x}^{1,2,3}\right)\left(\Sigma_x^{1,2,3}\right)^{-1}\left(\hat{\mathbf{x}}^{1,2,3} - \mathbf{x}^{1,2,3}\right) \quad .$$

A first order approximation to the maximum likelihood estimate of the point is

$$\hat{\mathbf{x}}^{1,2,3} = \mathbf{x}^{1,2,3} - \Sigma_x^{1,2,3}\mathbf{J}^\top \left(\mathbf{J}\Sigma_x^{1,2,3}\mathbf{J}^\top\right)^+ \mathbf{r} \tag{23}$$

which leads to

$$d^2|_{\Sigma_x^{1,2,3}} = \mathbf{r}^\top(\mathbf{J}\Sigma_x^{1,2,3}\mathbf{J}^\top)^+\mathbf{r} \quad . \tag{24}$$

Similarly, for lines: Consider a line correspondence over three images $\mathbf{l}^{1,2,3}$. Each line may be parametrized as $\mathbf{l}^1 = (l_1^1, l_2^1, 1)$ (for ease of exposition we shall temporarily ignore the singularity—lines through the origin) with covariance matrix [4, 6] Σ_l^1 etc. Then $\mathbf{l}^{1,2,3}$ may written as a 6 vector of the homogeneous coordinates. The constraints on lines given by equation (2), written as a vector equation $\mathbf{y}(\hat{\mathbf{l}}^{1,2,3}) = 0$, define a variety $Y(T) = \{\hat{\mathbf{l}}^{1,2,3} : y^i(\hat{\mathbf{l}}^{1,2,3}) = 0, i = 1..3\}$. Thus a situation analogous to that defined for point correspondences has been arrived at; each line correspondence may be represented as a point in $\mathcal{R}^6$ (the space of line parameters) which is constrained to lie on a variety of dimension four[2].

In general noisy lines $\mathbf{l}^{1,2,3}$ do not lie on the variety Y and an error measure for the MLE results from the Mahalanobis distance:

$$e^2|_{\Sigma_l^{1,2,3}} = \left(\hat{\mathbf{l}}^{1,2,3} - \mathbf{l}^{1,2,3}\right)\left(\Sigma_l^{1,2,3}\right)^{-1}\left(\hat{\mathbf{l}}^{1,2,3} - \mathbf{l}^{1,2,3}\right)$$

following a general formulation given in Kanatani [16].

A first order approximation to both the line errors can be obtained in a similar manner to that given for the point errors. The two may then be combined to give the overall robust cost function

$$D|_\Sigma = \sum_i \gamma\left(d|_{\Sigma_x^{1,2,3}}\right) + \sum_k \gamma(e|_{\Sigma_l^{1,2,3}}) \quad , \tag{25}$$

the minimization of which gives a first order approximation to the robust maximum likelihood solution.

References

[1] M. Armstrong, A. Zisserman, and R. Hartley. Self-calibration from image triplets. In B. Buxton and Cipolla R., editors, *Proc. 4th European Conference on Computer Vision, LNCS 1064, Cambridge*, pages 3–16. Springer–Verlag, 1996.

[2] 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]The reason that the variety has dimension three can be seen from the fact that each point on the variety also corresponds to a unique 3D line in the world; each of which has four degrees of freedom.

[3] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:679–698, 1986.

[4] J. C. Clarke. First order error propagation: A primer. OUEL report.

[5] R. Deriche, Z. Zhang, Q. T. Luong, and O. Faugeras. Robust recovery of the epipolar geometry for an uncalibrated stereo rig. In J. O. Eckland, editor, *Proc. 3rd European Conference on Computer Vision, LNCS 800/801, Stockholm*, pages 567–576. Springer-Verlag, 1994.

[6] O.D. Faugeras. *Three-Dimensional Computer Vision: A Geometric Viewpoint.* The MIT Press, 1993.

[7] O.D. Faugeras and B. Mourrain. On the geometry and algebra of the point and line correspondences between n images. In *Proc. 5th Int'l Conf. on Computer Vision, Boston*, pages 951–962, 1995.

[8] O.D. Faugeras and L. Robert. What can two images tell us about a third one. In J. O. Eckland, editor, *Proc. 3rd European Conference on Computer Vision, LNCS 800/801, Stockholm*, pages 485–492. Springer-Verlag, 1994.

[9] 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.

[10] 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.

[11] C. Harris and M. Stephens. A combined corner and edge detector. In *Proc. Alvey Conf.*, pages 189–192, 1987.

[12] R. I. Hartley. Lines and points in three views – a unified approach. In *ARPA Image Understanding Workshop, Monterey*, 1994.

[13] R. I. Hartley. Projective reconstruction from line correspondences. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 1994.

[14] R. I. Hartley. In defence of the 8-point algorithm. In *Proc. 5th Int'l Conf. on Computer Vision, Boston*, pages 1064–1075, 1995.

[15] P. J. Huber. *Robust Statistics.* John Willey and Sons, 1981.

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

[17] S. Laveau. *Geometry of a system of N cameras. Theory, estimation and applications.* PhD thesis, INRIA, 1996.

[18] 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.

[19] J. Mundy and A. Zisserman. *Geometric Invariance in Computer Vision.* MIT press, 1992.

[20] L. Quan. Invariants of 6 points from 3 uncalibrated images. In J. O. Eckland, editor, *Proc. 3rd European Conference on Computer Vision, LNCS 800/801, Stockholm*, pages 459–469. Springer-Verlag, 1994.

[21] A. Shashua. Trilinearity in visual recognition by alignment. In *Proc. 3rd European Conference on Computer Vision, LNCS 800/801, Stockholm*, volume 1, pages 479–484, May 1994.

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

[23] 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.

[24] P. H. S. Torr. *Outlier Detection and Motion Segmentation.* PhD thesis, University of Oxford, Engineering Dept., 1995.

[25] 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.

[26] 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.

[27] P. H. S. Torr and A Zisserman. Computing multiple view relations. OUEL Report, 1997.

[28] P. H. S. Torr, A Zisserman, and S. Maybank. Robust detection of degenerate configurations for the fundamental matrix. Accepted to CVIU, 1996.

[29] 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.

[30] B. 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.

[31] T. Viéville and Q-T. Luong. Motion of points and lines in the uncalibrated case. Technical Report 2054, INRIA, 1993.

[32] D. Weinshall, M. Werman, and A. Shashua. Duality of multi-point and multi-frame geometry: Fundamental shape matrices and tensors. In B. Buxton and Cipolla R., editors, *Proc. 4th European Conference on Computer Vision, LNCS 1065, Cambridge*, pages 217–227. Springer–Verlag, 1996.

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

[34] J. Weng, T. Huang, and N. Ahuja. Motion and structure from line correspondences: Closed-form solution, uniqueness and optimization. *IEEE PAMI*, vol.14(3):318–336, 1992.

[35] 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.

[36] A. Zisserman and Maybank S. A case against epipolar geometry. In J. Mundy, A. Zisserman, and D. Forsyth, editors, *Applications of Invariance in Computer Vision LNCS 825.* Springer-Verlag, 1994.